Accelerated Reader

Evaluation Report

July 2021

Authors:

Alex Sutherland, Miriam Broeks, Sonia Ilie, Megan Sim, Joachim Krapels, Elena Rosa Brown, Julie Belanger. Advisor: Professor Anna Vignoles

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

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

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

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

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

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

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from RAND Europe and the University of Cambridge:

Alex Sutherland, Miriam Broeks, Megan Sim, Joachim Krapels, Elena Rosa Brown, Julie Belanger, Sonia Ilie, Anna Vignoles (as trial advisor)

The lead evaluator was initially Alex Sutherland, then Sonia Ilie.

Contact details:

Name: Dr Sonia Ilie

Address: Faculty of Education University of Cambridge, 184 Hills Rd, Cambridge, CB2 8PQ

Email: **isi22@cam.ac.uk**

Name: Miriam Broeks

Address: RAND Europe, Westbrook Centre, Milton Road, Cambridge, CB4 1YG

Email: **mbroeks@randeurope.org**

Executive summary

This report should be read in conjunction with the Accelerated Reader addendum report, which is included as appendix G.

The Project

Accelerated Reader (AR), developed by Renaissance Learning, is a digital whole-class reading management and monitoring programme that aims to foster independent reading among primary and secondary pupils. The project initially aimed to assess the impact of one year of AR for children in Year 4 (ages 8 to 9) and to also look at the longer-term impact of the programme by collecting Key Stage 2 (KS2) SATs reading scores for pupils who started the programme in Year 5 (ages 9 to 10).

Due to problems with the original outcome measure delivered to pupils at the end of Year 4, the length of the trial was extended and the primary outcome measure was changed as well as the main cohort of interest. The primary outcome became KS2 SATs reading scores for the pupils who started the programme in Year 5. This meant schools were asked to continue delivering AR for one additional year and resulted in some schools having a gap (of up to six months) in their implementation of AR or not delivering AR during the second year.

The evaluation of AR involved 181 primary schools and 5,759 children. Schools were randomly allocated to either receive the intervention or to a waitlist 'business as usual' control group: trial cohorts in control schools continued as usual but control schools were able to use AR with non-trial younger pupil cohorts from the second year of the trial onwards. The implementation and process evaluation (IPE) involved observations of training, analysis of usage data, and surveys, workshops, and interviews in intervention and control schools.

Key conclusions
1. Children who started Accelerated Reader in Year 5 made, on average, no additional progress in reading compared to children in the comparison schools. This result has a very high security rating.
2. Children eligible for free school meals who started Accelerated Reader in Year 5 made, on average, no additional progress in reading compared to FSM children in the comparison schools. However, this result has high statistical uncertainty.
3. Data during the first year of implementation indicated that AR was implemented as intended in intervention schools. During the additional (second) year of the trial one third of pupils were no longer accessing AR. Findings suggest that higher fidelity was not associated with better outcomes.
4. The implementation and process evaluation suggests that most of the 'business as usual' comparison schools had similar amounts of dedicated reading time and some used other evidence-based reading schemes and reading interventions.
5. AR was very well received by the vast majority of teachers, teaching assistants, and librarians who perceived positive impacts on pupil reading ability, reading stamina, and attitudes.

EEF security rating

The findings for pupils who started AR in Year 5 have a very high security rating. This trial was an effectiveness trial which tested the intervention's success under everyday conditions in a large number of schools. The schools receiving AR were similar to comparison schools in terms of their baseline characteristics, including prior attainment at Key Stage 1. Relatively few pupils who started the trial (6%) were not included in the final analysis.

Additional findings

Children who started Accelerated Reader in Year 5, on average, made no additional progress in reading compared to children in the comparison schools. This is our best estimate of impact, which has a very high security rating. As with any study, there is always some uncertainty around the result: the range of possible impacts for AR on Year 5 pupils

included small negative effects of one month less progress to positive effects of up to one month of progress. Children eligible for free school meals (FSM) who started Accelerated Reader in Year 5, on average, made no additional progress in reading compared to children in the comparison schools. The range of possible impacts of AR for this result included small negative effects of one month less progress to positive effects of up to two months' progress.

A previous EEF-funded efficacy trial of AR showed positive impacts. The first trial, conducted in four secondary schools with 349 Year 7 pupils who achieved below age-expected levels in reading at the end of Key Stage 2, found that pupils who received AR made three months of additional progress in reading comprehension. It is possible that the difference in results may be due to the change in age group, the use of AR with the whole class rather than a targeted group of pupils, the change in outcome measure, or the small number of pupils and schools involved in the first trial.

Pupils in business as usual comparison schools in this trial were found to be experiencing dedicated reading time comparable to intervention schools (as reported by headteachers). Analysis also highlighted that at least two-fifths of comparison schools that responded to surveys were using reading schemes with their Year 5 cohort and almost half were using reading interventions (such as Lexia, Catch Up Literacy, and Every Child a Reader), some of which have evidence of positive impact on pupil outcomes. This may mean there was not enough 'unique' activity in AR schools to produce measurable differences compared to comparison schools. The evaluation team highlight that if a highly professionalised, well-implemented, and widely used intervention such as AR does not have an impact on reading outcomes, then other, similar, interventions aimed at these year groups also may not.

Initially, the trial intended to test the effectiveness of AR on Year 4 pupils' reading comprehension after one year of intervention. However, most pupils scored highly on the reading comprehension measure (New Group Reading Test) so it could not be used as an outcome. The subsequent change to the research design meant most pupils received a greater dosage of the intervention. However, many pupils did not experience AR as intended. Some schools did not take up the offer of additional delivery and some schools experienced a gap in delivery of up to six months before they were onboarded for their second year.

Results suggest that there was a high-level of implementation of AR during year one of the trial and a fairly high-level of implementation during year two. Implementation was calculated by considering pupils' average score on quizzes, engaged reading time, whether appropriately-levelled books were selected, and teacher responsiveness to pupil reading data. Although implementation overall was described as 'high', it is of interest that the ideal engaged reading time of between 15–30 minutes a day was only achieved by between 38% to 43% of pupils in year one and between 30% to 42% of pupils in year two. This could provide another potential explanation for the null result in this trial.
This trial does not consider secondary outcomes, such as attitudes to reading, reading habit formation, and enjoyment of reading, which are all possible impacts of AR. Teacher survey and interview responses do indicate that teachers perceived AR impacted well on pupils' attitudes to reading.

Cost

The average cost of AR for one school was around £10,044 or £48 per pupil per year when averaged over three years. The main costs were for the annual licenses, a one-off training cost, and purchasing books at start-up and in subsequent years. Schools also needed to meet the cost of staff cover for training.

Impact

Table 1: Summary of impact on primary outcome

Outcome/ Group	Effect size (95% confidence Interval)	Estimated months' progress	EEF security rating	No. of pupils	P value	EEF cost rating
Reading Assessment – Key Stage 2	0.00 (-0.08, 0.09)	0	🔒🔒🔒🔒🔒	5,759	0.94	£ £ £ £ £
Reading Assessment – Key Stage 2 FSM	0.02 (-0.08, 0.12)	0	n/a	2,163	0.69	£ £ £ £ £

Introduction

Background evidence

The Renaissance Learning: Accelerated Reader programme (Accelerated Reader, AR) is a reading programme that aims to foster reading for pleasure and improve reading ability. AR is a whole-class approach that is premised on pupils practicing reading regularly in school, reading books at an appropriate level of difficulty (incorporating 'stretch' reading),1 testing pupil comprehension, and providing regular feedback to teachers.

Accelerated Reader is a well-established approach, with Renaissance Learning reporting over 7,000 schools using AR in the U.K. and Ireland, covering over 1.6 million pupils. Accelerated Reader has previously been evaluated a number of times, including in two U.S. randomised controlled trials and one U.K. efficacy trial, and is included on the What Works Clearinghouse (WWC) literacy list.2 The WWC states that for beginning readers, AR has a mixed effect on comprehension (based on two studies of 265 pupils) and no discernible effects on reading fluency (based on one study of 32 pupils).

Some of the existing evaluations utilise AR's own reading test (the STAR Reading test) as an outcome (for example, Bullock, 2005; Nunnery et al., 2006). For the purposes of this trial, STAR Reading is 'treatment inherent', meaning that it is too closely aligned to the intervention and does not reflect the impact on more general outcomes (Slavin and Madden, 2011).3 Other AR evaluations carried out outside of the U.K. have found positive, as well as mixed, effects of the programme (Ross et al., 2004; Nunnery et al., 2006; Nichols 2013; Huang, 2012; Shannon et al. 2015). For example, Nunnery et al. (2006) evaluated the reading attainment of 978 grade 3 to 6 students in a U.S. urban school district. Using the STAR Reading test as an outcome measure, they found a positive effect (d = +0.36) in the reading achievement of grade 3 pupils, while for grade 4 to 6 pupils with the same outcome measure the effect was positive but gradually smaller in older grades.

The programme has previously been tested through an EEF efficacy trial involving four secondary schools that ran their own randomised controlled trials with 350 Year 7 pupils (the schools were advised by EEF evaluators; see Gorard et al., 2015). The unit of randomisation was pupil in the four secondary schools. The results of these four studies were aggregated by an independent evaluator who found that AR had a modest positive effect on the overall reading scores of the treatment group compared to the control group (effect size +0.24). When evaluated in terms of students eligible for free school meals ('FSM pupils')—bearing in mind that they were not randomly assigned to groups—the effect was larger (+0.38), suggesting that AR may particularly improve the reading ability of Year 7 FSM pupils (but this finding was affected by the small sample of FSM pupils). However, the study has strengths of pupil-level randomisation and minimum attrition (<3%).

This evaluation builds on this evidence base with the largest independent evaluation of Accelerated Reader ever undertaken using an RCT. The evaluation tested the effectiveness of AR at scale in a large number of schools and under normal school conditions. Additionally, this new RCT focused on younger children in primary school, namely Years 4 and 5, in line with evidence of the benefits of earlier intervention (Doyle et al., 2009). Table **2** outlines the key characteristics of the present trial, and the earlier efficacy trial (Gorard et al., 2015) undertaken of AR in the English context, as well as any differences in the programme. Consistent with the objectives of the EEF, and distinct from the previous EEF evaluation, this trial was powered to detect an impact on disadvantaged (FSM) children. This evaluation also took advantage of EEF work as part of the North East Literacy Campaign,4 meaning that the study included schools from the North East, where evidence has shown relatively poorer levels of literacy compared to the rest of England (Dunatchik et al., 2018).

[1] 'Stretch and challenge' techniques refer to teaching practices to expand the ability of more able students in a classroom. This can be done by, for example, asking tailored questions to probe pupils' thinking further. For AR, this relates to pupils reading within the appropriate difficulty range for their reading ability, slowly growing towards the more difficult end to encourage progression.
[2] https://ies.ed.gov/ncee/wwc/InterventionReport/12
[3] https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/within-class-attainment-grouping/technical-appendix
[4] More information available at: https://educationendowmentfoundation.org.uk/scaling-up-evidence/campaigns/north-east-literacy-campaign/

Table 2: Comparison with previous efficacy trial (Gorard et al., 2015)

	This trial (Sutherland et al., 2019)	Efficacy trial (Gorard et al., 2015)
Type of trial	Waitlist, cluster-randomised	Waitlist, school-led
Number of arms	Two-arm	Two-arm
Unit of randomisation	School	Pupil
Number of schools in trial	181	4
Number of pupils in trial	6,116 (Year 5) 6,311 (Year 4)	349
Pupil stages	Year 4 and 5 (at trial start)	Year 7 (at trial start)
Pupil targeting characteristic	All pupils in the trial year groups Specific focus on FSM pupils (trial powered for subgroup analysis)	Pupils who had not achieved a secure level 4 in Key Stage 2 English
Stratification variable(s) (if applicable)	Geography: Region School-level attainment: low/high KS1 Average points score	None
Primary outcome variable	KS2 reading score	New Group Reading test (NGRT 3-A)
Baseline testing measure	Yes (KS1 reading)	TA-assessed KS2 scores
Programme implementation duration	Year 5: up to two academic years	20 weeks
Programme implementation variants	As per 2016 AR guidelines from Renaissance Learning, specifically, the 30 minutes per day reading target	As per 2013 AR guidelines from Renaissance Learning, specifically a 30–60 minutes per day reading target
Developer involvement in data provision	Yes (no involvement in randomisation)	No
School dropout	3%	None
Pupil dropout	6%	<3%

Intervention

Why: Accelerated Reader (AR) is a whole-class reading management and monitoring programme that aims to foster the habit of independent reading among primary and early secondary pupils. It aims to improve reading by enabling children to read books and other materials commensurate with their reading level but also to stretch pupils' reading.

What: AR uses internet-based software, initially to test pupils in order to determine their reading level (defined within the AR framework as the 'zone of proximal development' or ZPD score), and suggests a ZPD score which helps guide pupils to appropriate books for their reading age and reading interest. These tests, used to define pupils' ZPD, are referred to as STAR Reading tests. The STAR Reading tests may be taken more than once per year to update pupils' ZPD scores. As preparation for AR, schools are meant to scan and label each book in their library with their reading level, a task which is facilitated by another web service. Pupils are then encouraged to read books within their ZPD. In addition, pupils take computerised quizzes on the books they have read and earn Accelerated Reader points; these points are related to the difficulty and length of the book read and also to the proportion of correct answers for each quiz and focus on the comprehension of the written text. Teachers are able to produce progress reports on a weekly basis that summarise data on pupil reading progress based on quiz results. These can be used by teachers to identify pupils that may need most support and adapt their teaching as necessary.

How: The intervention logic for AR involves:

- tailoring assigned reading to pupils;
- the importance of reading within a pupil's ZPD, which includes reading at their current level and so-called 'stretch' books that are just above their current level (for example, Vygotsky, 1978);
- a constant feedback loop between reading and assessment driven by comprehension testing and feedback to teachers via an online platform; and
- protected time in the daily timetable to allow for reading to take place.[5]

AR also means that schools need to have:

- a range of books at levels appropriate to their pupils (which often means buying new books);
- adequate resources in terms of IT infrastructure; and
- teachers to guide readers—especially the youngest and weakest—on book selection within their ZPD; there is an element of staff training required and therefore a need to understand how well the training has worked.

All the above is intended to increase the amount of time children spend reading at school and thus 'ingrain' the habit of reading, leading to improvements in reading ability.

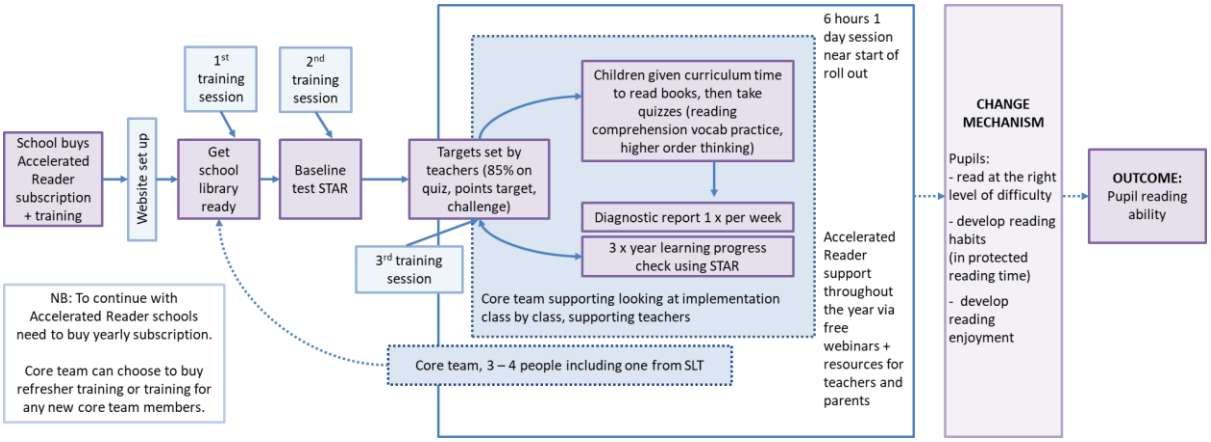
Figure **1** sets out the implementation logic model developed during kick-off meetings between the Education Endowment Foundation (EEF), Renaissance Learning (RL), the delivery partner National Foundation for Educational Research (NFER), and the evaluators (RAND Europe and University of Cambridge).

In the classroom, AR is implemented as follows:

- An initial adaptive online test (STAR Reading) is taken by pupils; this screens pupils for their reading levels and provides their ZPD score.
- Either the school or the classroom library is organised using AR by labelling each book with its corresponding ZPD score (which may be done through the ARbookfind website).
- Pupils are then recommended books to read that match their reading age and reading interest. Pupils are able to take quizzes on the books they have read on a computer or tablet and earn AR points related to difficulty. AR provides access to a database of more than 38,000 purpose-developed quizzes and taking the quizzes provides immediate feedback for pupils and reports for teachers.
- Teachers are then able to access data on how pupils are performing through their Interactive Reading Dashboard, which presents class-, group-, and individual-level pupil data. Both pupils and teachers can access these materials online via RL's website, Renaissance Place.

[5] The reading time non-negotiable aspect of AR is usually assumed to be 30 minutes' reading time per day, with an aim to achieve a minimum of 15 minutes engaged reading time. The latter figure is used in the Fidelity section.

Figure 1: Accelerated Reader logic model



Note: 'Core team' refers to the staff at the school designated to implement AR. This would mainly entail teachers but also teaching assistants, librarians, and a literacy coordinator (or another SLT representative).

When and how much: AR is delivered alongside the normal classroom curriculum but requires protected and adequate amounts of time for pupils to read on a daily basis, usually on the school premises, that often requires schools to change timetables (unless they already use dedicated, timetabled reading time).

The number of times AR is delivered during the treatment period is driven both by pupil willingness and teacher effectiveness/willingness (Sutherland et al., 2018, Evaluation Protocol). Although as part of the trial schools were required to include 20 to 30 minutes of reading time in their timetables, whether guided or individual reading time, the number of books read and quizzes taken by pupils would vary from individual to individual.

Who provided: Schools receive three one-hour remote training sessions and six hours of onsite training for staff implementing AR. These training sessions help schools ensure that their libraries are prepared for the intervention, show teachers how to use quizzes, and how to analyse the data that they generate.

During training, trainers emphasise the 'non-negotiable' aspects of AR, which consist of (a) reading time (recommended by AR as 20 to 35 minutes scheduled independent reading time per day), (b) resources (regular access to books and to the technology to enable quizzing), and (c) the analysis of data (monitoring of pupil performance, setting targets, and re-testing using the web platform).

Prior to the implementation of AR, RL delivered training for teachers and other school staff involved in the evaluation. These are assumed to be the 'core AR' team and it was a requirement of the training that the headteacher be present, at least briefly, at the training. The roll-out of training was planned to occur before the start of the 2016/2017 school year but in many cases overlapped with the start of the school year. Training was delivered in three stages:

1. Remote telephone sessions (mandatory): initial pre-implementation training with three remote telephone sessions for teachers and other school staff implementing AR for the first time. These sessions, accompanied by slides presented over the internet, were scheduled for individual schools and gave the opportunity for teachers to ask questions specific to their context. The first session focused on preparing the school's book stock for AR. The second was an introduction to the basics of Renaissance STAR Reading (website admin and using STAR Reading). The third session focused on analysing STAR Reading test data. Ideally, all three remote telephone sessions should take place before the onsite training day (see below).
2. Webinar (optional): an optional interactive webinar was also offered to AR users during the implementation stage and focused on AR best practices. RL webinars are not limited to new starters and therefore served as a platform to share experiences.
3. Onsite training day (mandatory): following the remote telephone sessions, all schools had a one-day, onsite training in which RL staff walked participants through the core aspects of AR (for example, 'the non-negotiables'), how to use the different features of the platform, pupil STAR Reading data, and how to use AR with their pupils (quizzes and testing).

This training day was targeted only at the core AR team in the school. As part of the onsite training, headteachers were required to attend the first session of the day or, if they were not able to, arrange to speak to the trainer at a convenient

time. This was considered by Renaissance Learning to be an important element to gain senior buy-in and support in order to ensure that any necessary changes can be implemented (for example, to timetables and other time allocations) to introduce AR.

In order to ensure that the core team was informed about the intervention, and in case of absences from the remote telephone sessions, the onsite training day was in part a recap of the remote telephone sessions. In preparation for these sessions schools were expected to have tested at least one of their classes so that, when discussing STAR Reading output data, they could explore their own data. Furthermore, schools needed to have labelled and input their book lists into their RL site.

Where: Although requiring involvement of librarians or equivalent post holder within schools, AR does not have to be implemented by the whole school and teachers can use it on their own in their classroom, provided that the library and IT systems are in place.

Tailoring: The use of a standardised online platform to assist with initial screening and ongoing testing means that there is a much lower risk of variation in how AR is supposed to work in practice (training or IT problems notwithstanding). Since schools will have pupils with different ranges of ZPD, the level of books read and required in school or classroom libraries will be different but the logic of AR applies regardless, with little school variation.

How well: Evaluations of AR and discussions with the developers suggest a more complex intervention. For example, as set out above, AR requires a number of factors to be in place as prerequisites but also involves a degree of teacher or teaching assistant (TA) involvement/guidance. The degree of teacher involvement is important, particularly in light of the previous EEF evaluation of AR (Gorard et al., 2015), which illustrated a high degree of variability in how the additional reading was actually implemented across schools. Similarly, the use of progress reports and STAR Reading results are intended to be used by teachers regularly to adapt pupil goals and the support that teachers provide students. Taken together, these elements suggest that schools' understanding of how to implement the programme and how effectively schools were able to schedule reading time could all affect treatment outcomes. Fidelity was assessed as part of this trial and will be discussed separately in the Results section.

Modifications to the trial: Although the trial was initially scheduled to take place over one school year (2016/2017), problems with the original outcome measure for Year 4 (New Group Reading Test 2A/2B) at the end of the first year led to the extension of the trial to cover one additional school year (2017/2018). While some treatment schools opted to continue implementing AR in the extended year, others did not, which meant that the intervention length varied across treatment schools (specific details are provided in the Participants section). More specifically, during the second year of the trial, ten of the schools initially randomised to receive the treatment discontinued their use of AR meaning that the 'dosage' of AR in the evaluation is contingent on individual pupil, teacher, and school factors. A summary of the problems encountered with testing at the end of the first year are given next and more detail is given in the Appendix of the updated trial protocol (Sutherland et al., 2018, Evaluation Protocol).

Issues: The outcome measure for Year 4 pupils was initially the paper version of the New Group Reading Test (NGRT) level 2A/2B. Unlike the digital version, the paper version of the NGRT is not adaptable. This was administered at the end of the 2016/2017 school year. Problems with this outcome measure in relation to the distribution of outcomes for all tested Year 4 pupils resulted in the evaluation team disregarding this outcome measure and, following extensive discussions with RL and the EEF, the extension of the trial for an additional school year. This change was reflected in an updated trial protocol and detailed in the Trial design update section. The new outcome measure for Year 4 pupils is the Key Stage 2 reading score, the same measure that was initially, and ultimately, the outcome measure for Year 5 pupils.

Evaluation objectives

The aim of the evaluation was to assess the extent to which AR leads to an improvement in reading ability (see Outcome Measures below) when comparing AR schools to business as usual schools. Consistent with the objectives of the EEF, and distinct from the previous evaluation, this trial has been powered to detect an impact on disadvantaged children—those eligible for free school meals. A full evaluation protocol may be found on the EEF's website:

https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Accelerated_Reader_Protocol_Update_2017.12.18.pdf

The three hypotheses tested in this trial are:

> Hypothesis 1: Accelerated Reader will have a positive effect on the reading comprehension of all eligible pupils (that is, pupils with a sight vocabulary of over 100 words) in the study year groups that are in randomly assigned intervention schools compared to all eligible pupils in control schools.
>
> Hypothesis 2: Accelerated Reader will have a positive effect on the reading comprehension of pupils who are in the study year groups and are eligible for free school meals who are in randomly assigned intervention schools compared to FSM pupils in control schools.
>
> Hypothesis 3: Accelerated Reader will have a positive effect on the reading comprehension of pupils who are in the study year groups and not eligible for free school meals who are in randomly-assigned intervention schools compared to non-FSM pupils in control schools.

Note that the research hypotheses were updated in 2017 at the request of the EEF to reflect a change in their approach to assessing FSM, non-FSM, and all pupils, as per the Statistical Analysis Plan (SAP). The SAP has been published on the EEF's website:

https://educationendowmentfoundation.org.uk/public/files/Projects/Accelerated_Reader_SAP_2018.05.02__FINAL.pdf

Ethics and trial registration

The trial has been registered under International Standard Randomised Controlled Trial Number (ISRCTN): ISRCTN17400885.

The study was reviewed by the RAND Europe ethics advisory board and approved—as was the extension of the study to cover an additional school year—by the University of Cambridge Faculty of Education ethics review process. The ethics and registration processes are in accordance with the ethics policies adopted by RAND Europe.

School recruitment was carried out by NFER. The sample of schools was drawn from NFER's Register of Schools. All eligible schools were invited to take part in the evaluation. The Methods section below provides further detail about school recruitment. The NFER school invitation letter and reply form are included in Appendices C and D respectively.

Data protection

This project was set up in 2016 under the Data Protection Act (DPA). Following the EEF's standard approach at the time, the trial used headteacher/senior leadership consent for participation in the trial, followed by parental opt-out to remove pupils from the research. The parental opt-out form is included in Appendix E: Parental consent form. The intervention was delivered within the school day and the intervention essentially was then standard practice in schools.

In terms of fair processing of personal data, the project fulfilled Condition 1 of processing personal data in Schedule 2 of the DPA as the data subjects gave their implicit consent in the form of an opt-out letter to parents at the beginning of the trial. NFER (contracted by the EEF) provided UPNs to RAND Europe using a secure platform compliant with DPA regulations.

RAND Europe obtained baseline and outcome data from the National Pupil Database. The first data request for baseline measures fell under the DPA. Following the trial design change resulting in a one year extension for the project, the second request (outcome data) fell under the General Data Protection Regulation (GDPR) as implemented by the Data Protection Act (2018). Therefore, NPD outcome data was requested on the legal basis of 'legitimate interest'. RAND Europe and University of Cambridge researchers gained access to this data as part of the project. NPD data was hosted and analysed onsite at RAND Europe, in compliance with NPD data security regulations.

RAND Europe adopts good industry practices regarding the protection of personal data as part of its obligations as a Data Controller under the Data Protection Act 2018 and takes appropriate technical and organisational measures conformant with ISO 27001 to protect personal data. Individuals targeted by the study have the right to oppose, have access to, rectify, or remove personal or sensitive personal data held by RAND Europe and the University of Cambridge.

Project team

The intervention was developed and implemented by Renaissance Learning. As such, RL was responsible for the installation of the AR system in all schools as per the terms of the sign-up information sheet. In addition, RL provided the evaluators with pupil-level data on any AR tests undertaken and on key indicators of implementation and compliance with the intervention.

The delivery team in RL comprised James Bell (project leader), Lauren Shapiro (senior programme manager), and Jessamy Hunter (senior project manager).

The evaluation was conducted by RAND Europe and the University of Cambridge in collaboration with NFER. RAND Europe and the University of Cambridge were responsible for the outcome and process evaluations, trial design, analysis, reporting, and quality assurance of the evaluation. As delivery partner, NFER was responsible for recruitment, outcome testing, and the distribution of process evaluation surveys.

The team in NFER comprised David Hereward (research manager) and Kathryn Hurd and Kinnery Koria (survey operations).

The evaluation team in RAND Europe comprised Dr Alex Sutherland (project leader), Miriam Broeks (project management, field work, and analysis), and Dr. Megan Sim, Dr. Julie Bélanger, and Dr. Joachim Krapels (formerly RAND Europe). The evaluation team in the University of Cambridge comprised: Dr Sonia Ilie (field work and analysis) and Professor Anna Vignoles (project advisor). The evaluation team also benefitted from advice and quality assurance from Dr. Jon Schweig (RAND Corporation) and Dr. Yulia Shenderovich (former RAND Europe).

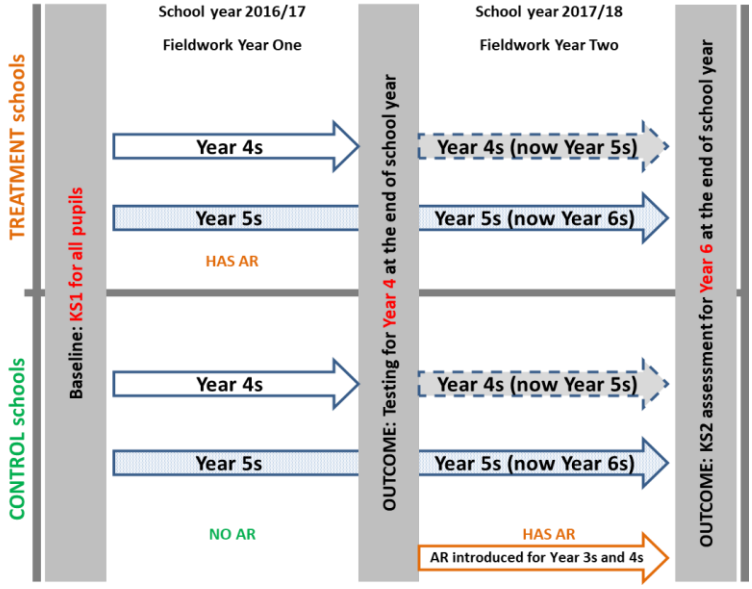
Methods

Trial design

The trial was a two-arm, waitlist, cluster-randomised controlled trial with schools as the unit of randomisation and pupils as the unit of outcome analysis. The aim of the trial was to assess the extent to which Accelerated Reader leads to an improvement in reading ability (see Outcome Measures below) greater than that observed in business as usual schools. Given the design of the trial, with two cohorts included in the study, this means, in effect, we conducted two trials at once—one for Year 4 pupils, the other for Year 5 pupils, as detailed below.

The waitlist element relates to schools being randomised to receive AR for the first time in the 2016/2017 school year or the 2017/2018 school year. For treatment schools, Year 4 and Year 5 pupils received AR in the school year 2016/2017. The initial trial design is illustrated in Figure 2.

Figure 2: Original trial design summary



Problems with the original outcome measure for Year 4 meant that the trial was extended for one year; this meant that treatment schools also received AR in the 2017/2018 school year, when treatment pupils were in Year 5 and Year 6. We know whether treatment schools continued to use AR in the second year of the trial and thus which pupils received extended exposure to AR. Through the process evaluation we attempted to capture what reading-related systems were in place in treatment schools that did not continue using AR.

For control schools, it was business as usual during the school year 2016/2017. During 2017/2018, Year 3 and Year 4 pupils in control schools received AR. This is akin to a cross-over design as control schools received the treatment. However, these cohorts were not directly assessed as part of this study. By monitoring pupils accessing AR in control schools in 2016/2017, 2017/2018, and 2018/2019, we were able to assess the extent of contamination in control schools in the relevant comparator year groups. Contamination was minimal.

Cluster randomisation was the preferred option (whereby the randomisation is done at the school level) to avoid the potential contamination effect of having intervention and control classes within the same school, given that some of the changes required to schools (such as the labelling of library books in accordance to AR level) could potentially spill over into the control condition.

Trial design update

As noted above, problems related to the younger cohort outcome measure (Year 4) resulted in the trial design being updated. More specifically, the outcome measure for Year 4s was initially the paper version of the New Group Reading Test (NGRT) level 2A/2B, administered at the end of the 2016/2017 school year. Problems with the original outcome measure for Year 4 led the evaluation team to disregard this outcome measure following extensive discussions with RL and the EEF. The Outcomes Measures section provides full details of the original NGRT measure. Under the updated trial design (Figure 3), outcomes for Year 4 pupils will be assessed after the third year of the trial using Key Stage 2 assessments. Year 5 treatment effects were assessed after two years using KS2 assessments at the end of the second year of the trial. Table 1 provides all relevant trial details.

Figure 3: Updated trial design summary

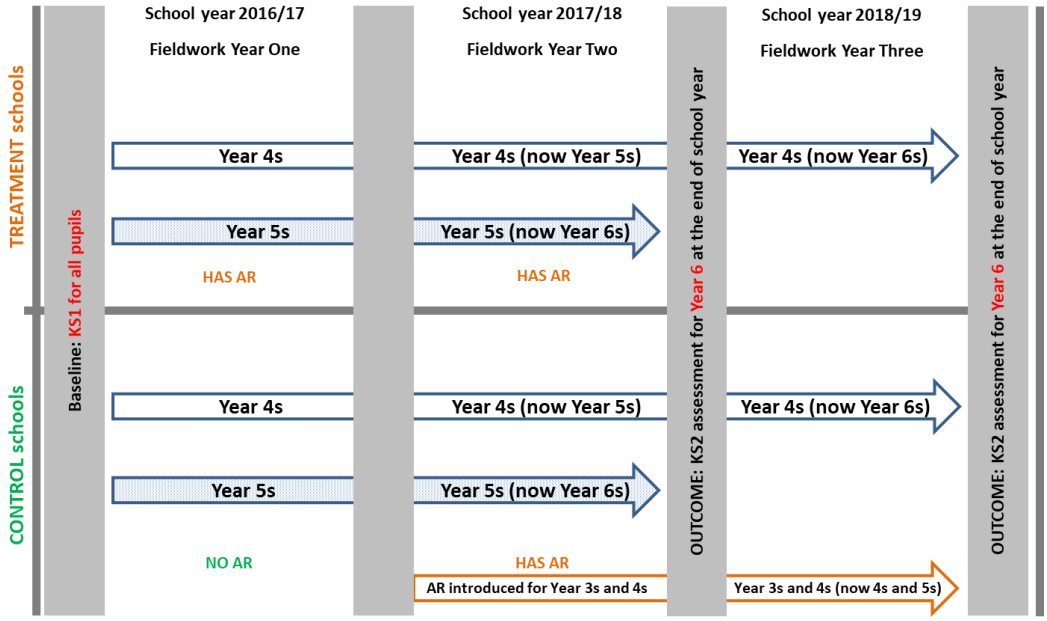


Table 3: Trial design details

Trial type and number of arms		Waitlist cluster-randomised two-arm trial.
Unit of randomisation		School.
Stratification variable(s) (if applicable)		Geography: region (North East, East and West Midlands, and North West/Merseyside and South East). School-level attainment: KS1 average points score (low/high).
Primary outcome	variable	Year 5 pupils: KS2 reading score (KS2_READMRKN), collected at the end of the 2017/2018 school year. Year 4 pupils: KS2 reading score, collected at the end of the 2018/2019 school year. This analysis will be reported separately in 2021.
	measure (instrument, scale)	KS2 reading test, as administered in line with national guidance.
Secondary outcome(s)	variable(s)	None.
	measure(s) (instrument, scale)	Not applicable.

Participant selection

Schools

The target number of schools required to take part in the evaluation was 200 in line with the sample size calculations (see below). NFER estimated that approximately 10% of the sample of schools invited might volunteer to take part in the evaluation and so requested that the size of the sample for invitation contain around 2,000 schools.

The sample of schools was drawn from NFER's register of schools comprising of the following school types: primaries (including 'middle deemed primary', infant and junior, first and middle, and junior), maintained schools, and academies. As this evaluation was funded as part of the EEF's North East Primary Literacy Campaign, the EEF requirement for this trial was that approximately half of intervention schools were located in the North East while ensuring that the sample of schools had an above average intake of pupils ever eligible for free school meals ('Ever FSM') which was set at 29%. Special schools, PRUs, and independent schools were excluded from the sample. The schools selected had pupils in all four of the following year groups: Year 3, Year 4, Year 5, and Year 6. The schools selected were from the local authorities (LAs) listed in Table 5.

Table 4: Geographical areas included in selection

LA name	LA number	Regions
Birmingham City Council	330	
City of Newcastle upon Tyne	391	North East Area
Darlington Borough Council	841	North East Area
Derbyshire County Council	830	
Durham County Council	840	North East Area
Gateshead Council	390	North East Area
Halton Borough Council	876	
Hampshire County Council	850	
Hartlepool Borough Council	805	North East Area
Knowsley Metropolitan Borough	340	
Liverpool City Council	341	
Middlesbrough Council	806	North East Area
North Tyneside Council	392	North East Area
Northumberland County Council	929	North East Area
Nottinghamshire County Council	891	
Redcar and Cleveland Borough Council	807	North East Area
Sefton Council	343	
South Tyneside MBC	393	North East Area
St Helens Council	342	
Stockton Borough Council	808	North East Area
Sunderland City Council	394	North East Area
West Sussex County Council	938	
Wirral Council	344	

The resulting sample comprised 2,401 schools. A total of 206 schools were removed from the sample due to their being previous or current AR customers leaving a remaining sample of 2,195 schools eligible for invitation to take part in the evaluation. Due to there being a finite number of (2,195) schools available for invitation it was necessary to contact all of the schools in the sample rather than adopt a staggered approach whereby schools in the sample with a higher percentage of Ever FSM would be contacted before schools with a lower percentage of Ever FSM. In the event that an excess of schools were recruited over and above the 200 required, it was proposed that the 200 schools that had the highest percentage of Ever FSM would be selected and put forward for randomisation.

Pupils

All Year 4 and Year 5 pupils in eligible schools were eligible for inclusion in the trial, with the following exceptions:

- Pupils unable to take the STAR Reading test were excluded. Schools using AR use the AR STAR Reading test to determine pupils' ZPD (measuring their reading range). The STAR Reading test requires a sight vocabulary of 100 words so pupils without this were unable to take STAR Reading and were excluded from the intervention.

- Pupils unable to read independently at the lowest level of the ZPD chart included in STAR Reading were excluded from the intervention, for example, a Year 4 pupil able to only read Year 1 books independently. Based on RL's experience, it was believed that such pupils would typically be unable to keep up with the requirements of AR and so the intervention is not suitable for them. However, these pupils were still able to use the AR system as to exclude them from the intervention would be unethical.

There is also the issue of student mobility once the trial has started (Vuchinich et al., 2012; Scochet, 2013). Our approach adhered to an intention-to-treat analysis in the event of students migrating between treatment and control schools. Pupils joining schools after the new school year had begun were excluded from the evaluation but were able to access AR.

Pupils or their parents were given the opportunity to opt out from the evaluation. There were two opt-outs by parents, both in the Year 4 cohort, whose results will be included in the Addendum report. The parental consent form is included in Appendix E: Parental consent form.

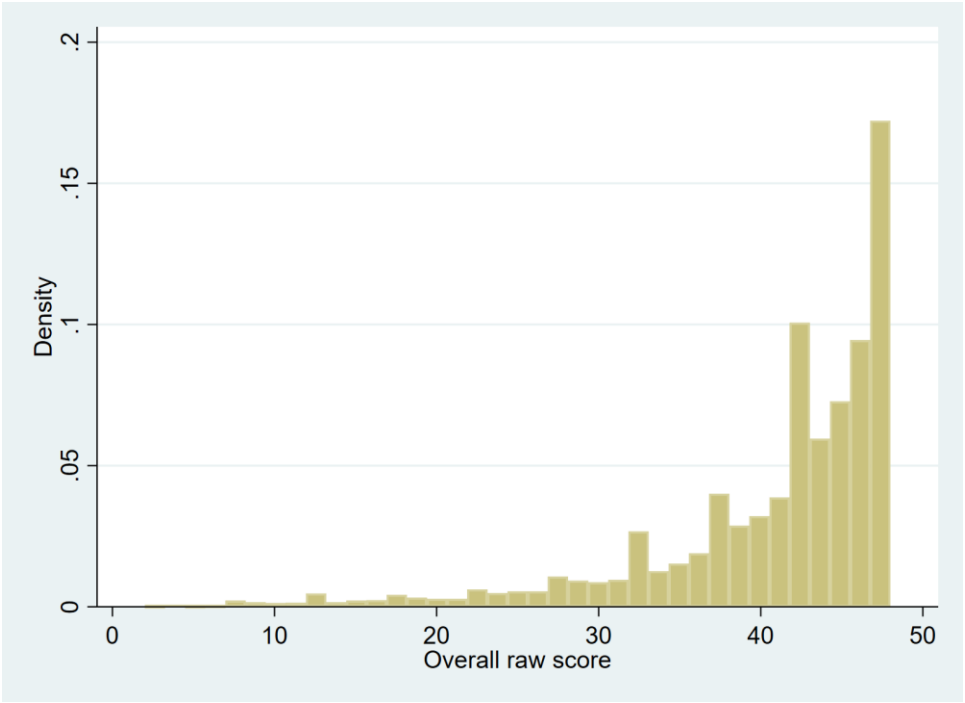
Outcome measures

The primary outcome measure for both Year 4 and Year 5 pupils is their Key Stage 2 fine grade score for reading (the KS2READMRKN variable in the National Pupils Database extract). This is collected when pupils are in Year 6, in this study at the end of the 2018/2019 and 2017/2018 school years, respectively, for Year 4 and Year 5 participants. The fine grade scores were requested from the National Pupil Database (NPD) for the Year 5 participants and form the basis of the analysis below, and were requested for Year 4 participants later in 2019.

For Year 4 pupils, the primary outcome measure was initially intended to be a measure of reading comprehension collected via GL Assessment's New Group Reading Test II (NGRT-II) level 2A/2B (paper version) at the end of the first year of the trial, 2016/2017. This test is available for Year 2 to Year 4 pupils (ages 7 to 9), and lasts between 45 and 50 minutes and was deemed appropriate for use in this trial having consulted with GL Assessment. The NGRT-II provides standard overall scores that can be broken down into separate scale scores for sentence completion and context comprehension (GL Assessment, 2015). It also delivers age equivalent scores and National Curriculum levels for reading (NGRT, 2015). The trial analysis would have used the most fine-grained measure available, namely the raw score on the NGRT Test 2A/2B.

Testing of Year 4 pupils with NGRT-II was undertaken in the summer of 2017. To minimise assessor bias in outcome measures, the evaluation team only shared information about allocation with senior researchers from NFER who did not inform NFER field staff overseeing test administration of school allocation. Assessor bias was also minimised through the use of standardised and national testing of outcome measures. The test was invigilated by independent invigilators from NFER and there were no reported problems with test administration. We originally intended to follow up the Year 4 pupils to assess the impact of AR over time, using Key Stage 2 data. However, the distribution of the NGRT-II test data achieved was strongly suggestive of a problem with the test. Figure 4 shows the distribution of the test results for approximately 5,500 Year 4 pupils. The test was scored between 0 and 48. Ordinarily, one would expect to find a measure of educational ability or aptitude to be approximately normally distributed, that is, with the majority of observations clustered around a mid-point and following a bell-curve. We observe that the majority of observations are distributed to the right, meaning that the majority of pupils scored highly on this test (around 30% of pupils scored 47/48 or 48/48). Having investigated that the test that was administered correctly, and the scoring of the test was accurately undertaken, we concluded that the test was too easy for the pupils in this specific trial, resulting in ceiling effects such that the test was not suitable for assessing the effectiveness of AR on pupil outcomes.

Figure 4: NGRT-II test data for approximately 5,500 Year 4 pupils



Trial update: new outcome measure for Year 4

Given the above issue related to the original Year 4 outcome measure and following extensive discussions with RL and the EEF, the EEF's grants committee decided to continue funding AR for the target cohorts of pupils for another academic year (2017/18). The primary outcome for Year 4 pupils was amended to be Key Stage 2 fine grade score for reading (KS2READMRKN) in the 2018/2019 academic year. Further information on the issues with the NGRT-II outcome testing is available in the Appendix of the updated trial protocol (Sutherland et al., 2018).

For Year 5 pupils, the primary outcome measure is the Key Stage 2 fine grade score for reading (KS2READMRKN) when the pupils are in Year 6 (2017/2018 academic year). Initially, the trial was intended to test the effectiveness of AR after one year of intervention (when pupils were in Year 5). The extension of the trial means that the analysis using the KS2READMRKN variable above now measures the effectiveness of AR after up to *two years* of intervention. Renaissance Learning data on school uptake of AR tracks whether treatment schools continued to use AR in the second year of the trial, and thus which pupils received extended exposure to AR.

Secondary outcomes

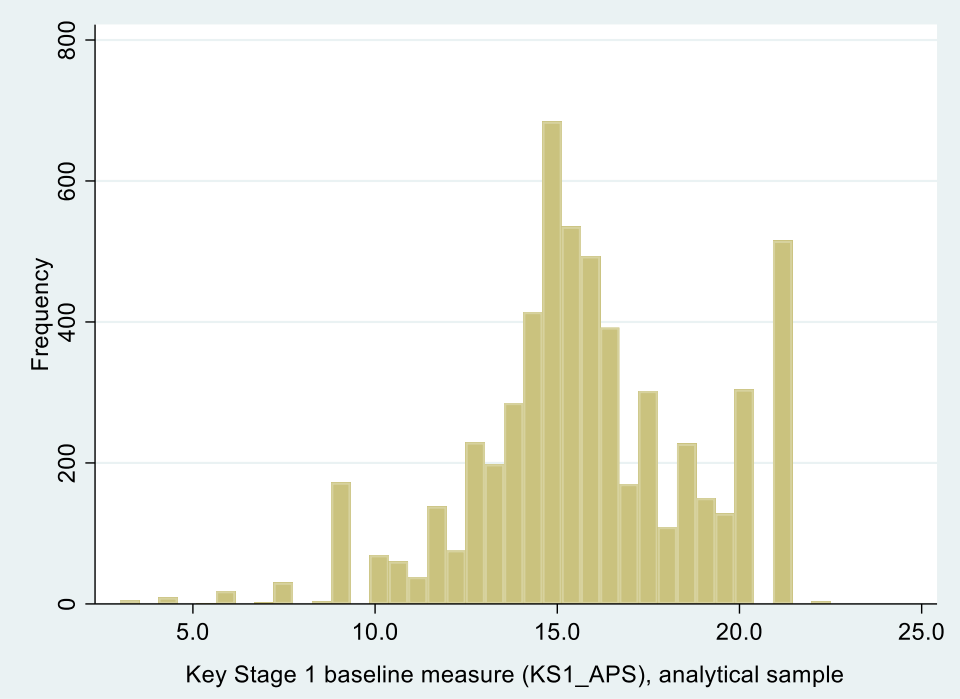
There were no secondary outcomes in this trial.

Baseline measure

The baseline measure was Key Stage 1 attainment, measured through the KS1 average points score variable also obtained from the National Pupil Database for pupils in the trial. KS1 attainment is a valid predictor of later attainment in school. The raw correlation between KS1 and KS2 attainment ranged between 0.69 and 0.73 in a sample of large EEF trials analysed for the EEF by Allen and colleagues (2018), and indeed it was used as a benchmark to compare the performance of other commercially-available attainment measures for the same year group in that study.

The manner of recording KS1 information in the NPD has recently changed, with the granularity of the measure decreasing. Even so, KS1 attainment still provides a good amount of information for use in trials. Given its timing, this AR trial was able to use the previous version of the measure. The distribution of the measure is not perfectly normal (Figure **5** below), but sufficiently so as to not raise either substantive or methodological concerns.

Figure 5: Key Stage 1 baseline measure, analytical sample



Sample size

At protocol stage, we assumed that the trial would consist of 200 schools, 100 in each treatment arm. This was the agreed upper limit placed on the trial in terms of both funding and the capacity of the developer to install and implement AR in time for a September 2016 start. We assumed that there would be on average 33 pupils per class and average Ever FSM of 28% to 29% as reported in DfE statistics (2015). With the assumed (average of) 33 pupils per year and 200 schools, the overall estimated sample for Year 4 was 6,600 pupils and 6,600 pupils for Year 5, including approximately 2,000 Ever FSM pupils in each Year. Based on the intra-cluster correlation coefficient (ICC) analysis using 2014/2015 NPD data provided by the EEF at the time, we assumed an ICC of 0.13. Finally, for comparability across EEF trials, only pupil-level covariates were used in analyses; specifically, based on KS1-KS2 correlation of r = 0.73 (Treadaway, 2013), we assumed level 1 variance explained of 0.73^2 = 0.53. We also assumed an alpha of 5% and an intended 80% power to detect effects. Power and minimum detectable effect size (MDES) calculations were performed using the PowerUp tool (Dong and Maynard, 2013). Calculations were performed both for the overall sample and for the sample of pupils eligible for FSM. Further updates to MDES at randomisation and at analysis are presented in Table **7**. At design stage, the MDES for the overall primary outcome analysis was 0.15; at randomisation, the overall MDES was 0.16; for FSM-eligible pupils, the MDES at design stage was 0.17' at randomisation the FSM-group MDES was 0.18. At analysis for the Year 5 cohort only, the MDES was 0.137 for the overall analysis and 0.137 for the FSM-

only analysis. The MDES was smaller than originally planned because the observed ICCs were smaller than originally expected (the pre-post correlations were also smaller, but not sufficiently so to drive the MDES up).

At randomisation stage, NFER recruited 181 schools to participate in the trial, encompassing more than 12,000 pupils. These schools were randomised (using the procedure outlined below) to either the control or treatment conditions. Following random allocation, there were 89 schools in the treatment group and 92 schools in the control group. At the time of enrolment in the trial, the 89 treatment schools included more than 6,000 pupils (3,188 pupils in Year 4 and 3,088 pupils in Year 5). The 92 control schools included 3,123 pupils in Year 4 and 3,028 pupils in Year 5. Additionally, three pupils in the control schools were identified as Year 2 and therefore not included in analyses.

Randomisation

A stratified cluster-randomised waitlist design was planned and undertaken with schools as the unit of analysis. Allocation to treatment or control was conducted on 5 May 2016 by Dr Sutherland with advice from Prof Anna Vignoles. We ruled out simple randomisation because of the risk of imbalance on key factors such as geographical location. NFER recruited 181 schools for randomisation, from an initial target of 200 schools. These 181 schools were allocated to the intervention and control groups on a 1:1 ratio through stratified randomisation.

Strata were created by creating groups combining geographical region (based on government regions) and a dichotomised measure of school-level prior KS1 attainment. KS1 prior attainment data was taken from the DfE website and used to mean-split the data, with the threshold being a KS1 average points score (KS1aps) of 15.2. Schools were categorised as 'low' for prior attainment if they had below 15.2 KS1aps (the sample mean was 15.098), and 'high' if they were equal to or greater than 15.2 KS1aps. Two schools did not have KS1aps data but were still included in the randomisation as a separate stratum. This was retained in the later analysis. Regions were grouped together as follows:

- North East;
- East Midlands and West Midlands ('Midlands'); and
- NW/Merseyside and South East.

Within each region-KS1aps stratum, simple randomisation was conducted using a random number generator (in Stata). Each school within a stratum was assigned a random number; schools were then sorted by the random number and the first half in each strata were allocated to treatment. Random allocation resulted in 89 schools in the treatment group and 92 schools in the control group. The evaluation team informed NFER and Renaissance Learning of allocations on 5 May 2016. As discussed below, trial analysis will incorporate stratum variables.[6] Table **5** below shows the allocations by strata. Later analyses were not undertaken blind to randomisation.

Table 5: Allocations by strata

Region	KS1 high or low	Control	Treatment	Total
NE	Low	17	16	33
NE	High	15	15	30
Midlands	Low	12	11	23
Midlands	High	17	17	34
NW/M/SE	Low	16	15	31
NW/M/SE	High	14	14	28
Missing KS1	Missing KS1	1	1	2
Total		92	89	181

Statistical analysis

The analysis of primary outcomes (as detailed above) for Year 5 pupils (relevant to this report) is based on an intention-to-treat (ITT) framework. This means that once randomised, schools and participants are analysed according to their allocation regardless of whether they received or participated in the intervention or not.

The ITT approach is particularly relevant for future policy-making stakeholders and practitioners who may roll out or implement a particular programme without much control as to how that programme is actually taken up in the system. Therefore, the ITT approach allows for estimating the effects of *offering* that particular programme. Furthermore, the ITT approach is inherently conservative as it captures the averaged effect of offering the intervention for those who do

[6] Strata help to reduce differences between treatment and control groups so omitting them increases between-group differences at baseline and a subsequent loss of power. See Kahan, B. C. and Morris, T. P. (2012) 'Reporting and Analysis of Trials Using Stratified Randomisation in Leading Medical Journals: Review and Reanalysis', *BMJ*, 345. http://www.bmj.com/content/bmj/345/bmj.e5840.full.pdf

not comply and the effect for those who do comply. To ensure that we capture the effects of the intervention in relation to the appropriate implementation of AR, as per initial specifications from Renaissance Learning, a fidelity analysis has also been undertaken. This is detailed in the Fidelity section.

To answer the research questions around the effects of AR on the reading outcome, the treatment and control arms were compared in terms of the difference in means between groups at follow-up, conditional on baseline measures (KS1) and all stratification variables. Since randomisation was stratified and undertaken at the school level, this requires that we explicitly account for the clustering of pupils in schools. Therefore, the analyses that have been conducted take the form of a multilevel model. The general equation for the multilevel model is

$$yij=\propto+ X_{ij}\beta+Zj+\delta ARj+uij +uj\ i=1..,=1..M, \text{ (1)}$$

where y_{ij} denotes the pupil-level outcome, i and j denote pupil and school indexes respectively, and X_{ij} is the $1\times k$ vector of individual characteristics (here, prior attainment, KS1APS). AR_j is a dummy variable denoting treatment/control group at school level, β and δ are the $k\times 1$ and 1×1 vectors of regression coefficients, Z_j is a vector of school-level characteristics formed from the stratification variables mentioned previously (that is, region and low/high average KS1 scores), u_{ij} is the pupil-level error term, and u_j is the school-level error term. In Equation 1, α denotes the overall model intercept.

The coefficient δ associated with the AR_j dummy constitutes the main result of the trial. 'All pupils' represents the main analysis group (with the FSM-eligible and non-FSM eligible samples representing subgroups) so to obtain the main result of the trial, Equation 1 above will be estimated for all pupils.

Effect sizes were estimated using the effect sizes for cluster-randomised trials given in the EEF evaluator guidance. Equation 2 provides one such example, for when the population variance is available:

$$ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sigma^2} \text{ (2)}$$

where ($\bar{Y}_T - \bar{Y}_C$) is the mean difference between intervention groups adjusted for baseline characteristics and σ^2 is the population standard deviation (variance). The ES therefore represents the proportion of the population standard deviation attributable to the intervention (Hutchison and Styles, 2010). A 95% confidence interval for the ES, that takes into account the clustering of pupils in schools, will also be reported and will be calculated by dividing the upper and lower confidence interval bounds by the population standard deviation.

The population variance is not, however, available for this analysis and therefore we have substituted the sample-based, pooled unconditional variance (s*) in the denominator, which is an estimate of the population variance. In multilevel models this variance will be the weighted average of the variance of treatment and control groups, as per EEF guidance (EEF, 2018 p. 4 n. 11). Equation 3 below provides the formula used for calculating this effect size in this report. Respective confidence intervals were calculated by dividing the confidence interval of the treatment effect coefficient (δ in Equation 1) by the pooled standard deviation.

$$s^* = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}} \text{ (3)}$$

Primary analysis model

Given the clustered nature of the data (pupils clustered in schools), we used multilevel models to analyse the results from the trial. The equation is:

$$yij=\propto+ X_{ij}\beta+Zj+\delta ARj+uij +uj\ i=1..,=1..M, \text{ (4)}$$

where y_{ij} denotes the pupil-level outcome, i and j denote pupil and school indexes respectively, and X_{ij} is the $1\times k$ vector of individual characteristics (here, prior attainment, KS1APS). AR_j is a dummy variable denoting treatment/control group at school level, β and δ are the $k\times 1$ and 1×1 vectors of regression coefficients, Z_j is a vector of school-level characteristics formed from the stratification variables mentioned previously (region and low/high average KS1 scores), u_{ij} is the pupil-level error term, and u_j is the school-level error term. In Equation 1, α denotes the overall model intercept.

The vector of individual characteristics only includes prior attainment, KS1APS. For the 5,759 pupils in the analytical sample, this variable has a mean of 15.86 (CI: 15.8, 15.9) and a standard deviation of 3.07. The vector of school characteristics includes the variables used for randomisation: region and low/high average KS1 scores. Region 1 (North East) represents the reference category with Region 2 (Midlands) and Region 3 (Northwest/Merseyside and South East) entered as school-level dummies in the model. A further categorical variable (AVG_KS1) at the school level is also

included to account for the average KS1 score grouping at randomisation with the three categories used at randomisation (high/low/missing). Two schools had missing information on this variable at randomisation and were allocated using a simple random draw. Therefore, the AVG_KS1 variable is coded at 0 if a respective school was in the low average KS1 score group, 1 if the respective school was in the high average KS1 score group at randomisation, and 2 if the respective school had missing data on this variable at randomisation.

No other control variables were included in the final model.

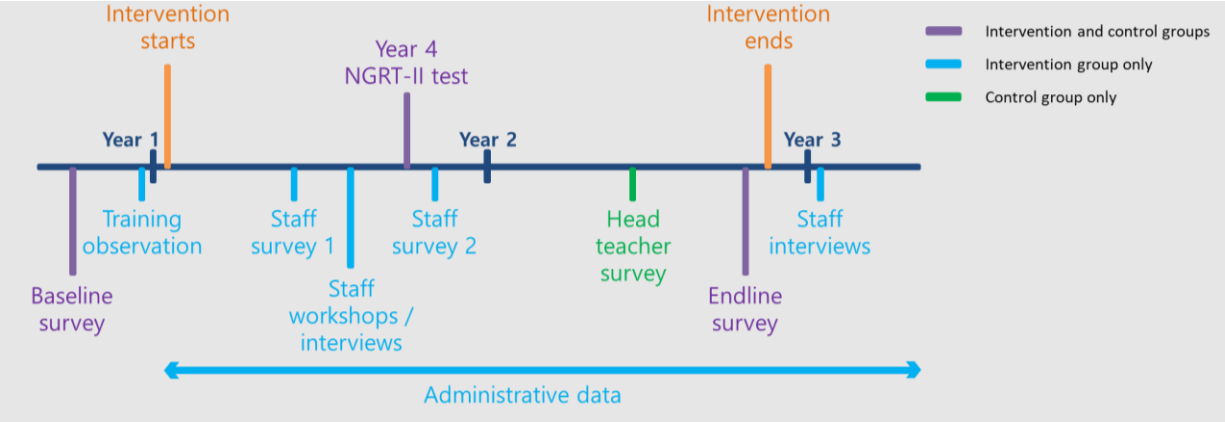
Implementation and process evaluation methods

Following Oakley et al. (2006) and informed by both implementation science, realist approaches to evaluation, and the EEF's own Implementation and Process Evaluation Guidance (Humphrey et al., 2015), we conducted a process evaluation to better understand the complexities of implementing AR. Implementation science emphasises thinking about 'dosage' or 'exposure', fidelity or adherence, and implementation quality. The realist approach emphasises thinking about what works, for whom, and under what conditions. The purpose of the process evaluation was to address the following questions:

- Was the intervention implemented with fidelity in the intervention schools?
- What factors and initial conditions appear to explain variation in fidelity of implementation?
- What appear to be the necessary conditions for success of the intervention?
- What were the barriers to delivery?

A series of data collection approaches were used to answer these process evaluation research questions, as follows (Figure **6**).

Figure 6: Implementation and process evaluation activities



In what follows we provide a brief overview of the different data collection methods used and the rationale behind them. Further details are provided in the Implementation and process evaluation section (page 38). Semi-structured protocols were used for all observations, interviews, and workshops to ensure that researchers systematically captured information (for more information on aspects captured in the protocols please see page 38).

The *observations of a sample of the training sessions* for staff in intervention schools followed the standard training pattern for AR and included participation in three remote telephone and web-based sessions. Seven remote telephone training sessions, four webinars, and eight onsite training sessions in schools were part of the observations in May and June 2016 and June to September 2016 respectively.

The *school survey* included a baseline survey and an endline survey. The baseline survey collected information about the use of literacy-focused programmes and the status of school libraries prior to the implementation of the programme. All schools completed this at the point of *recruitment* in February and March 2016 (treatment: n = 89; control: n = 92; an additional 43 other schools also completed the survey but were later excluded to prioritise those with higher FSM proportions). The endline survey collected information about how practice had changed in both treatment and control schools. The endline survey was administered to teachers, librarians, and senior leaders in treatment schools (n = 87; please see

Appendix F: Breakdown of survey responses by role) and headteachers in control schools (n = 42) in June and July 2018. Both of these surveys were administered by NFER.

Further to the school surveys, *staff surveys* were also administered in treatment schools only: intervention staff surveys took place in February 2017 (during the first trial year) and *again* during June and July 2017 to elicit information regarding the implementation of the intervention. These surveys were administered by NFER in February 2017 (n=147) and in June and July 2017 (n=169) and included several types of stakeholders in each school: teachers, librarians, and senior leaders (please see

Appendix F: Breakdown of survey responses by role). The first staff survey gathered information on perceptions regarding the effectiveness of the training and resource materials, perceptions of the implementation of the intervention, and explored any barriers or challenges faced by the staff in the implementation. The second staff survey, at the end of the first year, gathered information regarding the implementation of the intervention and any barriers or challenges faced by the staff.

A further *control headteacher midline survey* was administered in control schools only in February 2018 exploring other types of literacy and reading programmes being used in these schools. Responses were collected from headteachers (n = 37). This survey gathered information on the kinds of literacy programmes that were in place in the control group schools and how—and to what degree—these differed from (or are similar to) the AR programme in practice.

The *staff workshops* undertaken by the evaluation team with staff in a subset of treatment schools explored key findings from the staff surveys by facilitating in-depth discussion around these topics. The workshops took place in Liverpool and Newcastle locations in April 2017.

A second round of staff workshops was originally intended to be carried out at the end of the trial, however, high attrition from the first set of workshops led to the decision to make these phone *staff interviews* instead. These took place with two teachers in May and June 2017 and with one teacher in July 2018.

Finally, *administrative data*[7] was supplied by RL in relation to the uptake and usage of AR in treatment schools and to any potential contamination in control schools. This was provided from the beginning of the evaluation up to June 2019. This data included pupil-level information on the quizzes they had sat, the average percentage of correct responses achieved across sat quizzes each respective implementation term, as well as information from STAR Reading regarding their ZPD (updated when a new test was taken) and whether the books pupils were reading matched their ZPD range. We refer to these aspects as 'pupil-level AR benchmarks'. The AR platform required (and the delivery team supplied) a unique identifier for each pupil so that their information could be tracked over time. Some teachers provided UPNs as part of this, others variations on pupils' (given) names, and others used consecutive numbers (for example, 1, 2, 3, …) to identify the pupils. All these are sufficient for the purposes of tracking but do not allow for one-to-one matching with pupils for the purposes of impact analysis. Therefore, in the IPE section this data is used to assess whether there was any contamination in control schools while in the Fidelity sub-section (in Impact evaluation) it is used to assess fidelity to the intervention at the school level.

Implementation and process evaluation analysis approach

As part of the analysis of implementation and process evaluation (IPE) data, we initially analysed each data source separately (surveys, training observations, workshops, and interviews) and then compared them to draw out emerging themes. Analyses of survey data focused on looking for any emerging patterns in responses while bearing in mind methodological limitations (such as the inability to track an individual's responses across surveys). Surveys provided the most comprehensive data source as they covered the largest number of views across different participant groups

[7] Following the extension of the trial it was decided to continue collecting administrative data on the use of AR throughout Year 3. This data was provided by RL to the evaluation team.

(teachers, TAs, librarians and library coordinators, literacy coordinators, and headteachers). Surveys also covered questions on the training, early set-up, and implementation of the intervention, as well as questions around daily use (enablers and barriers) and costs. Analyses of survey data informed the focus of the workshops and interviews since these provided the opportunity to delve deeper into aspects that were salient from survey responses. In this way, the different data sources (including administrative data provided by Renaissance Learning) complemented each other. By triangulating different data sources, the evaluation team was able to obtain a more comprehensive picture of the implementation and use of the intervention in practice.

Costs

We estimated average marginal costs per pupil per year for schools for the intervention in two stages: an initial discussion with the developer and then asking headteachers as well as librarians and literacy coordinators to populate a template, outlining key cost categories in both start-up and operation of the AR intervention. Average marginal costs per pupil per year over a three-year period were estimated in line with the convention followed by other Education Endowment Foundation projects (EEF, 2016). The EEF cost rating is included in Appendix A: EEF cost rating. Separate estimates of prerequisite costs and of additional staff time were requested from headteachers and other school staff (such as teachers and TAs). These estimates together were intended to provide a general indication of the costs involved with implementing AR, though we recognise that costs may differ between schools; in particular, prerequisite costs will vary substantially depending on the existing library resources and books available.

We collected data on the following specific cost categories:

- direct financial costs, including direct costs paid to AR if the intervention were to be offered without subsidy, salary costs for additional staff required, direct costs for books and other materials, and so forth;
- direct time costs, including staff and volunteer time; we also collected data on average supply cover time (rather than cost) required for the intervention; and
- we also collected additional data from the delivery organisation to validate the above estimates and to ascertain whether any changes to costs are anticipated in the future if the intervention were to be made more widely available.

Timeline

Table **6** below presents a timeline of the main activities related to the evaluation and intervention delivery.

Table 6: Timeline

Date	Activity
January 2016	Evaluation kick-off meetings
February–March 2016	Recruitment of schools
April–May 2016	Randomisation
May–July 2016	Training provided for intervention schools
September 2016	Intervention begins in Year 4 and Year 5 in intervention schools
February 2017	First intervention staff survey in intervention schools
April 2017	Workshops in selected intervention schools
June 2017	Outcome tests for Year 4 pupils (NGRT-II)
June–July 2017	Second intervention staff survey in intervention schools
Apr–July 2017	Training provided to control schools
Sept 2017	Intervention begins in Year 3 and Year 4 of control schools Intervention continues in (now) Year 5 and Year 6 of intervention schools

February 2018	Survey with headteachers in control schools
June–July 2018	Endline staff survey in intervention and control schools
July 2018	Interviews with staff in intervention schools
May 2018	Originally Year 5 pupils (now Year 6) take KS2 tests
May 2019	Originally Year 4 pupils (now Year 6) take KS2 tests
June 2019	First report: older cohort (Year 5) KS2 outcomes and process evaluation
December 2019	Second report: younger cohort (Year 4) KS2 outcomes update

Impact evaluation

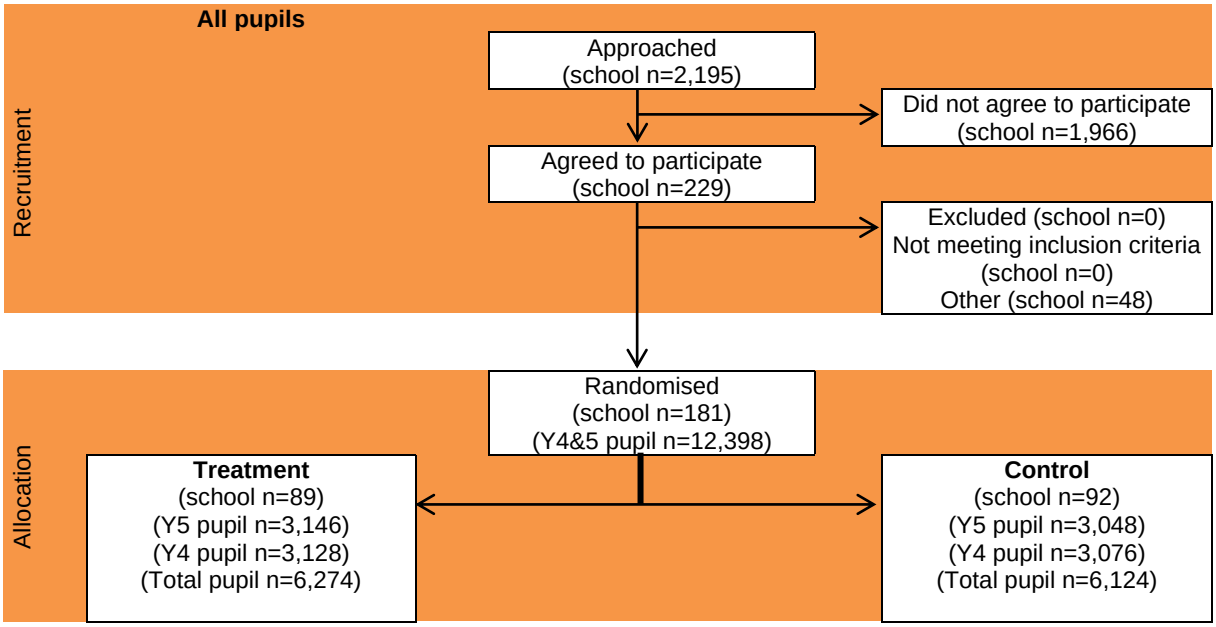
Box 1: Impact evaluation key findings

- We did not find support for the hypothesis as to the causal effect of AR on pupil reading outcomes compared to pupils in business as usual schools: pupils in treatment schools did no better than pupils in control schools.
- The trial was also powered to detect an effect for the FSM subgroup of pupils. We found no support for this hypothesis: FSM-eligible pupils in AR (treatment) schools did as well as FSM-eligible pupils in control schools.
- A continuous implementation fidelity measure was derived from administrative data. Implementation fidelity was high across treatment schools. However, our quantitative assessment of fidelity using this measure also showed no difference between reading outcomes of pupils from treatment and control schools.

Participants

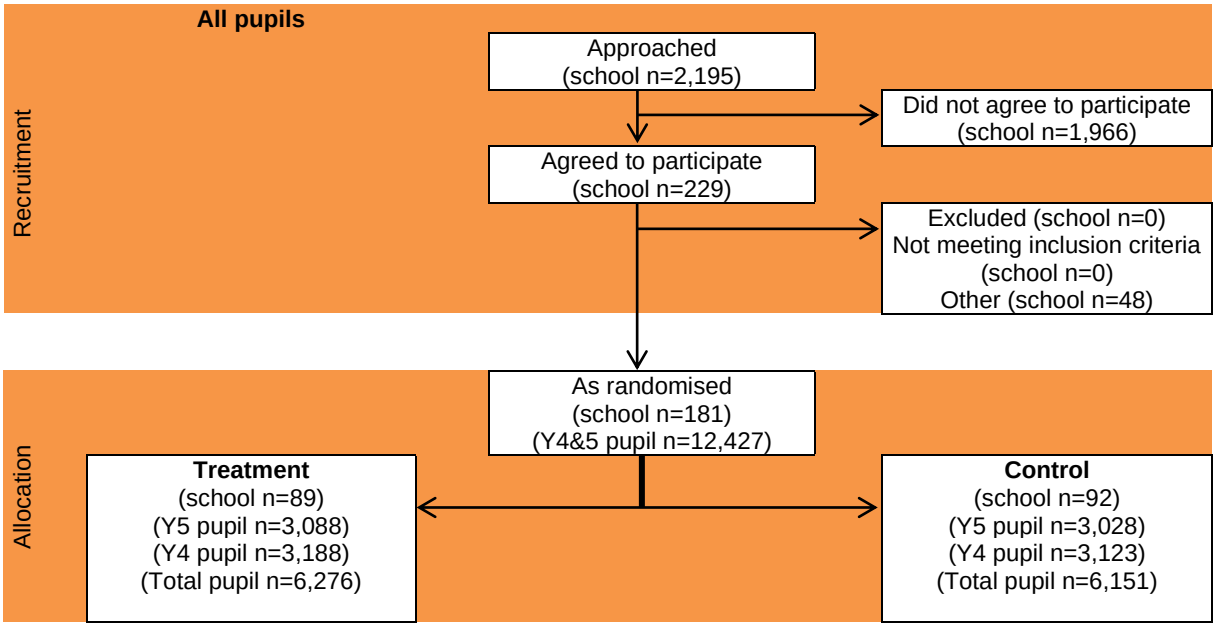
The trial design treats Year 4 and Year 5 pupils as two separate trials, with the overall participant flow diagram up to randomisation presented in Figure 7. The Year 4 outcomes will be reported at a later date in an addendum report so only the Year 5 participant flow diagram is presented in full here, in Figure 9 below.

Figure 7: Participant flow diagram both Year 4 and Year 5, as at randomisation



After randomisation, a year group identification mistake was discovered by NFER whereby some pupils had been misclassified by two control and two intervention schools in the original data return, so that a total of 78 pupils were initially labelled as Year 5 when they were actually in Year 4, and a further 31 Year 4 pupils (29 consenting to take part) were not included in the original data return. As a result, Figure **8** below illustrates the final, definitive numbers for the whole trial.

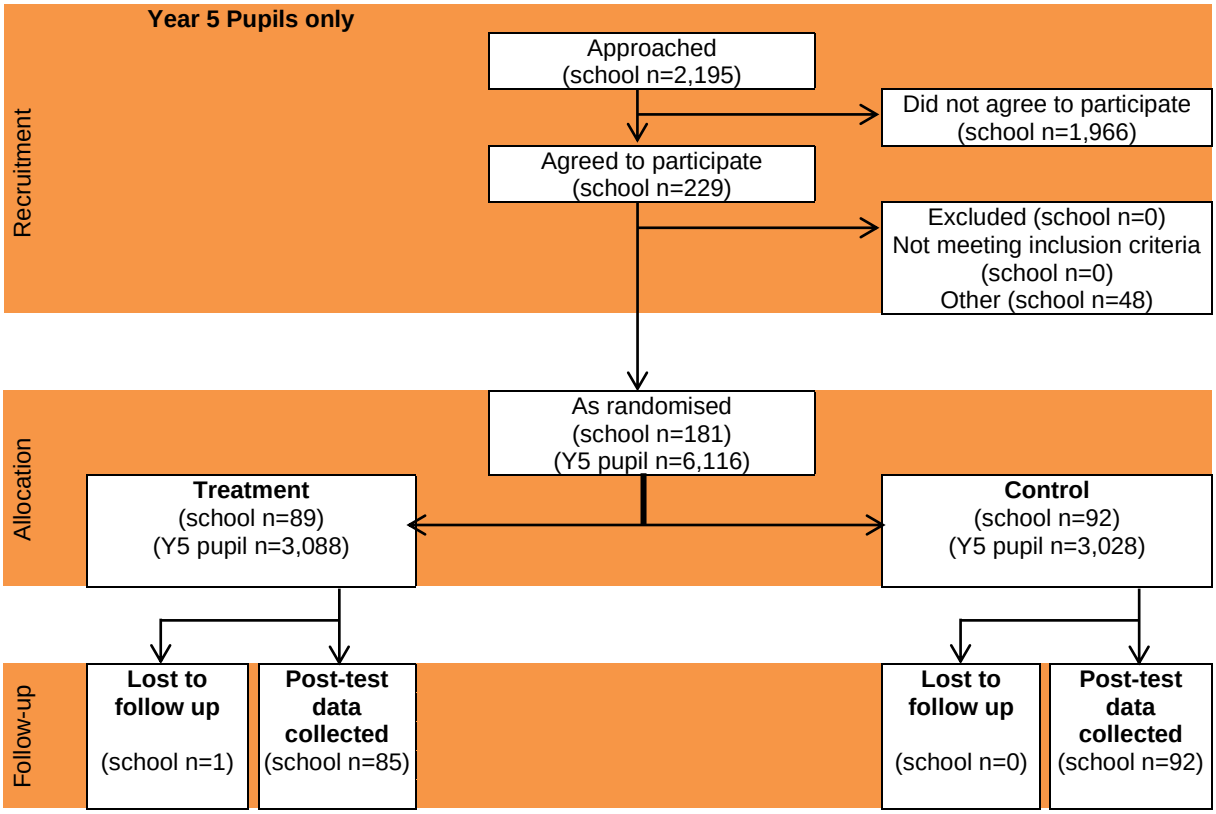
Figure 8: Participant flow diagram both Year 4 and Year 5, after correction



As Figure 9 illustrates, at randomisation stage, NFER had recruited 181 schools to participate in the trial. These schools were randomised (using procedure outlined above) to either the control or treatment groups. Following random allocation, the Year 5 arm of the trial consisted of a total of 6,116 pupils in 89 treatment schools (3,088 pupils) and 92 control schools (3,028 pupils).

After the school-level randomisation, three of the 89 intervention schools were observed to not have any Year 5 pupils. They continue to be part of the trial, with the Year 4 pupils included in the addendum report.

Figure 9: Participant flow diagram for Year 5 pupils



Implementation stoppage and attrition

At the original trial conclusion date (June 2017), six of the intervention schools and one control school had withdrawn from the intervention, but not from the trial as a whole—that is, they had stopped implementing AR but did not withdraw from the trial, therefore, under the ITT approach, were still included in the final analysis and are not considered to have been lost to follow-up. The most-often cited reason for this withdrawal from the intervention was not having the resources or budget to continue implementing it.

Further withdrawals from the intervention only occurred in the second year of the trial, so that by December 2018, the overall school attrition consisted of:

- one intervention school that had withdrawn from the intervention as well as from the trial (addressed below, listed under 'withdrawn from follow-up'); and
- a further seven intervention schools that had withdrawn from the intervention only, but again remained in the trial for the purposes of outcome analysis.

We consider that the latter two categories of school withdrawals represent withdrawals from the intervention but not the trial, and schools are retained in the ITT analysis.

In the Statistical Analysis Plan, we reported that by June 2017 a total of six intervention schools and one control school had withdrawn from the trial. This did not distinguish between stopping to implement the intervention and withdrawal from the trial, which was inaccurate. However, reviewing the information provided by NFER and RL as to the reasons for withdrawal, and considering the difference between stopping to implement the intervention, only one school had actually withdrawn from both the intervention and the trial (the intervention school mentioned above) with the rest of the schools withdrawing from the intervention only (where relevant).

Additionally, in the second year of the trial, in which control schools were in a position to receive AR for younger pupils, 23 control schools decided not to take up the programme. These schools remained in the trial and are therefore included in the flow chart in Figure 8. The

Implementation and process evaluation did not independently collect reasons for the decisions to stop implementing AR in the second year. However, the delivery team were informed of reasons for implementation stoppage by schools: the top-cited reason was a change in school leadership (specifically headteacher); a lack of communication about the trial in the first year was cited as a reason by schools in the control group.

Given the trial extension, treatment schools were offered AR in the second year of the trial as well. Of the 89 schools randomly assigned to receive the treatment, ten chose not to renew their AR license, that is, they did not implement AR in the second year of the trial. They continue to be part of the evaluation as per the ITT approach, but initially-Year 5 pupils in these schools will have received AR for only one of the potential two years.

The timing of the trial extension also meant that some schools continuing to implement AR in the second year did so with a gap in implementation at the beginning of that second year while extension agreements were being finalised. The

Implementation and process evaluation includes further information on this issue, as well as on the proportion of pupils who met the implementation standards during the second year.

At analysis stage, data was available for 2,920 Year 5 pupils in 85 intervention schools and 2,839 Year 5 pupils in 92 control schools. Year 4 data is not yet available and will be reported in the addendum report.

For the Year 5 cohort, the overall ratio between pupils analysed and pupils randomised is 5,759 to 6,116—a retention rate of 94.2% and an attrition rate of 5.8%. In the control group, the retention rate is 93.4% (an attrition rate of 6.7%) with 2,839 pupils in the analysis from the 3,028 randomised. In the intervention group, the retention rate is 94.6% (an attrition rate of 5.4%) with 2,920 pupils analysed of the 3,088 randomised to the intervention condition.

Sample size calculations

At the time of the *analysis* of the Year 5 pupil outcomes, there were 85 schools in the treatment group and 92 schools in the control group. This included 2,920 Year 5 pupils in the treatment schools and 2,839 Year 5 pupils in the control schools. For the purposes of the at-analysis MDES calculations for both the overall sample and FSM group (performed using PowerUp; Dong and Maynard, 2013), the ICC and level 1 explained variance by the pre-test were derived empirically from data held, and are presented in Table **7**. At design stage, the MDES for the overall primary outcome analysis was 0.15; at randomisation the overall MDES was 0.16; at analysis for the Year 5 cohort only, the MDES was 0.137.

Table 7: Sample size and minimum detectable effect sizes for different stages of the trial

		Protocol		Randomisation		Analysis: Y5 only	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.15	0.17	0.16	0.18	0.137	0.137
Pre-test/post-test correlations (KS1_READWRITPOINTS and KS2READMARKN)	level 1 (pupil)	0.73	0.73	0.73	0.73	0.662[e]	0.655[e]
	level 2 (class)	NA	NA	NA	NA	NA	NA
	level 3 (school)	0.00	0.00	0.00	0.00	0.00	0.00
Intracluster correlations (ICCs)	level 2 (class)	NA	NA	NA	NA	NA	NA
	level 3 (school)	0.13	0.13	0.13	0.13	0.089	0.057
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two	Two	Two
Average cluster size		66	20	68.5[a]	22.8[b]	32.5[f]	13[f]
Number of schools	intervention	100	100	89	89	85	79
	control	100	100	92	92	92	88
	total	200	200	181	181	177	167
Number of pupils	intervention	6,600	2,000	6,274	2,032[c]	2,920	1,092
	control	6,600	2,000	6,124	2,101[c]	2,839	1,071
	total	13,200	4,000	12,398	4,133[d]	5,759	2,163

[a] Based on summing the average number of pupils per school for treatment and control arms (70.5 and 66.6 respectively) and dividing by two.
[b] This is estimated by dividing the overall figure by three.
[c] Based on multiplying the number of schools by the average number of pupils per school.
[d] The sum of the two cells above.
[e] Correlation between KS1_READWRITPOINTS and KS2READMARKN.
[f] Average observations per group in analysis.

Subgroup sample size: pupils eligible for free school meals

As Table 7**Error! Reference source not found.** illustrates, separate sample size calculations were carried out for the subgroup in this analysis, namely pupils eligible for free school meals.

The Ever-6 FSM variable was used to identify FSM pupils as follows. The FSM variable was sourced from the 2017 Spring census, the school census closest to randomisation (EVERFSM_6_P_SPR17). FSM data from the 2017 Spring census was available for 6,047 of the total 6,100 pupils with observed data at the analysis stage (99%). FSM data for a further 46 observations was sourced from the 2016 Spring census (EVERFSM_6_P_SPR16). Given the nature of the Ever-6 FSM variable, the Spring 2016 and Spring 2017 iterations of the variable should be equivalent for pupils in Year 5. This takes the total proportion of pupils with FSM data to 99.9% (6,093 observations of 6,100).

School characteristics

Characteristics of the 181 randomised schools are presented in Table **8** below using data obtained pre-randomisation and for school years preceding the trial. The differences between the control and intervention groups are very small, under one percentage point both in terms of school background and in terms of the proportions relating to pupil attainment.

Table 8: Baseline comparison of school characteristics

Variable	Treatment (n = 89)		Control (n = 92)		Overall (n = 181)	
	Mean (missing/N)	Standard deviation	Mean (missing/N)	Standard deviation	Mean (missing/N)	Standard deviation
Proportion eligible for free school meals (2014/2015)	20.09% (0/92)	14.10%	20.49% (0/92)	14.80%	20.29% (0/92)	14.42%
Proportion speaking English as an additional language	6.80% (13/89)	13.42%	7.53% (12/92)	12.99%	7.18% (25/181)	13.17%
Number of pupils on roll	284 (13/89)	297	239 (12/92)	115	261 (25/181)	223
Proportion achieving Level 4 or above in reading at KS2 in 2013/2014[8]	81.92% (3/89)	13.81%	79.10% (2/92)	13.03%	80.47% (5/92)	13.45%
Proportion OFSTED rating = Good	74.64% (18/89)	43.81%	70.66% (17/92)	45.86%	72.60% (35/181)	44.75%

Pupil characteristics

Table **9** shows the baseline distribution of school and pupil[9] characteristics across the control and intervention schools, using all data available on each respective variable. There were 12,264 observations matched to the NPD at baseline in 2017, of which six were in the wrong year group. This match constituted:

- Year 5: a total of 6,034 matched records at baseline, 98.7% of the 6,116 pupils randomised; matched baseline data was available for 3,050 pupils in intervention schools and 2,984 in control schools; and
- Year 4 (results to be reported in the addendum report): a total of 6,224 matched records, 3,127 pupils in intervention schools and 3,097 in control schools, to be reported on at a later stage.

To assess the balance of pupil characteristics at baseline in accordance with EEF and CONSORT10 guidelines we provide descriptive tables of the distribution of pupil characteristics at baseline in the control and treatment schools using the baseline data as matched from the NPD in 2017—after randomisation but before the outcome data collection point. Only results for the original Year 5 cohort are presented here; we will update to present the total AR study sample in the addendum report.

Given that appropriate randomisation procedures were followed, any differences between control and treatment groups at baseline will be by definition due to chance; classical statistical testing is, therefore, unnecessary. Imbalance can be assessed by looking at differences in means and distributions of variables. There is a convention in some disciplines that a ten percentage point or larger difference in intervention and control means at baseline constitutes 'imbalance' and would justify including those measures in sensitivity analyses. To summarise: neither eligibility for FSM or gender show a difference higher than two percentage points, with all measures displaying differences of below one percentage point.

Using the matched baseline data, the Ever-6 FSM eligibility (Ever-6) rate was 36.43% for intervention schools and 36.60% for control schools. The missing data rate for FSM eligibility was 0.08% across both control and intervention groups, with four observations missing in the control group and one in the intervention group, respectively. Please note that these analyses are not carried out using outcomes data; they refer to Ever-6 FSM eligibility as at baseline. The

[8] The Statistical Analysis Plan listed a baseline comparison on KS1 attainment at the school level. This, however, was not undertaken at the point of data availability and was replaced by KS2 attainment. The original KS1 data is no longer available.
[9] All baseline equivalence analyses referring to pupils use samples as randomised (but post-correction as the NPD source contained the correct allocation of pupils in year groups).
[10] http://www.consort-statement.org/checklists/view/32-consort/510-baseline-data

gender distribution saw very similar proportions of girls in the intervention and control schools. In the Year 5 cohort, girls made up 49.28% of pupils in the intervention schools and 49.03% of pupils in control schools.

Table 9: Baseline comparison of pupil characteristics—categorical variables only

Variable	Treatment group		Control group		ES
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage	
Ever-6 eligible for FSM (2015/2016)	1,111/3,050 (1)	36.4%	1,092/2,984 (4)	36.6%	0.001
Female (2015/2016)	1,503/3,050 (1)	49.3%	1,463/2,984 (3)	49.0%	0.008

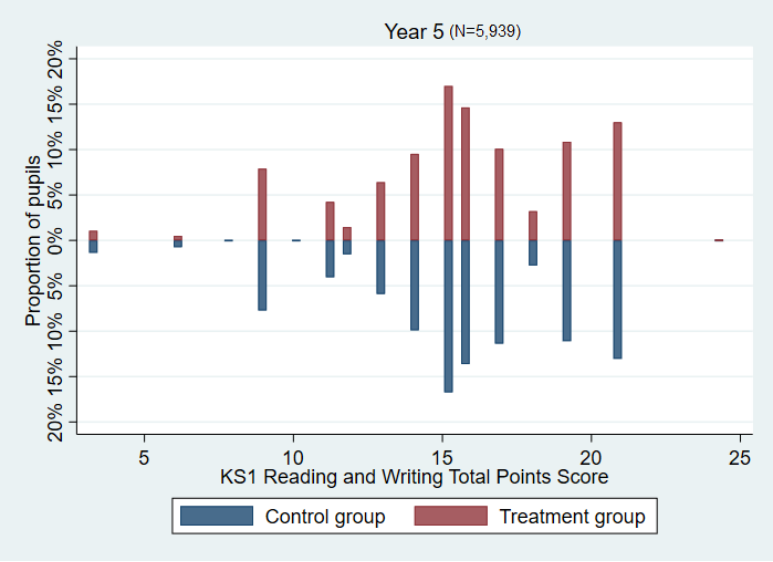
In addition to pupil characteristics, equivalence at baseline was also investigated in terms of prior attainment, specifically the reading and writing score at the Key Stage 1 (NPD: KS1_READWRITPOINTS).[11] Table **10** below presents the mean scores and standard deviations per intervention group and control group. The mean difference between the intervention and the control group was less than 0.2 of a KS1 point in the Year 5 cohort.

Table 10: Baseline comparison of pupil characteristics—continuous variable

Variable	Treatment group		Control group		ES
Pupil-level (continuous)	Mean (missing/N)	Standard deviation	Mean (missing/N)	Standard deviation	
KS1 R&W Score (2015/2016)	15.58 (39/3,050)	3.59	15.55 (56/2,984)	3.68	0.009

Figure 10 presents the distribution of scores, which again suggest equivalence between the control and intervention groups for each Year cohort.

Figure 10: Distribution of prior attainment KS1 reading and writing scores for Year 5 pupils



It is important to understand the characteristics of the pupils in the sample at analysis stage, in addition to the baseline stage reported previously. Table **11** presents the characteristics of pupils in the analysis sample, as at final analysis stage. For both treatment and control groups, the percentage of Ever-6 FSM pupils is roughly 37%. Just under 50% of the sample is female and a majority of the sample (78% in the control group, 85% in the intervention group) are White

[11] The Statistical Analysis Plan listed a baseline comparison on KS1 attainment at the pupil level. At the time of the baseline comparison, it was undertaken using the KS1 reading and writing attainment variable, which was justified given that the KS2 reading attainment variable is the primary outcome measure. However, at analysis stage, the KS1_APS variable was chosen instead of KS1_R&W because of the better distribution of data. The baseline comparison could not be re-done because the original NPD data that was used for the baseline comparison (which included all pupils as after the randomisation and the year group correction as opposed to the data used for analysis, which does not include data on the four treatment schools lost to follow-up) was no longer available at analysis stage.

British pupils. For attainment at Key Stage 1 (average points score), the treatment and the control group were very similar to each other (ES of difference = 0.000).

Table 11: Pupil characteristics at analysis stage—analytical sample

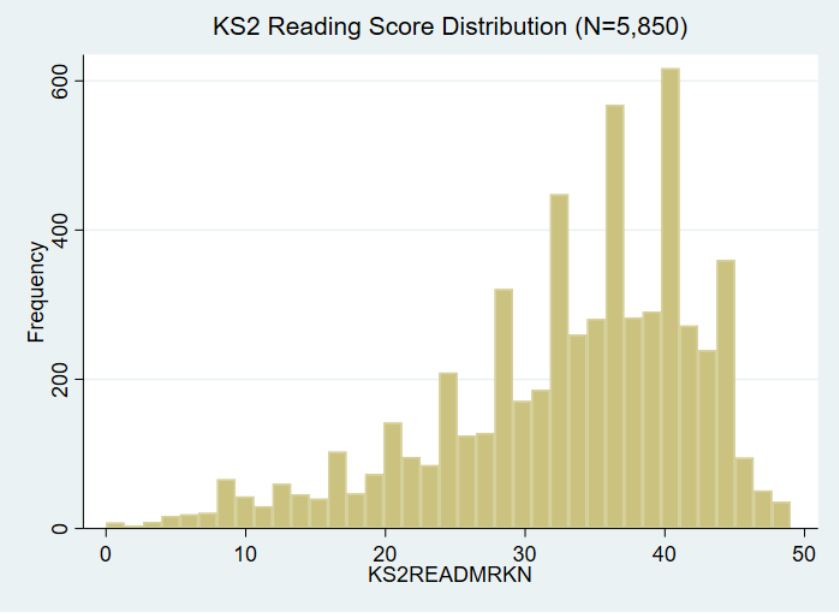
Variable	Treatment group		Control group		
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage	
Ever-6 FSM	1,092/2,920 (0)	37.40%	1,071/2,839 (0)	37.72%	
Female	1,456/2,920 (0)	49.86%	1,440/2,839 (0)	49.28%	
White British	2,498/2,920 (0)	85.55%	2,226/2,839 (0)	78.41%	
Pupil-level (continuous)	Mean (missing/N)	Standard deviation	Mean (missing/N)	Standard deviation	ES
KS1 Average Points Score	15.9 (0/2,920)	3.0	15.9 (0/2,839)	3.1	0.000

Outcomes and analysis

Primary outcome measure

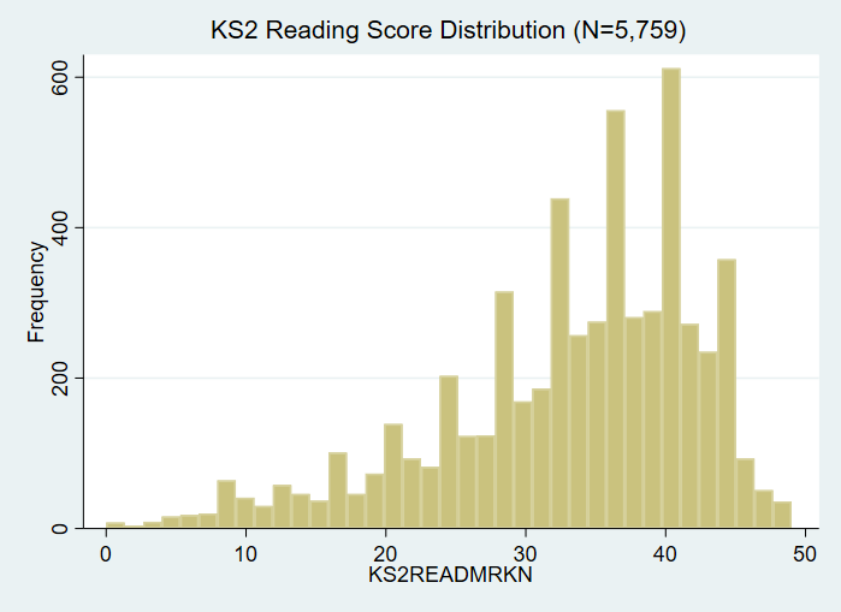
The primary outcome measure in the Year 5 analysis is the raw fine-grained Key Stage 2 reading score (KS2READMARKN). The distribution of this variable is presented in Figure 10. Of the 6,100 pupils across the control and intervention groups observed at follow-up (the 'Follow-Up' line in Figure 9), 5,850 pupils have a recorded reading score (95.9% of sample). For pupils with observed data (5,850), the outcome measure has a mean of 33.1 and a standard deviation of 9.4, from a range of 0 to 49.

Figure 11: Distribution of Key Stage 2 reading score outcome variable—all sample with data



For the analysis sample (5,759 pupils with observable data on all analysis variables), the primary outcome measure has a mean of 33.2 and a standard deviation of 9.3 from a range of 0 to 49 points. The distribution for the analytical sample is illustrated in Figure **12**.

Figure 12: Distribution of KS2 reading score outcome variable—analysis sample



Missing data

All variables included in the analysis display patterns of missing data that confirm that more than 95% of the relevant data is present and analysable:

- outcome measure, KS2READMRKN: 95.9% of data available;
- prior attainment measure, KS1_APS: 98.1% of data available;
- school region: 100% of data available;
- school average KS1 category: 100% of data available; and
- FSM eligibility (for subgroup analysis only): 99.9% of data available.

Therefore, no imputation or other adjustment for missingness was carried out.

Results

Results of the ITT analysis point to an effect size (adjusted for the multilevel nature of the model above) of 0.00 with a significance level of 0.936. This indicates that pupils' reading levels were similar in both treatment and business as usual schools at the end of KS2.

Table 12: Primary analysis (ITT)

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Primary outcome	2,920	33.31 (32.97, 33.64)	2,839	33.06 (32.71, 33.40)	5,759 (2,920, 2,839)	0.003 (-0.08, 0.09)	0.936

Table 13: Effect size estimation—primary analysis

Outcome	Unadjusted difference in means (intervention-control)	Adjusted difference in means	Intervention group		Control group		Pooled variance
			n (missing)	Variance of outcome	n (missing)	Variance of outcome	

Primary outcome	0.248	0.032	3,074 (154)	85.73	3,026 (187)	88.55	87.14

Subgroup analysis

The subgroup analysis focuses on pupils identified as eligible for free school meals. The analysis takes the same form as Equation 2 but subsets the sample by the FSM variable.

Table **14** and Table **15** present the results of the analysis for the FSM group, while Table **16** and Table **17** present the results of the analysis for the non-FSM group.

Table 14: Subgroup analysis (ITT)—FSM group

	Raw means				Effect sizes		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95%CI)	n (missing)	Mean (95%CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Primary outcome	1,092	30.62 (30.01, 31.23)	1,0.71	30.33 (29.72; 30.94)	2,163 (1,092; 1,071)	0.020 (-0.08, 0.12)	0.685

Table 15: Effect size estimation—FSM group

Outcome	Unadjusted difference in means (intervention-control)	Adjusted difference in means	Intervention group		Control group		Pooled variance
			n (missing)	Variance of outcome	n (missing)	Variance of outcome	
Primary outcome	0.292	0.209	1,166 (74)	104.42	1,151 (85)	103.07	103.72

	Raw means				Effect sizes		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95%CI)	n (missing)	Mean (95%CI)	N in model (intervention; control)	Hedges g (95% CI)	p-value

Primary outcome	1,828	34.91 (34.53; 35.29)	1,768	34.71 (34.32; 35.11)	3,596 (1,828; 1,768)	0.014 (-0.08, 0.11)	0.778

Table 16: Subgroup analysis (ITT)—non-FSM group

Table 17: Effect size estimation—non-FSM group

Outcome	Unadjusted difference in means (intervention-control)	Adjusted difference in means	Intervention group		Control group		Pooled variance
			n (missing)	Variance of outcome	n (missing)	Variance of outcome	
Primary outcome	0.198	0.115	1,906 (78)	67.73	1,865 (97)	72.55	70.09

The results of the subgroup analysis mirror those of the primary ITT analysis. In both the FSM subgroup and the non-FSM subgroup, pupils in the treatment schools perform as well as pupils in control schools (that is, the difference between the treatment and control groups is not statistically significant). In this large-scale trial that has used national assessment scores as the main outcome measure, the effect sizes for the FSM subgroup and the non-FSM subgroup are very small but positive, ranging between 0.01 and 0.02.

Fidelity

The main assumption behind an Intention to Treat approach is that the analysis follows directly from the allocation and does not allow for various levels of implementation fidelity. To address this, we have undertaken an analysis of fidelity using a continuous measure derived in consultation with Renaissance Learning.

The derived fidelity measure addresses two separate but related questions:

1. Pupil fidelity. To what extent are pupils in a school reaching a level of AR best practices implementation?
2. Teacher fidelity, taken to mean responsiveness. To what extent do we believe teachers are acting on AR data to help pupils improve their daily reading practice behaviours, particularly when pupils had previously demonstrated low fidelity?

The fidelity variable reflects a total score that combines both pupil and teacher fidelity, for reasons outlined below.

Key to the AR programme is the extent to which teachers review and act upon reading practice data via AR. Documenting teacher behaviour with precision is difficult in the present study: we lack classroom observations, cameras, teacher interviews, or other means to know to what extent they are reviewing and taking action based on AR data. The proposed responsiveness measure captures this in the absence of more direct (but practically unfeasible) observation. The measure focuses on pupils who have demonstrated low fidelity in a prior term. Teachers would be credited with responsiveness if those pupils improved their performance over time. Although this approach has limitations, pupils who do little reading or do not read challenging material are unlikely to change course without some involvement from a teacher or librarian, so using teacher responsiveness data seems like a reasonable proxy and the best we can do at this time.

Some teachers may also achieve a high degree of fidelity early in the year and may be able to maintain it throughout the year. They may not be recognised, therefore, in the teacher responsiveness calculation since their pupils are already performing well. Thus, direct accounting for pupil fidelity credits teachers who maintain strong performance.

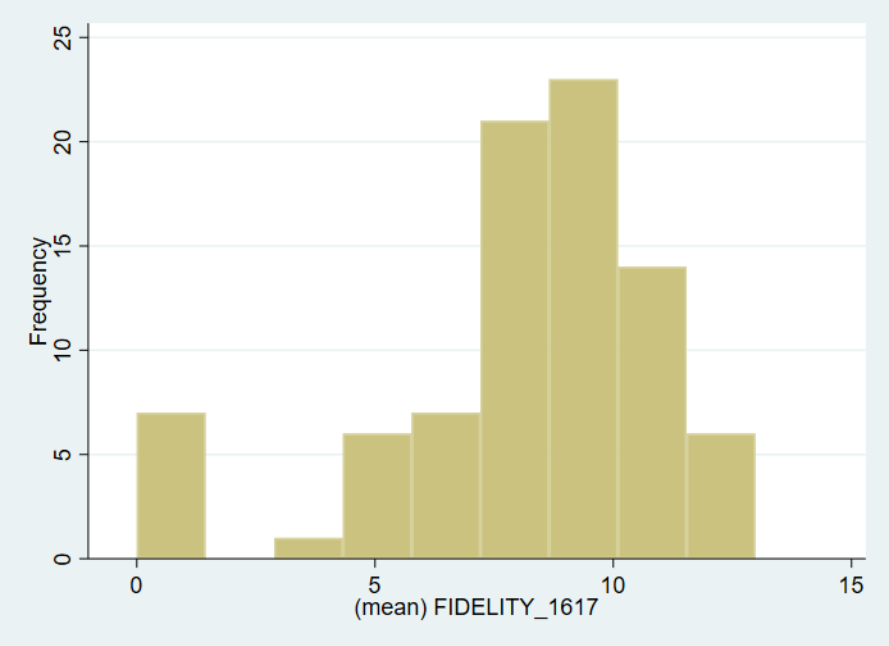
The fidelity measure is calculated as follows, only for the 2016/2017 year (the first year of implementation):

1. Determine pupil fidelity, separately for each term, by coding each of the three pupil-relevant ‘non-negotiable’ aspects of AR: (1) average percent correct (APC), coded as 1 if APC ≥ 85%, else 0; (2) engaged reading time (ERT), coded a 1 if ERT ≥ 15 minutes, else 0; and (3) reading within the zone of proximal developed (ZPD), coded as 1 if the average book difficulty level read by pupils is within, or above, the ZPD, else 0.

2. Determine teacher fidelity by identifying pupils from Term 1 who were coded as 0 in APC. Determine if their APC score was higher in Term 2 compared to Term 1, coding the teacher APC as 1 if this is the case, else coded as 0. For pupils whose APC was 1 in the first term, assign teachers a value of 0.75, which still assumes that teacher responsiveness may be higher in schools where initial fidelity is not achieved, but does not bias the score. Repeat procedure for ERT and ZPD, and for the shift from Term 2 to Term 3.
3. Then sum the above score to create a final fidelity score where a score of 0 would indicate lack of fidelity across all aspects. A score of 15 would suggest perfect achievement of all pupil and all teacher fidelity criteria (score distribution in Figure **13**).

This raw fidelity score indicates a high degree of implementation fidelity with a skew towards higher levels of fidelity at school level (Figure **13**).

Figure 13: Distribution of 2016/2017 fidelity score in treatment schools (school level)

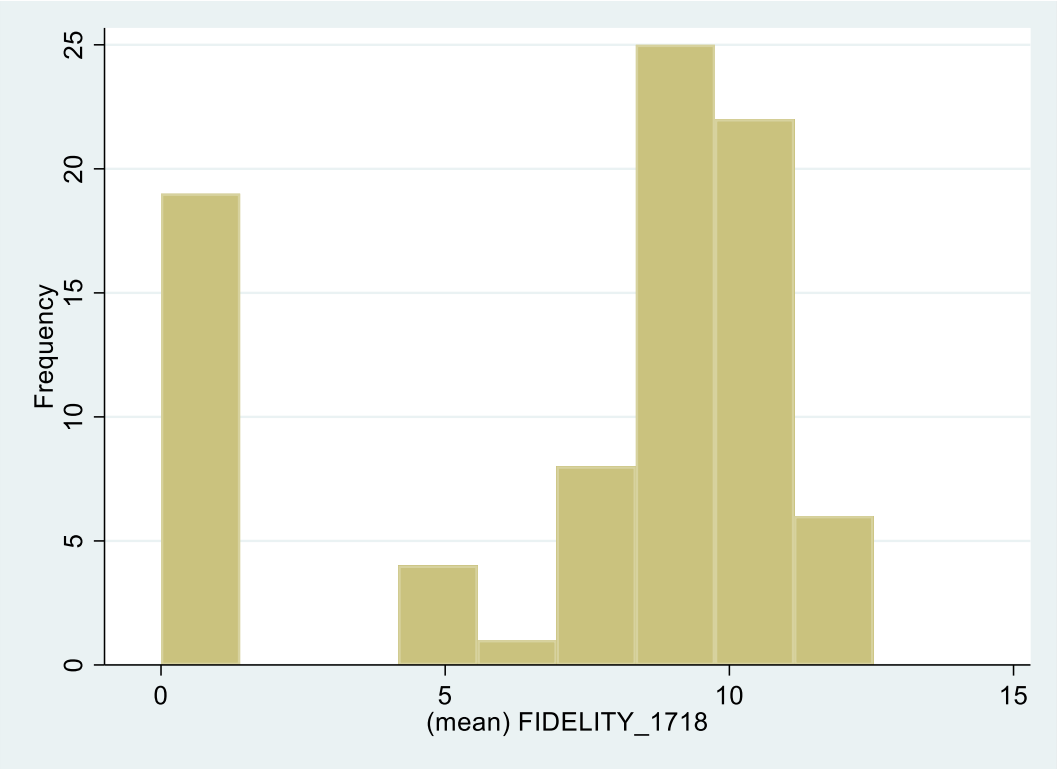


Once the continuous fidelity score above was calculated (at the school level), a two-stage least squares analytical approach was then taken whereby the fidelity score above was regressed on the treatment allocation variable and residuals predicted. Control schools were assigned a fidelity score of 0 (no implementation of AR). The predicted standardised residuals were then entered into a multilevel model as per Equation 2, replacing the treatment variable with the newly-estimated standardised predicted residuals fidelity variable. Results suggest a high level of implementation fidelity throughout but a small, non-statistically-significant impact of increased fidelity on the pupil KS2 reading outcome (coefficient for continuous fidelity variable = 0.11, p = 0.281, N = 5,759; 95% CI: -0.09, 0.311).

Second year fidelity

As a separate, exploratory-only analysis, the relationship between second year (2017/2018) fidelity and the primary pupil reading outcome was also explored. The fidelity measure for 2017/2018 was calculated as above, using exactly the same formula as for 2016/2017. The distribution of the raw fidelity score for 2017/2018 is illustrated in Figure **14**. This suggests that although lower than in the first year, overall fidelity of implementation in the second year was still high. The higher number of schools with fidelity 0 reflects the choice of several schools not to continue to implement AR in the second year of the trial, as discussed in the Implementation stoppage and attrition section. The pair-wise correlation between first and second year fidelity measures at school level is $r = 0.88$, $p < 0.05$.

Figure 14: Distribution of 2017/2018 fidelity score in treatment schools (school-level)



The same two-stage least squares (2SLS) approach was then applied, again with control schools assigned a fidelity score of 0 (no AR implementation) ahead of it. The predicted standardised residuals were then entered into a multilevel model, also mirroring the first year fidelity score. Again, results suggest a fairly high level of fidelity in the second year of implementation but no statistically significant relationship between second year fidelity and KS2 reading outcomes: coefficient for continuous fidelity measure after 2SLS = 0.11, p = 0.122, N = 5,759; 95% CI: -0.023, 0.25.

Cost

Box 2: Costs—key findings

- The per pupil per year marginal cost of implementing AR over three years is estimated to be £47.83. The cumulative per pupil cost over three years is £143.49. The marginal cost calculation of the intervention includes costs for the annual license for AR, a one-off training cost, and costs for purchasing books at start-up and in subsequent years.
- Costs not included in the marginal cost calculation:
  - Costs for new equipment are considered part of prerequisite costs since these were reported by few survey respondents. There was considerable variation in the level of costs reported in both the first year (for start-up of AR) and second year (running of AR). When looking at the range of costs reported (more specifically the higher end of the range), in the first year, higher costs were reported than in the second year.
  - Staff time and costs are not included in the marginal cost calculation. However, additional time spent on AR, whether in the start-up phase or during the weekly running of AR, was the most commonly reported additional cost of AR. The level of costs reported varied widely across participants. Other staff costs were less commonly reported (for example, additional direct staff costs, extension of staff contract hours, or supply cover).

Marginal cost

The Accelerated Reader programme was implemented over two school years (2016–2018) in 89 schools. For the estimated average marginal cost per pupil, we included the cost of the AR license (non-discounted cost of buying annual licenses) and books. Per pupil per school costs are based on an average of 70 pupils per school in the intervention arm of this trial.12

The cost of the AR license was provided by Renaissance Learning to the evaluation team. The cost of purchasing AR varies depending on the number of students that will be using AR in a school. Estimates provided in Table **18** are based on schools purchasing an AR license for 70 pupils as this was the average number of pupils in intervention schools in this trial. (Please note that the license cost per pupil reduces as the number of pupils for which it is purchased increases.) These costs are both for accessing Accelerated Reader and the Star Assessment, which are sold as a package. Renaissance Learning always includes training during the first year (three remote sessions and a full day onsite) to guide schools through the initial set-up and help them implement the programme according to best practices. Schools can then choose whether to include additional training during later years. They can also reduce the per student cost by signing up to multi-year deals. A three-year license (10% discount if paid up front) would amount to £2,873.70 or £13.68 per student per year for a license for 70 pupils. This information is summarised in Table **18** and incorporated in the calculations of Table **19**.

Table 18: Cost of AR license per year based on 70 pupils (example provided by Renaissance Learning)

Primary Schools	Example of 70 pupils	Per student per annum	Training cost (3 remote sessions and 1 onsite training day)
First year	£991.00	£14.16	£1,101
Subsequent years	£1,101	£15.73	
Three-year license (10% discount if paid up front)	£2,873.70	£13.68	

Books were a key component of the Accelerated Reader programme as schools had to ensure they had books that aligned with those in the programme. Book costs were reported by headteacher and librarian/literacy coordinator survey respondents in the first, second, and endline staff surveys, and were then averaged and rounded to the nearest £50 to provide an estimate for both the start-up phase and additional cost in the second year; please see Table **19**, where the range of costs is also reported.

Books could be considered a prerequisite rather than marginal cost, but the high proportion of respondents reporting book-related costs in our survey led us to decide that books should be classified as a marginal cost instead (as per EEF,

[12] In this trial, at randomisation there were 6,276 intervention pupils (Year 4 pupils: n = 3,088; Year 5 pupils: n = 3,188) in 89 intervention schools meaning that on average there were 70 pupils per school (6,276 pupils divided by 89 schools).

2016). In total, 57 headteacher or librarian/literacy coordinator respondents reported book costs for the first year of AR implementation; 21 reported book costs for the second year.

Table 19: Estimated cost per school for delivering AR

Item	Type of cost	Cost	Total cost over 3 years
Annual license fee	Running cost per school (for 70 pupils; non-discounted)	£991 (Year 1)	(Year 1 £991 + Year 2 and 3, 2 x £1,101) = £3,193
One-off training cost	Start-up cost (covering 3 remote sessions and 1 onsite training day)	£1,101	£1,101
Start-up cost for books	Start-up cost per school	£1,350 (range: £0 to £7,000; mean: £1,237; SD: £1,549) (n = 57)	£1,350
Additional cost for books in subsequent years	Running cost per school	£2,200 (range: £200 to £10,000; mean: £2,159; SD: £2,650) (n = 21)	(£2,200 x 2) = £4,400
Total			£10,044

Table 20: Estimated average cost per pupil over three years

	Year 1	Year 2	Year 3
Approximate cost per pupil per year	£49.17 (£991/70 + £1,101/70 + £1,350/70)	£47.16 (£1,101/70 + £2,200/70)	£47.16 (£1,101/70 + £2,200/70)
Cumulative cost per pupil per year	*£49.17*	*£96.33*	*£143.49*

Costs not included under marginal cost

In contrast, we excluded from the cost-per-pupil calculation the cost of equipment as well as labels (and related costs) as these were reported by a small proportion of survey respondents and appeared more likely to fall under prerequisite than marginal costs. For instance, equipment costs listed by 16 headteacher or librarian/literacy coordinator respondents for the first year of AR implementation included existing or additional tablets, laptops, tablet trolley, scanner, bar code reader, and investment into WiFi. These costs ranged from nothing (as tablets were already owned by the school) to £25,000 (for 'technology'). For the second year of AR implementation, four headteacher respondents reported equipment costs such as tablets and docking stations ranging from £1,000 to £3,200. Costs related to labelling included the labels themselves and labelling kits and ranged from £20 to £1,000. These costs were reported by four headteacher and five librarian/literacy coordinator respondents for the first year of AR implementation and by one of the former and three of the latter for the second year of implementation.

Staff time and cost

As per the EEF's guidance on cost evaluation (EEF, 2016), we excluded school staff time from the cost per pupil per year estimate but present our findings on these costs in the narrative below. In summary, by far the cost most frequently reported by respondents was additional time spent on AR, whether in the start-up phase or during the weekly running of AR. However, it should be noted that the amounts of time reported ranged widely across participants. In contrast, in relation to other staff costs (such as additional direct staff costs, extension of staff contract hours, or supply cover) fewer participants reported these, though there was still a wide range of costs reported. Moreover, it should be noted that the findings presented are informed by responses in the first, second, and endline staff surveys. Findings from the surveys

relied on self-report and participants who responded to the surveys may not have been representative of all staff in intervention schools. Taking these limitations into account, further details on staff time, direct staff costs, extension of staff contract hours, and supply cover are presented below.

Staff extra time

First, we asked survey respondents to provide estimates for staff extra time (though some respondents provided estimated *cost* of staff extra time instead). For the first year of AR, 38 headteacher respondents (across the first and second staff surveys) reported that staff spent an average of 31.3 additional hours for the start-up phase of AR while 149 teacher, TA, or literacy coordinator respondents (across the first and second staff surveys) reported that staff spent an average of 12.9 additional hours for the start-up phase of AR. For the weekly running of AR for the first year of implementation, across the first and second staff surveys, 59 headteacher respondents reported that staff spent an average of 1.6 additional hours per week while 242 teacher, TA, or literacy coordinator respondents reported that staff spent an average of 0.9 additional hours per week. For the second year of implementation, 14 headteacher respondents reported that staff spent an average of 0.7 additional hours each week, while 65 teacher, TA, or literacy coordinator respondents reported that staff spent an average of 1.0 additional hour each week. These findings are presented in Table **21**. In addition, ten respondents (headteachers or librarian/literacy coordinators) provided estimated costs across both years of AR implementation, ranging from £130 to £2,250 for staff additional time (for example, time to reorganise book stock, time to label books, and staff overtime).[13]

Table 21: Estimated average additional time spent by staff on AR

	Average additional time in start-up phase	Average additional time spent running AR each week – Year 1	Average additional time spent running AR each week – Year 2
Headteachers perceptions	31.3 (range: 0–150 hours) (n = 38)	1.6 (range: 0–10 hours) (n = 59)	0.7 (range: 0–4 hours) (n = 14)
Teachers/TAs/librarians/literacy coordinators perceptions	12.9 (range: 0–240 hours) (n = 149)	0.9 (range: 0–5 hours) (n = 242)	1.0 (range: 0–5 hours) (n = 65)

Staff direct cost

In addition, we asked survey respondents to provide estimates for direct staff costs. For these, we asked survey respondents about new staff hired and extension of staff contract hours in both years of AR implementation. In terms of new staff hired in the first year of AR implementation, one headteacher respondent and five headteacher respondents reported hiring new staff for the start-up phase and weekly running of AR, respectively (though one did not provide a cost figure). One librarian/literacy coordinator respondent reported hiring extra staff (at the cost of £300) in the first year of implementation, though it was unclear if that was in the start-up phase or subsequently. In the second year of AR implementation, only one headteacher survey respondent reported hiring new staff for the weekly running of AR, at the cost of £5,000. These findings for direct staff costs reported by headteachers are presented in Table **22** (librarian/literacy coordinator responses are not included as it was not clear during which phase the costs were incurred).

Table 22: Estimated average additional direct staff costs for AR

	Average direct staff costs in start-up phase	Average direct staff costs for running AR each week—Year 1	Average direct staff costs for running AR each week—Year 2
Headteachers perceptions	£600 (n = 1)	£4,000 (range: £0–8,000) (n = 4)	£5,000 (n = 1)

In terms of extending contract hours, five headteacher respondents reported extending staff contract hours for the start-up phase of AR, and reported costs ranging between £250 and £1,500. Two headteacher respondents reported extending staff contract hours for the weekly running of AR for the first year, and reported costs of £30 and £2,000. For

[13] While it may be desirable to have a single estimate of additional staff time for implementing and running AR to which schools can refer to, the evaluation team considers that the two measures of additional time should not be aggregated. The main reason being that there is substantial variation between answers, which would potentially inflate a summed estimate of additional time.

the second year of AR implementation, no headteacher respondents reported extending staff contract hours for the weekly running of AR. These findings on staff contract hour extensions are presented in Table **23**.

Table 23: Estimated average costs of extending staff contract hours for AR

	Average cost of extending staff hours in start-up phase	Average cost of extending staff hours for running AR each week – Year 1	Average cost of extending staff hours for running AR each week – Year 2
Headteachers perceptions	£810 (range: £250-1,500) (n=5)	£1,015 (range: £30-2,000) (n=2)	NA (n=0)

Staff supply cover

Finally, we asked survey respondents to provide estimated cost for supply cover needed for the start-up phase of AR as well as for running AR in both years of implementation. Twenty-four headteacher and one librarian/literacy coordinator respondents reported needing to hire supply cover for the start-up phase of AR, with costs ranging from £70 to £1,400. For the first year of implementation, two headteacher respondents reported requiring staff supply cover for the weekly running of AR but did not provide associated costs. One librarian/literacy coordinator respondent reported requiring staff supply cover (at the cost of £450) in the first year of implementation, though it was unclear if that was in the start-up phase or subsequently. For the second year of implementation, one headteacher respondent reported needing supply cover to release time for teaching training and one headteacher respondent reported needing supply cover for the weekly running of AR; the costs were £1,000 and £800 respectively. These findings for staff supply cover costs reported by headteachers are presented in Table **24** (librarians/literacy coordinators are not included as it was not clear which phase the costs were incurred in).

Table 24: Estimated average costs of staff supply cover for AR

	Average cost of extending staff hours in start-up phase	Average cost of extending staff hours for running AR each week – Year 1	Average cost of extending staff hours for running AR each week – Year 2
Headteachers	£556 (range: £70–£1,400) (n = 24)	Not reported (n = 2)	£900 (range: £800–£1,000) (n = 2)

Implementation and process evaluation

In the following sections we discuss the findings in relation to implementation, fidelity, perceived outcomes, formative findings, and control group activity. It should be highlighted here that the findings reported in this section draw largely on survey, workshop, and interview findings, which represent respondents' self-reports and perceptions and may not be representative of the study sample as a whole. With the surveys, while the same question was often asked over time in multiple surveys, it was not possible to track individual respondents.

Methods

Baseline contextual information was collected via a survey for both control and intervention schools. This included information about the use of literacy-focused programmes and the status of the school libraries.

The evaluation team conducted *observations of a sample of the training sessions* for staff in the intervention schools, which, as described in the introduction, included three remote telephone sessions via WebEx/telephone and one onsite session per school. The evaluation team observed seven of the remote telephone sessions (by remote login)[14] and the onsite training session in eight schools. These observations were accompanied by a review of the training materials and a review of the extensive online resources. These observations enabled the team to better understand the aims of the training, its delivery, how it was initially received by participants, to examine possible differences which might be due to different trainers in different locations, and the extent to which training matched up to the intervention logic model. Semi-structured observation protocols were used to capture this information.

[14] The evaluation team intended to observe an additional remote telephone session but it was cancelled. The only participant who was going to attend the session overbooked herself. This session was rescheduled but we did not observe it.

A *first intervention online staff survey* was administered by NFER in all intervention schools in February 2017. The survey was administered to headteachers, Year 4 and Year 5 teachers, TAs, librarians, and literacy coordinators. In this and all subsequent surveys, participants were informed that their responses would be kept confidential and would not be shared with their schools or outside the delivery and evaluation team. The survey was intended to be conducted approximately half-way through the intervention period (that is, halfway through a single intervention year) and serve as a midline data collection point, but the extension of the trial meant that the survey collected data during the first-third of the intervention period.

Following initial analysis of the first intervention staff survey, the evaluation team undertook two *workshops with staff* in a subset of treatment group schools in April 2017, selected from the original regional focus of the EEF (that is, North East and North West regions). These workshops served two main purposes. First, they explored key points from the survey results to either validate or challenge survey findings through better understanding the underlying dynamics of the implementation of AR in practice. Second, the workshops explored these issues in further depth than is possible through a survey. They encouraged workshop participants to elaborate on issues identified in the early stages of the process evaluation such as teachers' experiences of the AR training sessions and the AR help resources, staff's experiences with the implementation of the intervention, barriers to the implementation of the intervention, and staff's perceptions of students' experience with AR. Researchers captured information from the workshops using semi-structured protocols designed to capture the information described above.

The evaluation team planned the workshops so that they would be located in areas with more than one (and ideally three to five) intervention schools in nearby commuting distance of one another and invited two to three staff members from each school to attend. For each workshop, one school was asked to host, ideally the most centrally-located school or the one most easily accessible by transit. We did not expect to be able to randomise the teacher selection for these workshops, but hoped to work with each school to identify appropriate participants who were both available and willing to participate. Participants were provided with refreshments and, where appropriate, certain travel costs could be covered by the project to ensure adequate participation.

A *second intervention online staff survey* was administered by NFER in June and July 2017. Similar to the first online staff survey, this survey was administered to headteachers, Year 4 and year 5 teachers, TAs, librarians and literacy coordinators, and headteachers in all intervention schools. The survey was intended to be conducted at the end of the intervention period (that is, halfway through a single intervention year) and serve as an end-line data collection point, but the extension of the trial meant that the survey collected data during the mid-point of the intervention period.

Additionally, we administered an *online survey for headteachers in control schools* in February 2018. The survey sought to provide indicative data that may explain any differences (or similarities) in performance identified between treatment and control group schools. The control school headteacher survey was intended to be conducted at the end of the intervention period (that is, at the end of a single intervention year) and collect information about practices in control schools during the year. However, the extension of the trial led the evaluation team to conduct the survey in February 2018 and collect data about both intervention years (2016/2017 and 2017/2018) while minimising the data collection burden for control schools.

At the end of the two-year intervention period (during June and July 2018), *endline staff and headteacher surveys* were administered by NFER in intervention and control schools respectively to understand whether, and how, practice has changed in treatment and control schools, as well as to further explore the barriers or challenges faced by the staff.

While we initially intended to hold further workshops with staff in intervention schools after the second survey, similar to the approach described above, our experience of conducting the first round of workshops (there was high participant drop-out) led us to arrange *staff interviews* instead so that we could better target participants. As the planned workshops were intended to provide an opportunity, at the end of the intervention period, to gather in-depth information regarding staff's perceptions regarding the overall implementation of the intervention (and barriers to implementation) and of its effectiveness and impact, we instead conducted the interviews at the end of the extended intervention period (at the end of school year 2017/2018). Semi-structured interview protocols were used to explore this information. Potential interviewees were approached to ensure that respondents covered a range of school factors including whether their schools continued to use AR in the second year and, if so, the school subscription type and level of engagement with AR.

In the second year of the trial, the routine administrative data collected by AR will form the basis of understanding whether there has been any 'contamination' (that is, whether the Year 5 and Year 6 pupils in control schools have accessed AR). One interesting question is whether treatment schools continue to use AR—something that can be tracked using routine data. Similarly, in those schools that do not continue with AR, the question will be what else they are doing, which is something we aim to capture in the surveys.

Implementation and fidelity

Box 3: Implementation and fidelity—key findings

- IPE evidence on the set-up of AR and daily running (non-negotiables) suggests that intervention schools were in general implementing AR as intended.

Set-up

- IPE data shows that most intervention teachers and relevant literacy coordinators and librarians (part of the core AR team implementing the intervention) received AR training. Only a small minority of teachers, teaching assistants, headteachers, and one literacy coordinator reported that they had neither attended the training sessions conducted by RL nor any alternative training.
- Reports from survey responses and the two workshops showed that participants perceived the training as useful. Participants reported feeling prepared for implementing AR after receiving the training.
    - The vast majority of respondents to the intervention staff surveys reported having clarity over the four non-negotiables of AR—investment on book stock, dedicated daily reading time, regular quizzing, and regular use of AR data.

Non-negotiables in practice

- *Book stock*: IPE data showed that there was some discrepancy in respondents' feelings about the schools' investment in books. Librarians or literacy coordinators were more likely to report feeling that there was insufficient funding to purchase new books, compared to headteachers and teachers. In the endline survey, there was less discrepancy with most survey participants responding that there was sufficient funding for books.
- *Reading time*: most survey respondents reported having no barriers to providing dedicated reading time for AR during class time. On average, survey respondents reported that pupils had an average of 21–30 minutes protected reading time per day (while this is slightly less time reading time than the recommendation of at least 30 minutes for using AR, it is higher than the 15 minutes of engaged reading time which RL and the evaluators agreed was a measure of pupil fidelity).
- *Regular quizzing*: data collected through the IPE suggests that most teachers implementing AR in their classrooms made regular use of STAR Reading tests and quizzes. Some teacher interviewees reported that the quizzing element of AR had, in particular, helped pupils to develop a habit of reading.
- *Regular use of AR data*: in the intervention surveys almost all teacher and TA respondents reported having looked at the reports generated by AR at least once, either on their own or with colleagues in their school.

As described above, the AR intervention is aimed at increasing student reading progress as well as their engagement in, and motivation for, reading. Crucial to achieving this goal are the four non-negotiables of AR—making sure there are adequate resources in terms of books, dedicating protected and sufficient time for pupils to read daily in school, creating a feedback loop between reading and assessment through regular quizzes, and teachers regularly using the data generated through AR. This latter aspect is included in the calculation of the fidelity score whereby teacher fidelity includes actions being taken on observing that the pupil fidelity (in relation to reading within the ZPD, quizzing, and engaged reading time) was not met, possible only through regular use of the AR-generated data. There is also an important element of staff training, including mandatory and optional training, to ensure that staff understand how to implement AR as intended.

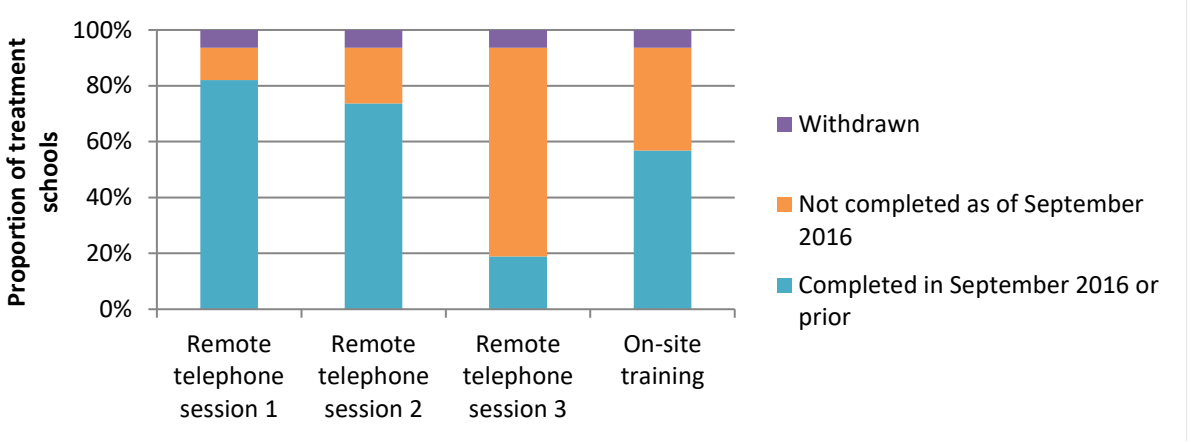
This section discusses the implementation of AR with respect to three factors: (i) the training offered and delivered to schools implementing AR, (ii) the perceived effectiveness of the training, and (iii) the four 'non-negotiable' elements of AR and perceived barriers to those elements.

Overall, training sessions were well attended

To implement AR, schools are offered a number of training opportunities of which the three remote telephone sessions and one onsite training day are mandatory. In addition, schools can dial in to webinars and workshops. Ideally, all schools should have had their remote telephone sessions and onsite training in the summer term of 2016 before the intervention started at the start of the school year in September 2016 as the training helps schools to understand the intervention and to prepare their library. However, the administrative data on training uptake from RL showed that not all schools had received their mandatory training sessions. In particular, by September 2016, only 20% of schools (18 schools out of 89) had taken part in their third remote telephone session (see Figure **15**). In practice, the third remote telephone session also often took place after the onsite training day. The latter was likely due to schools needing to have tested at least some pupils with STAR Reading to use this data during the third remote telephone session. Administrative data provided by Renaissance Learning up to November 2016 showed that 44% of schools (30 out of 89) had completed the third remote telephone session by this time. The evaluation team did not revisit data on uptake of the third remote telephone session beyond this date as the evaluation team considers that receiving the third remote

telephone session—which focused on analysing STAR Reading test data—after the onsite training (or not at all) is relatively less central to a full understanding of AR since a larger proportion of schools reported receiving the onsite training. The onsite training also covered how to analyse STAR Reading test data and provided a refresher of the aspects covered in the first two remote telephone sessions. As discussed below, only a small proportion of staff survey respondents reported not receiving any AR training at all, suggesting that a majority of AR school staff received the necessary guidance to set-up and run the intervention.

Figure 15: Overall school training uptake as of September 2016



n = 89 schools. Source: Renaissance Learning administrative data.

We observed seven remote telephone training sessions. Overall, attendance levels at the telephone training sessions varied (range: one to seven teachers). The level of prior knowledge and basic familiarity with AR by attendees also varied (for example, in one session some participants had never visited their AR school site (logged in); in others, teachers had specific questions about how to use aspects of AR).

We also observed eight onsite training days between June and September 2016 in schools in the clusters of Nottingham, Newcastle, and Liverpool. At all the training days we observed, there was good attendance from Year 4 and Year 5 teachers. In addition—and while this was not mandatory—in nearly all cases the literacy coordinator was present with one exception in which the deputy head attended instead. In one instance librarians and teaching assistants also joined.

As part of the onsite training, headteachers were required to attend at least part of the session as per RL requirements. In six of the eight training days we attended, headteachers had contact with the trainer either during the initial session or during lunchtime. However, in two instances the headteachers were not involved at any point during in the training day. In one case, while the headteacher was present at the school, it became evident that there were competing priorities (an imminent Ofsted inspection) that hindered her ability to make time to speak to the trainer. In the second case we observed, the AR trainer did not insist on fulfilling this requirement.

The fact that not all staff had received the required training was also reflected in the findings of the first and second intervention staff surveys. In these surveys a small number of staff—8 out of 145 and 6 out of 53 respondents respectively, including teachers, teaching assistants, headteachers, and one literacy coordinator—reported that they had neither attended the training sessions conducted by RL nor any alternative training offered. However, these numbers are small and likely did not have a significant impact on the overall implementation of AR in the trial.

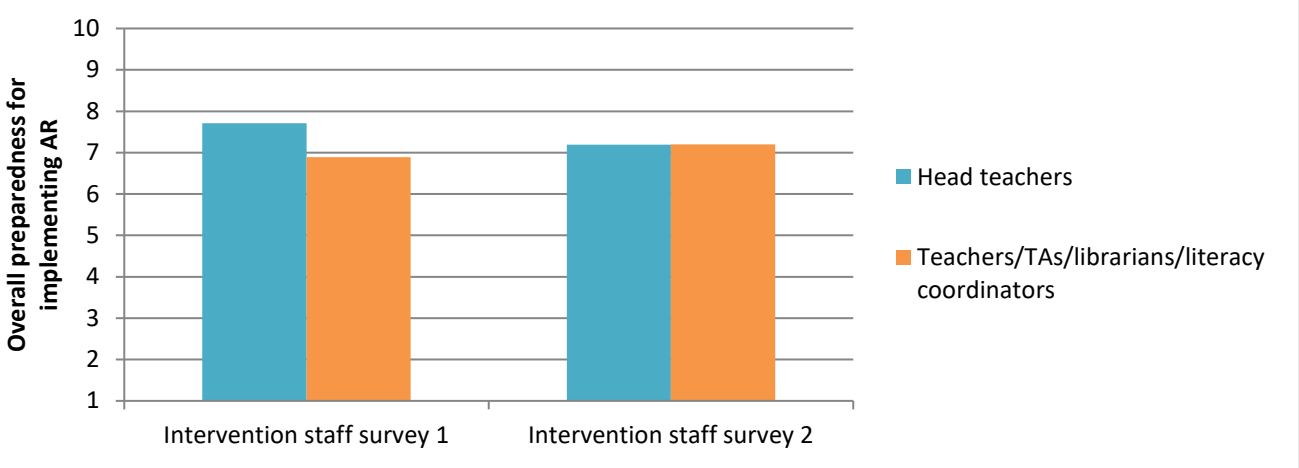
Training was found to be of good quality and effective

Overall, respondents to the intervention staff surveys reported that in the training they had attended (webinar, remote telephone session, or onsite training day) the information provided was clear and useful, participants were able to receive helpful information in response to questions specific to their school, and the trainers were knowledgeable.

After the onsite training day, the vast majority of respondents to the intervention staff surveys reported having clarity over the four non-negotiables of AR. However, survey responses indicated that in relative terms the least clear non-negotiables were investments in book stock and the regular use of the data generated through AR. In the survey, 18 out of 99 and 17 out of 99 respondents, respectively, reported that these two elements were not clear to them following the training; for the other two non-negotiables (reading time and quizzing) nearly all participants reported having clarity.

Nevertheless, across both surveys, when asked to rate their overall preparedness for implementing AR (with 1 representing 'completely unprepared' and 10 being 'fully prepared'), staff's reported preparedness to implement AR was moderately high (as presented in Figure **16**).

Figure 16: Survey respondents' average reported overall preparedness for implementing AR



Intervention staff survey 1: teacher/TA/librarian/literacy coordinator, n = 107; headteacher:, n = 35. Survey 2: teacher/TA/librarian/literacy coordinator, n = 139; headteacher, n = 26.

In addition to the surveys, the IPE also captured information about the quality of the training through training observations. Apart from eight onsite training days observed between June and September 2016, we also observed seven remote telephone training sessions and four webinars between May and June 2016. The remote telephone sessions were mandatory and delivered to individual schools and served as a space to ask questions specific to the participant's context while webinars were optional and open to both new and experienced AR users where different experiences could be shared.

In each of the remote telephone sessions and webinars we observed, trainers came across as knowledgeable and the training provided was consistent with AR guidelines. The trainers displayed a good understanding of the programme and were able to answer participants' questions. Furthermore, during all sessions, trainers ensured participants could clarify doubts by checking at the end of a topic whether the content just delivered was clear. Similarly, at the onsite training days, we observed that trainers were systematic in providing the core information about AR. They made clear the importance of the four 'non-negotiables' of AR. In one instance a minor discrepancy was observed when a trainer told participants 20 minutes of dedicated reading time would be sufficient while AR guidelines outline that at a minimum reading time should last for 30 minutes daily.

Overall, trial participants were aware of the four non-negotiables and put them into practice

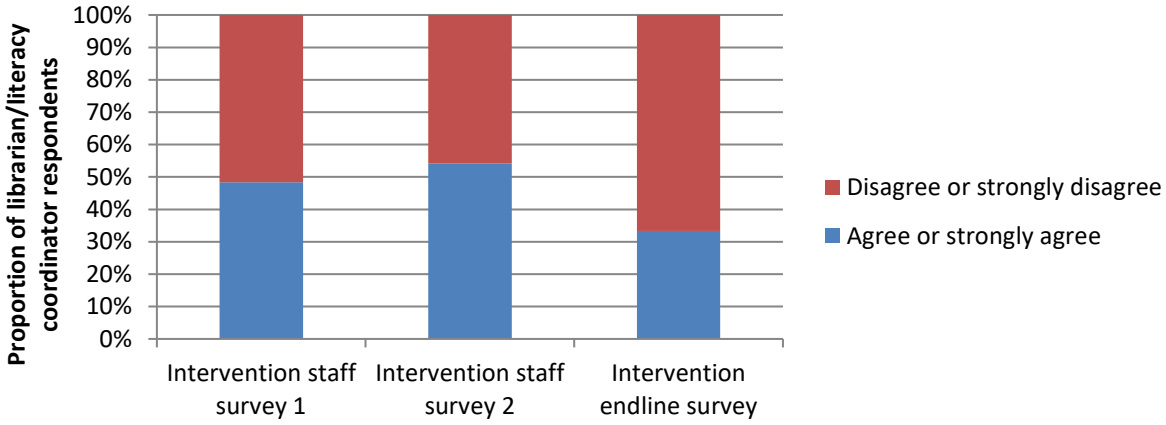
In the remainder of this section we report on observations and participants' perspectives of the non-negotiable elements and the barriers to achieving them.

Reports on whether investments in book stock was sufficient varied but at endline, higher proportions of survey respondents disagreed that it was insufficient

IPE data presented in this section shows that there was some discrepancy in respondents' feelings about the schools' investment in books, particularly at the beginning and middle of the evaluation. Librarians or literacy coordinators were much more likely to report feeling that there was insufficient funding to purchase new books, compared to headteachers and teachers. By the time of the endpoint survey, there was less discrepancy with most survey participants responding that there was sufficient funding for books.

In the intervention staff surveys conducted in the first year of the trial, there was no clear indication from librarians and literacy coordinators as to whether they had sufficient funding to purchase new books for AR, as presented in Figure **17**. Those who responded to this question in the first intervention staff survey (n = 31) were almost equally split between agreeing and disagreeing on whether there was sufficient funding. A similar split in responses was found in the second intervention staff survey (n = 24). Other anecdotal evidence also suggested that the cost of new books might have been a barrier: informants from two different schools stated that books were purchased from charity shops to reduce costs (INT 1; workshop 2). Informants from one additional school stated that they had borrowed books from the local public library when they did not have books in their own stock. However, in the endline survey, compared to the first and second staff surveys, the librarian and literacy coordinator respondents (n = 18) were more likely to disagree than agree that funding was insufficient.

Figure 17: Librarian and literacy coordinator respondents' perceptions on whether there was insufficient funding to purchase new books for AR

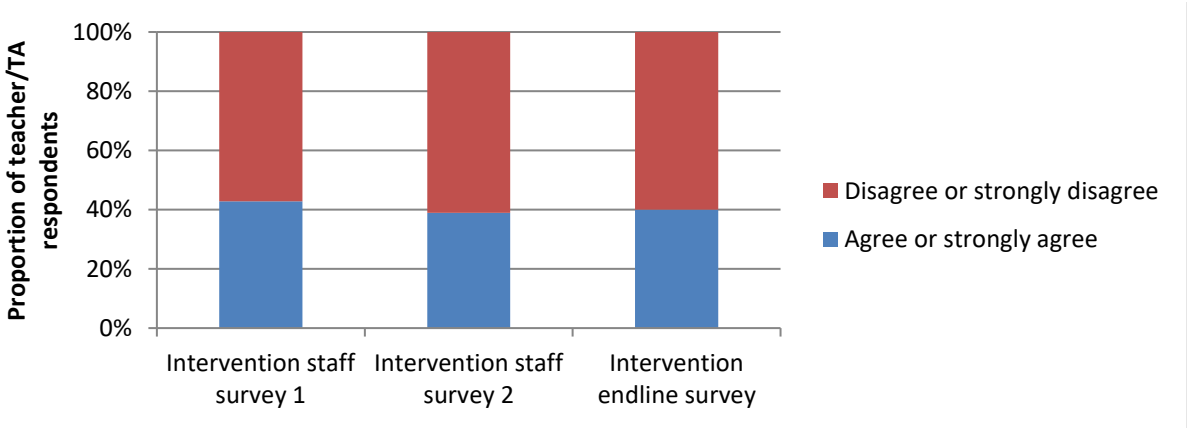


Intervention staff survey 1: n = 31; survey 2: n = 24; endline survey: n = 18.

A similar trend can be noted in responses from headteachers across the three staff surveys (survey 1, n = 36; survey 2: n = 26; endline survey: n = 14) whereby between the first and second surveys, the proportion of headteacher respondents who felt that their schools did not have sufficient resources (such as books or tablets) to use AR effectively fell from one-third to about one-ninth—though in the endline survey this proportion had risen to about a quarter.

Across the three intervention staff surveys, the majority of teachers and TA respondents (survey 1: n = 77; survey 2: n = 112; endline survey: n = 50) disagreed or strongly disagreed that it was difficult to request or buy additional books for the library, as presented in Figure **18**. In addition, the majority of respondents across these three surveys (teachers/TAs, headteachers, librarians, and literacy coordinators) disagreed or strongly disagreed that there were insufficient books or labelled books for pupils to read and progress. At baseline, approximately a quarter of treatment schools had reported having between 1,001 and 2,499 books, and just over another quarter reported having between 2,500 and 5,000 books in the school.

Figure 18: Teacher and TA respondents' perceptions on whether it was difficult to request or buy additional books for the library



Intervention staff survey 1: n = 77; survey 2: n = 112; endline survey: n = 50.

A majority of intervention schools had dedicated reading time for AR

Survey responses showed that most teacher, TA, and headteacher survey respondents reported having no barriers to providing dedicated reading time for AR during class time but that the total time spent per day on AR was slightly under the daily 30 minutes that was recommended. As reported in the baseline survey, the majority of schools already had dedicated reading time for the intervention cohorts in their timetable, as shown in Table **25**.

Table 25: Proportion of schools that reported having dedicated reading time prior to AR implementation

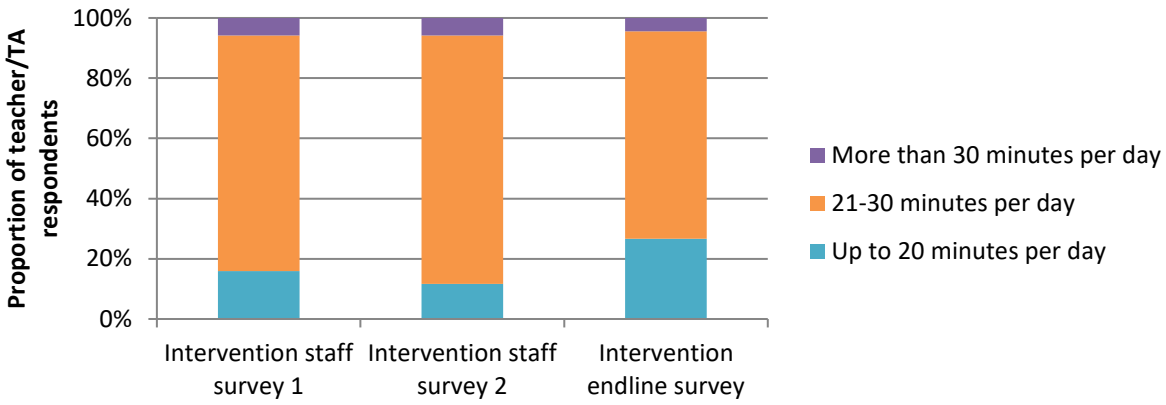
Year group	Treatment (n = 89)	Control (n = 135)
Year 4	88.8%	94.1%
Year 5	89.9%	93.3%

Source: baseline survey.

Despite there being dedicated time for reading prior to AR being introduced, survey respondents reported making changes to timetables to fit in AR. However, this was not perceived as a difficulty. In all surveys, almost all teachers and TAs reported making such changes and that there was daily protected reading time in their class that could be used for AR. Across these three surveys, the majority of teachers and TAs disagreed or strongly disagreed that difficulty in adding dedicated reading time to the timetable was a barrier to using AR. The majority of headteachers also disagreed or strongly disagreed that teachers had indicated it was difficult to add dedicated reading time to the timetable. Across the three surveys, the majority of teacher, TA, and headteacher respondents agreed or strongly agreed that the school encouraged teachers to add dedicated reading time to the existing timetable.

However, according to the majority of teachers and TAs in the surveys, pupils have an average of 21–30 minutes protected reading time per day, as shown in Figure **19**. This is slightly less time than the AR recommendation of at least 30 minutes, but given that RL suggest that an engaged reading time of at least 15 minutes is an indication of good intervention fidelity, the range of protected reading time reported here would allow for that amount of engaged reading time (please see Fidelity section, page 32).

Figure 19: Amount of protected reading time per day in intervention schools, as reported by teacher and TA respondents



Intervention staff survey 1: n = 77; survey 2: n = 112; endline survey: n = 50.

Across the three surveys, the majority of teacher and TA respondents disagreed or strongly disagreed that low pupil interest in reading time was a barrier to using AR.

Most teachers implementing AR in their classrooms made regular use of STAR Reading tests and quizzes

Across all three staff surveys, almost all teacher and TA respondents reported having used the STAR Reading test to test their pupils' reading levels. In addition, almost all teacher and TA respondents reported that pupils in their classes were taking AR quizzes once they had finished a book. Low pupil interest in the quizzes and tests were not reported to be a barrier to using AR, however, in one interview, a participant mentioned that insufficient computers, laptops, or tablets in the classroom could be a barrier since pupils then had to wait for equipment to be freed up before they could do their tests or quizzes (INT2).

Data from the workshops and interviews we conducted indicated that teachers felt that pupils liked STAR Reading testing and quizzing. For example:

> *'Yeah, it [doing more than one STAR Reading test per term] is useful but also the kids were just begging me to do the STAR testing to get their numbers up' (INT1).*

Two teachers reported that pupils were eager and motivated to take quizzes when they had finished a book. In one interview, staff reported that pupils would even seek to take quizzes for books they had read at home:

'The quizzes are done every time the child has completed a book […] they're pretty good [at taking a quiz when they have finished a book] […], when they've completed a book, they go and get a laptop independently, access it and complete the quiz' (INT 3).

'They absolutely love it. What I do before each day, I put a pile of iPads in front of the class and they don't have to tell me they finished the book. When they finish the book they can go and get the iPad and do the quiz and then go and get another book from the library and just crack on' (INT1).

'Children are now more inclined to read at home […] They know that they can then bring in the book from home and do a quiz in school […] if they are reading books from home they usually come in and ask me if there is a quiz to go with the book that they have, and I'll check it for them' (INT 3).

These interviewees felt that the quizzing element of AR had in particular helped pupils to develop a habit of reading:

'There's always the element of competition as to how well you do in the quizzes. […] I think they enjoy doing the quizzes, and seeing how well they do. And it's not necessarily a competition with other children, I think it's a competition with themselves […] and some of them get quite disappointed when they haven't done so well' (INT 3).

However, data from the workshops (one and two) suggested that a small subset of pupils may have struggled with the intervention because they did not perform well, and for these pupils the AR quizzes may have been a discouraging experience. Informants highlighted that it was then important to manage pupils so that reading did not feel like a punishment; in some matters this was simply a matter of ensuring that pupils were indeed reading books within their ZPD.

Overall, intervention participants reported making regular use of AR data

In the three intervention staff surveys, almost all teacher and TA respondents reported having looked at the reports generated by AR at least once, either on their own or with colleagues in their school. AR guidelines suggested teachers should look at STAR Reading data once a term to reassess children's ZPD and make adjustments if needed on the levels of the books read by pupils.

Data from the workshops indicated that staff (including teachers and the senior leadership team) often used AR reports, for instance to monitor and reward pupils' reading or to set overall reading targets for the school. Data from interviews suggested that the AR reports provide information that helps teachers identify children who may need further support (INT 2). Informants in one interview noted that the AR data provided useful information that often corroborated what staff already knew:

'[The data] highlights which children need intervention; quite often, it just confirms what we already know […] it also gives us a clearer indication of those that have made progress, and how much progress has been made. It also flags up the dips, and it's quite interesting to see that most of them don't make progress in a straight line […] For most of the children it is an upward trend, but it's quite a gradual slope […] over the school year' (INT 3).

Taken as a whole, IPE evidence on AR non-negotiables suggests that intervention schools were, in general, implementing AR as intended.

Pupil-level fidelity varied by AR benchmark

As also relates to the Fidelity subsection in the Impact evaluation, tracking data on pupil usage of AR was available via the dedicated online platform used by AR. This data included pupil-level information on the quizzes they had sat, the average percentage of correct responses achieved across sat quizzes each respective implementation term, as well as information from STAR Reading regarding their ZPD (updated when a new test was taken) and whether the books pupils were reading matched their ZPD range. We refer to these as 'AR benchmarks' for the sake of simplicity. As explained in the IPE methods section, matching with the attainment data was not possible. It was, however, possible to explore how pupils met each respective AR benchmark.

Across the data extracts from RL for 2016/2017 and 2017/2018, we observed AR usage data for 10,250 pupils—in both Year 4 and Year 5 and in both intervention and control schools, from the second year of the trial onwards.

For the purpose of this report, we only focus on the Year 5 pupils as Year 4 pupils will be included in the Addendum report.

For 2016/2017, we observed all relevant data for 5,901 pupils; 2,947 of these were in Year 5 in treatment schools (since control schools could not yet have received AR). Of these pupils, 85% sat quizzes in Term 1 and 50% reached the average percent correct benchmark in Term 1. For the same term, 43% reached the engaged reading time benchmark, and 89% read within their respective ZPD. In Term 2, 98% of pupils took quizzes and 51% of them reached the average percent correct benchmark. For this term, the engaged reading benchmark was met by 38% of pupils and 90% read

within their ZPD range. In Term 3, 100% of pupils in the data took quizzes and 49.6% of them reached the average percent correct benchmark. In relation to engaged reading time, 38% reached this benchmark with 91% reading within their ZPD.

For 2017/2018, we observed only 1,925 older cohort (now Year 6) treatment condition pupils with relevant data in the RL data extracts, showing a clear reduction in AR usage at the pupil level in this second year of the trial. Of these pupils, 87% are observed taking quizzes in Term 1 with 56% of these pupils reaching the average percent correct benchmark in Term 1. Also in this term, 42% of pupils reach the engaged reading time benchmark, with 94% of them reading within their ZPD. In Term 2, 89% of pupils took quizzes, with 62% of them meeting the average percent correct required level. For this term, the engaged reading time benchmark was met by 37% of pupils, and 91% read within their ZPD range. In Term 3, only 81% of pupils sat quizzes, reflecting evidence from the delivery team (as reported by schools) that usage of AR was lower around the SATs (in Term 3 of Year 6). Of these pupils, 61% met the average percent correct target, 30% reached the engaged reading time benchmark, and 92% read within their ZPD.

This suggests that overall pupil-level fidelity was very high in relation to reading with their respective zone of proximal development score fairly high in relation to the success rate in relation to quizzes, though not very high in relation to the engaged reading time that pupils exhibited. This is consistent with the school-level fidelity results presented in the Fidelity section that also included a measure of teacher responsiveness to these elements. The latter point, of lower than benchmarked engaged reading time, would suggest a potential partial explanation for the overall result of this trial whereby despite clear communication on the non-negotiable aspect of protected reading time and its fairly high level of implementation according to surveyed school staff, the realised *engaged* reading time may not have been high enough.

IPE: Perceived outcomes

Box 4: Perceived outcomes—key findings

- The vast majority of teachers, TAs, and librarians responding reported that participating in AR had helped pupils to increase their reading stamina and reading ability, and had improved pupils' attitudes to reading.
- Survey responses suggest teaching time for other subjects might have been affected by introducing AR, but the proportion of respondents agreeing with the statement reduced over time.
- A teacher interviewee suggested that with time, teachers began to see the value of the data generated by AR. This might have changed teacher views from seeing AR as 'taking away time from teaching' to a tool that facilitates teaching practice.

Survey respondents noted that they perceived AR had some positive outcomes on pupils

At the end of the first year of AR being implemented in intervention schools, we asked respondents how they perceived participating in AR had impacted pupils (intervention staff survey 2). The vast majority of teachers, TAs, and librarians responding reported that participating in AR had helped pupils to increase their reading stamina and reading ability, and had improved pupils' attitudes to reading. This finding was repeated in the endline survey at the end of the second year of AR in which almost all teacher, TA, and librarian respondents indicated that participating in AR in that year had helped pupils to increase their reading stamina and reading ability and had improved pupils' attitudes to reading. In both surveys, the majority of headteachers also reported that teachers in their schools had indicated these positive impacts.

Anecdotally, data from the workshops and interviews we conducted indicated that AR was perceived to have helped to increase pupils' enthusiasm for reading in that pupils no longer considered reading 'to be "work" anymore' (workshop 1). AR was also felt to have improved pupils' reading stamina (workshop 2), their reading ability (INT 1), and their vocabulary and decoding skills (workshop 2). One informant also reported that the school had received positive feedback from parents about their children's reading habits:

> *'They're now becoming more enthusiastic about reading […] Children have a reading record book that goes backwards and forwards to home […] there's been some positive comments from parents saying how nice it is that their child is reading voluntarily and that it's no longer a chore or a battle to get them to read' (INT 3).*

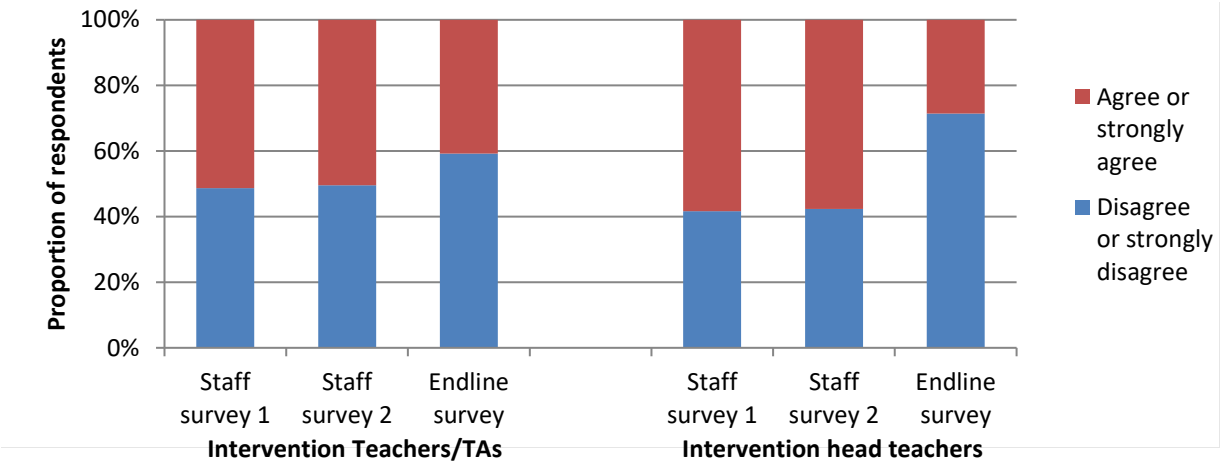
Interestingly, headteachers from control schools also reported observing positive impacts of having dedicated reading time at their schools on pupils' reading practices. In the control headteacher midline and endline surveys (those pertaining to the control cohort in both years of the trial), the majority of control group headteacher respondents reported positive impacts of dedicated reading time, with at least two-thirds reporting that teachers had indicated improvements in pupils' reading stamina, reading ability, and attitudes to reading.

The fact that both intervention participants and control headteachers reported seeing positive changes in pupils' reading practices and attitudes may serve as an indication that the situation in schools in the different intervention groups may not have been too different, however, we are mindful of a potential social desirability bias in response to this question. We provide further information on control schools in Control group activity' (page 49).

There were some perceived unintended consequences of using AR

In terms of unintended consequences, some survey and workshop respondents reported that using AR on a daily basis meant that time spent on teaching other subjects was reduced (intervention staff surveys 1 and 2, endline survey, workshop 2), but the proportion of survey respondents agreeing with the statement reduced over time. As presented in Figure **20**, slightly more than half of teachers and TAs in both intervention staff surveys agreed or strongly agreed that teaching time for other subjects had been reduced, while more than half of headteachers reported that teachers in their schools had indicated this. In the endline survey, this pattern had reversed slightly such that a smaller proportion of teacher or TA respondents (about two-fifths) agreed or strongly agreed that teaching time for other subjects had been reduced as a result of using AR. Less than a third of headteacher respondents in the endline survey reported that teachers had indicated that time spent on other subjects had been reduced.

Figure 20: Teacher, TA, and headteacher respondents' perceptions on whether using AR on a daily basis meant that teaching time for other subjects was reduced



Intervention staff survey 1: teachers and TAs, n = 76; headteachers, n = 36. Survey 2: teachers and TAs, n = 113; headteachers, n = 26. Endline survey: teachers and TAs, n = 49; headteachers, n = 14.

The change in the negative perception that AR takes away time from other subjects (captured through the surveys) is reflected in one informant's view that engaging in AR generated data that could benefit teachers and their teaching:

> *'I think initially some of the staff thought it was taking away teaching time, but then once they got into […] realising the data can then be extracted, and how useful that is, I think they sort of turned around and begun to realise the benefit of having that dedicated time' (INT 3).*

It is worth noting that while the surveys did not provide information about how much time was taken away from other subjects (in the regular timetable) or which subjects were the ones affected, in the workshops participants were asked whether time from other 'core' subjects such as English, maths, or science were affected. They noted that this was not the case and that time was taken away from subjects such as music or art. During the workshops it was not specified how much time was taken away from these subjects.

It should be noted that it was not possible to independently corroborate how the use of AR benefited teaching (as noted in the interview quote presented above). Also, the evaluation did not explore if pupil outcomes in other subjects had been impacted as this was out of scope.

Formative findings

Box 5: IPE formative findings summary

- IPE data suggests that:
  - Improvements can be made in how information is shared with teachers implementing AR. Survey and interview respondents noted that sometimes too much information was given, which was difficult to navigate.
    - A potential solution for this is to provide a document that summarises all key information and that provides step-by-step guidance for implementation.
    - One-to-one support provided through the online AR chat was identified as useful by some participants, while one survey respondent noted wanting more individualised support for their school.
  - IPE data also suggests that the way AR translates from a U.S. context to an English or U.K. setting could potentially be further improved. IPE participants commented that AR used old National Curriculum scores, it may too strongly focus on quizzing, and that it misses other important elements of reading (such as inference making).

During training observations, workshops, and interviews, some teachers, literacy coordinators, and librarians commented that an element of the intervention that could be improved is around the amount of information given and the format in which it is provided.

Some respondents felt that the onsite training was too heavily loaded with information (staff intervention surveys, endline survey, INT1). One of the interviewed teachers noted that the onsite training was helpful but they felt 'overwhelmed' with the large amount of information available, there was 'too much' and they 'felt confused' (INT1); nonetheless, to clarify their doubts the teacher used the online AR chat and found this helpful. Workshop participants also noted finding the online chat useful (workshops 1 and 2). In an open-ended survey response, another teacher expressed wanting to receive more individualised support from the AR RL team at the implementation stage, noting that it would be useful to have 'phone calls to catch up on progress and tell us what we are supposed to be doing every few weeks'. These two views suggest that individualised support was considered to be valuable by participants. In the case of the latter suggestion, the question arises as to whether it would feasible to provide more targeted support to each individual school (from a cost and capacity perspective of the RL team). However, the former reports that the online chat provides useful support suggests that this is a feature worth maintaining and is a good alternative should more individualised support not be feasible. Further to this, other participants provided suggestions on how to improve the way information is conveyed and made available to teachers and staff implementing AR. These are outlined below.

Respondents to the online surveys identified potential solutions to help them digest the large amount of information provided during the training. Some suggestions were that it would be helpful to provide participants with 'a training booklet going through each process step by step'. Another survey respondent noted that having 'a comprehensive timeline so that the process is more clearly mapped out from the start' would also be beneficial. Moreover, a different respondent noted that while 'there is a wealth of training and support materials available, they could perhaps be a little more accessible and user friendly'. These responses suggest that providing a single 'to go' document that captures the main information would help teachers using AR have greater clarity of the necessary steps that need to be taken, all resources that are available, and how best to make use of them. It is worth noting that in the RL website there is a training timeline available, but these open-ended survey responses suggest that not all AR users rely on this resource.

IPE data also suggests that the way AR translates from a U.S. context to an English or U.K. setting could potentially be further improved. During interviews and workshops some teachers fed back that in their view AR is too much about quizzing (INT2, workshop 2). Also, some workshop participants noted that AR lacks elements such as a focus on guided reading and the 'inference' component of reading. Furthermore, during one of the onsite training observations and in the second intervention staff survey, participants noted that the diagnostic reports produced by the AR site used the old National Curriculum levels to score pupil progress. When this issue was raised at one of the observed onsite training sessions, the trainer reassured participants that AR would transition to using the new National Curriculum levels. Nonetheless, this problem was raised again in the second staff survey (which was administered nearly a year after the onsite training took place). While this issue on AR's alignment with the U.K. context may be symptomatic of the fast changing English education system and the pace at which RL can adapt its software, it can also suggest that more attention may need to be placed on ensuring the appropriate translation of AR into the an English or U.K. context. This mismatch may also serve as a potential explanation for the null effects found by the impact analysis of this evaluation.

Control group activity

Box 6: Control group activity—key findings

- Control headteacher survey respondents reported investing in books during the trial, similar to intervention schools where books were also purchased as part of setting-up AR. However, IPE data is unable to compare whether intervention schools were acquiring more or fewer books than intervention schools.
- Control headteachers reported in the surveys that their schools were using reading schemes and reading interventions throughout the duration of the trial. Some of the reported reading interventions have evidence to suggest they are effective in improving pupil reading outcomes. Their use in control schools might serve to partially explain the null effect results found in the impact evaluation, especially if every control school were implementing interventions that have proven effective.
- There were few instances of contamination identified throughout the trial, and none that extended beyond the use of the STAR test with pupils. This suggest that contamination was minimal and likely not a factor influencing trial results.

The waitlist design of the trial meant that schools randomised into the control group would continue with 'business as usual' in the school year 2016/2017 and would not implement AR. Instead, pupils in Year 3 and Year 4 received AR in the school year 2017/2018. This section reports on what happened in control groups in the first two years of the intervention.

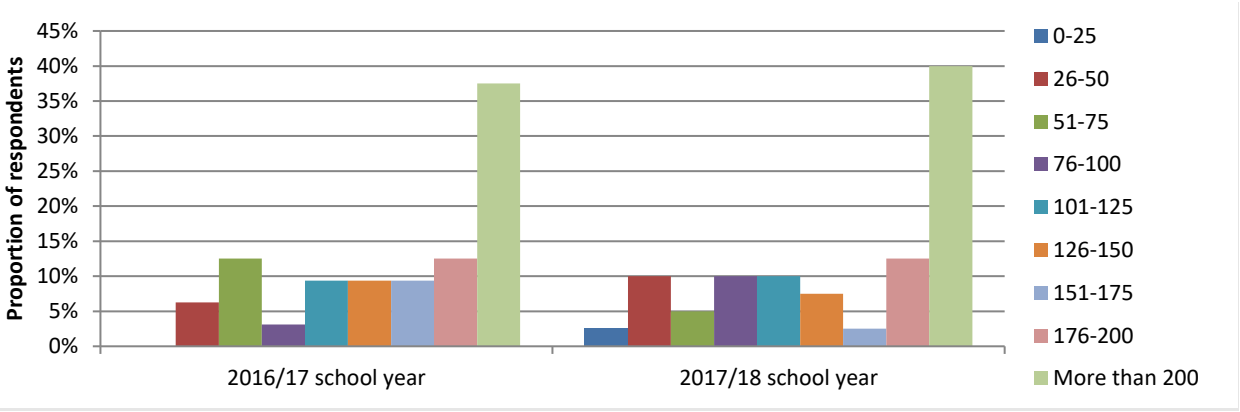
The findings in this section are informed by administrative data from RL on the use of AR in control schools as well as by the three surveys conducted with staff in control schools: the baseline survey conducted in February and March 2016, the control group headteacher midline survey conducted in February and March 2018, and the control group headteacher endline survey conducted in June and July 2018. The baseline survey, conducted at the point of recruitment, asked schools about their then current practices for Year 4 and 5 pupils. The midline survey asked headteachers to report on the Year 4 and 5 cohorts in the 2016/2017 school year (that is, the control group in the first year of the trial). The endline survey asked headteachers to report on the Year 5 and Year 6 cohorts in the 2017/2018 school year (that is, the control group in the second year of the trial).

Below we present what 'business as usual' looked like in terms of school investments in book stock, dedicated reading time for pupils, literacy activities implemented in the school, and the perceived impact of these activities. We also report on 'contamination', in other words, whether the control cohorts accessed AR resources. Unusually for an education trial, contamination data was both available and of very high quality as it was based on RL administrative data derived from the online STAR Reading test and quizzing system. Upon notification that a control-group pupil was using AR, RL intervened promptly to stop any contamination in the very few situations where control schools were observed to have used STAR Reading.

Investments in book stock were comparable to treatment schools

The majority of control schools reported having between 1,001–2,499 or 2,500–5,000 books at baseline (similar to the reports of intervention control headteachers at baseline). While these are fairly wide margins, the pattern remained the same during the first two years of the trial. Most of the headteachers also reported buying new books in each of these years: 86% (32 out of 37) of headteacher respondents reported that their school had bought books in the 2016/2017 school year while 95% (40 out of 42) reported that their school had bought books in the 2017/2018 school year. In each of these years, more than a third of headteacher respondents reported buying at least 200 books (see Figure 21 below).

Figure 21: Control schools—number of books purchased in 2016/2017 and 2017/2018

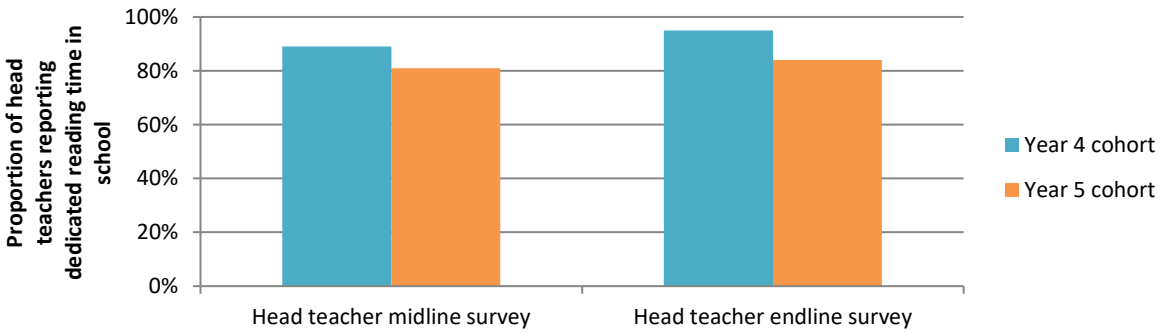


Source: control headteacher midline survey, n = 37 (2016/2017); control headteacher endline survey, n = 42 (2017/2018).

Dedicated reading time and its perceived impact

Based on the surveys, it appeared that the practice of dedicated reading time was also common in control schools (see Figure **22**). As reported above, at baseline, over 93% of control schools reported having dedicated reading time in their timetables for Year 4 and Year 5 pupils (compared to about 90% of treatment schools). During the two years of the intervention, a large proportion of the control group cohorts experienced dedicated reading time in their schools, with most having between one to two hours per week of reading time, which is also broadly in line with the reading time in intervention schools (see Figure **23**). While the National Curriculum does not provide a recommendation of how much time should be dedicated to reading in schools for Key Stage 1 and Key Stage 2 pupils, it does, however, emphasise the importance of developing pupil's reading skills: 'Schools should do everything to promote wider reading. They should provide library facilities and set ambitious expectations for reading at home' (DfE, 2013, p.10). This suggests that control schools also had motivations to encourage reading despite not using AR, and potentially means that there was a lack of programme differentiation between 'business as usual' and AR, with practices in intervention and control schools not being sufficiently distinctive. This may serve as an explanation for the null effects of AR identified in the impact evaluation.

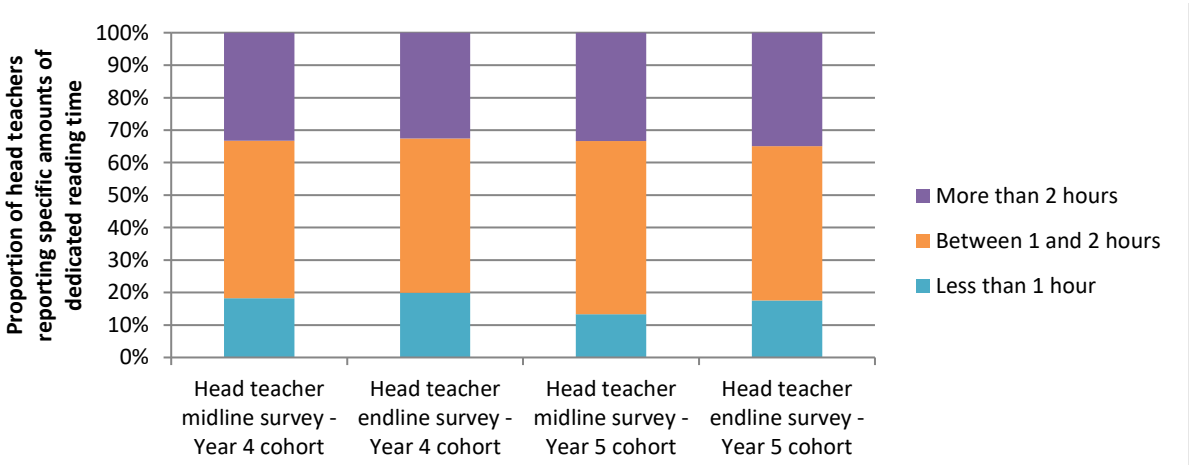
Figure 22: Control schools—prevalence of dedicated reading time for Year 4 and Year 5 pupils



Note: Control

Source: headteacher midline survey 1, Year 4: n = 37; Year 5: n = 37; control headteacher endline survey, Year 4: n = 42; Year 5: n = 42.

Figure 23: Control schools—weekly reading time for Year 4 and Year 5 pupils



Source: control headteacher midline survey 1, Year 4: n = 37; Year 5: n = 37; control headteacher endline survey, Year 4: n = 42; Year 5: n = 42.

Control school literacy activities and their perceived impact on pupils

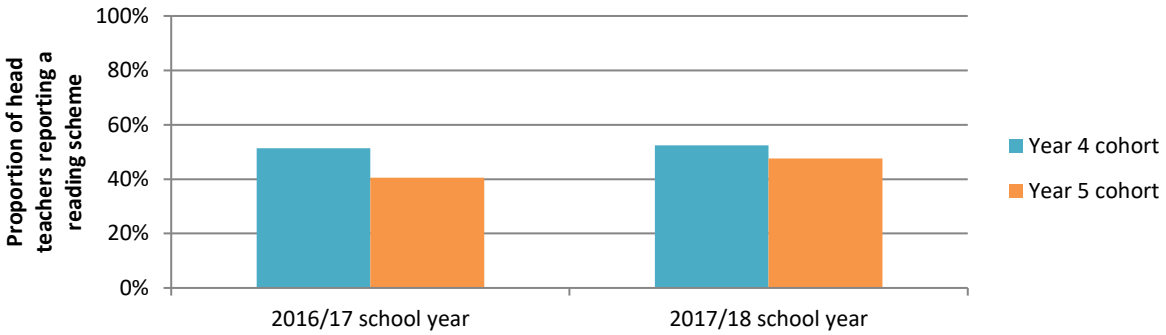
Control school headteachers responding to the midline and endline surveys were also asked if their schools implemented reading schemes or reading interventions for control group cohorts (that is, Year 4 and Year 5 in 2016/2017, Year 5 and Year 6 in 2017/2018) and how teachers in their schools perceived the impacts of these activities. AR is a reading intervention. Reading schemes and reading interventions can be defined thus:

- A **reading scheme** is characterised by the provision of reading material that is structured carefully to support teaching in the classroom and to build pupils' reading confidence when reading at home. They are non-competitive and rest on the principle that children learn at their own pace.
- A **reading intervention** is a programme seeking to improve reading skills with clear stages and attainable goals. Reading interventions may focus on phonics, comprehension, vocabulary, or other aspects of reading.

Reading schemes were prevalent in control schools

More than half of headteacher respondents reported that their school had implemented a reading scheme for the Year 4 control cohort over two years, while at least two-fifths of respondents indicated that their school had implemented a reading scheme for the Year 5 control cohort (see Figure 24). The most common reading scheme implemented was the Oxford Reading Tree, a programme popular with primary schools in the UK, which focuses on reading for pleasure, synthetic phonics and reading at appropriate levels.[15] Other reading schemes mentioned include the Bug Club, Big Cat and Collins books.

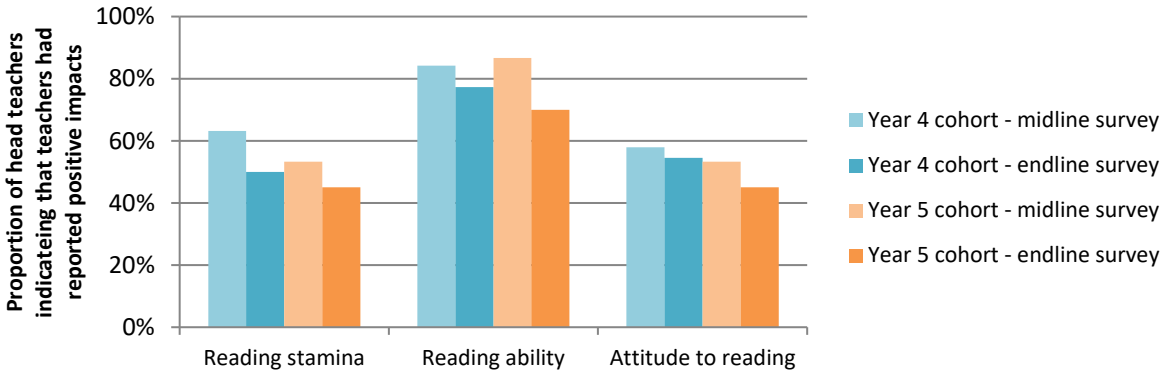
Figure 24: Control schools—prevalence of reading schemes for Year 4 and Year 5



Source: control headteacher midline survey 1, Year 4: n = 37; Year 5: n = 37; control headteacher endline survey, Year 4: n = 42; Year 5: n = 42.

Headteachers whose schools implemented a reading scheme reported that teachers were most likely to have indicated that the reading scheme had had a positive impact on pupils' reading ability, compared to their reading stamina or attitudes to reading. Progress was also perceived in these other outcomes by a majority of teachers at all measurement points.

Figure 25: Control schools—perceived impact of Year 4 and Year 5 reading schemes



Source: control headteacher midline survey 1, Year 4: n = 33; Year 5: n = 30; control headteacher endline survey, Year 4: n = 22; Year 5: n = 20.

Reading interventions were used in control schools

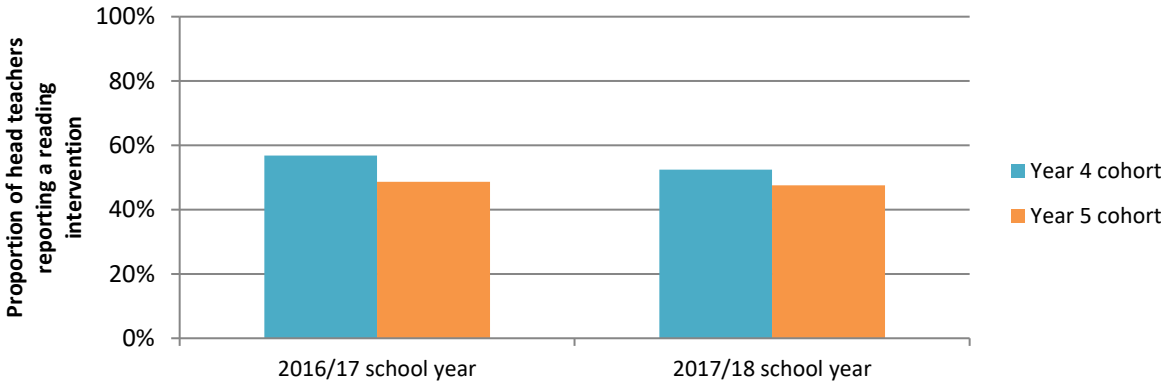
In addition to reading schemes, more than half of headteacher respondents reported that their school had implemented a reading intervention for the Year 4 control cohort in the previous two years, while almost half of respondents indicated that their school had implemented a reading intervention for the Year 5 control cohort (as presented in Figure **26**). Headteachers reported that they used reading interventions such as Lexia,[16] Catch Up Literacy,[17] Every Child a Reader,[18] Read Write Inc. Phonics,[19] and Inference Training.[20]

It should be noted that when asked about what reading interventions were being implemented for Year 4 pupils in control schools in 2016/2017 (and to a lesser extent, when asked about reading schemes in the same context), some respondents from control schools reported using Accelerated Reader. We believe that this is more likely to be a response error than contamination because (1) the question was being asked in the school year 2017/2018 and respondents may have mistakenly thought they were being asked about the current school year (and in that year, Year

[15] https://www.oxfordowl.co.uk/for-home/find-a-book/oxford-reading-tree-levels/
[16] https://www.lexiauk.co.uk/cohorts/primary-schools/
[17] https://www.catchup.org/interventions/literacy.php
[18] https://www.gov.uk/government/publications/evaluation-of-every-child-a-reader-ecar
[19] https://ruthmiskin.com/en/programmes/phonics/
[20] http://www.targetliteracy.co.uk/comprehension

3 and Year 4 pupils in control schools should have received AR) and (2) none of the respondents reported that Accelerated Reader was being used with Year 5 control pupils, meaning they could have been using AR with younger pupils.

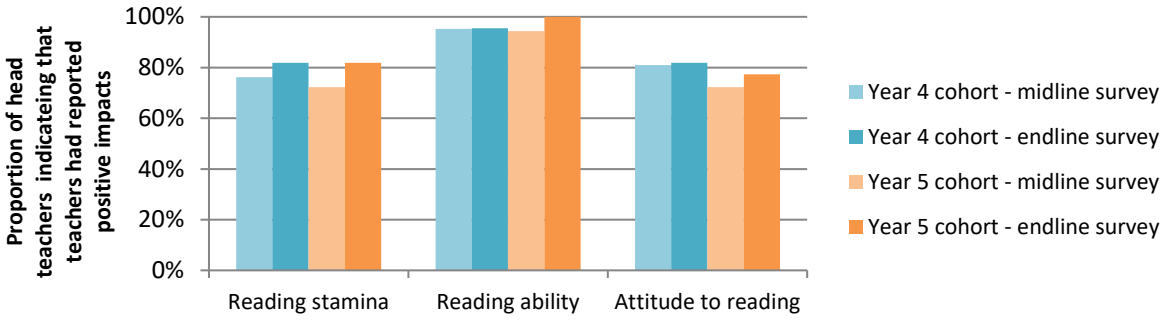
Figure 26: Control schools—prevalence of reading interventions for Year 4 and Year 5 pupils in control schools, as reported by head teacher respondents



Source: control headteacher midline survey 1, Year 4: n = 37; Year 5: n = 37; control headteacher endline survey, Year 4: n = 42; Year 5: n = 42.

Similar to reports from control headteachers on the effects of reading schemes, a large proportion of headteachers whose schools implemented a reading *intervention* reported that teachers indicated that the reading intervention had had positive impacts on pupils' reading ability, reading stamina, and attitudes towards reading.

Figure 27: Control schools—perceived impact of Year 4 and Year 5 reading interventions for Year 4 and Year 5 pupils in control schools, as reported by head teacher respondents



Note: control headteacher midline survey 1, Year 4: n = 21; Year 5: n = 18; control headteacher endline survey, Year 4: n = 22; Year 5: n = 22.

Implications of control schools implementing other reading interventions

In what follows we briefly discuss the other reading interventions reportedly being used in control schools and compare them to AR. We also explore the evidence available on their effectiveness.

Of the five reading interventions reportedly being used (based on survey responses), three have been found to have some positive effects on pupil reading outcomes and one is currently being evaluated in a U.K. setting (while one currently has no evaluations available).

- *Lexia* is a computer-based reading programme that gives pupils practice in basic reading skills and was evaluated in three studies that found that the intervention has potential positive effects on alphabetics and comprehension (What Works Clearing House, 2009).
- *Catch Up Literacy* is delivered by teaching assistants and consists of two 15-minute sessions per week with pupils to improve their reading. This intervention has been evaluated by the EEF, first in an efficacy trial and later in an effectiveness trial. The former found positive effects for reading progress for FSM pupils while the latter (larger) evaluation did not find positive effects on reading comprehension (EEF, 2019a).
- *Every Child a Reader* (ECaR) is a layered intervention made of three components: quality class-based teaching, small-group support, and one-to-one support. Each component focuses on improving reading through different

levels of targeted support and includes 'Reading Recovery21' daily half-hour sessions for pupils needing most support. These sessions aim to equip pupils to become independent readers. ECaR has also been evaluated in a U.K. setting. The independent evaluation commissioned by the Department for Education found significant positive effects of the intervention on pupils' reading ability using a difference-in-difference methodology (DfE, 2013; also note the references therein to Reading Recovery and ECaR evaluations).

- *Read Write Inc. Phonics* is a CPD and teaching programme that uses phonics to improve attainment at KS1 reading and is currently being evaluated by the EEF (EEF, 2019b).
- Finally, *Inference Training* is a widely used group intervention that aims to boost reading comprehension and enjoyment in reading. To the knowledge of the evaluation team, this intervention has not been evaluated yet (Institute for Effective Education, 2019).

Beyond the existing evidence, limited as it may be, when we looked at the format or goals of these interventions we observed some similarities with AR. Like AR, *Lexia* is a computerised intervention that also focuses on improving reading comprehension while *ECaR* seeks to foster independent reading. *Inference Training* has not yet been evaluated but focuses on improving reading enjoyment. Should all of these interventions actually be effective, this would serve to shrink the difference between treatment and control schools.

We do not have systematic data across all control schools, but it seems reasonable to infer that many primary schools in the study would be using a *reading scheme* at least, even if stopping short of using a branded intervention (approximately 50% of heads reported one of these were in use). But, this means that many schools may not be using reading schemes *or* interventions and yet we still do not observe differences between treatment and control groups. This leads us to infer that whatever routine practice is in primary schools—that is, whatever 'business as usual' activities actually were—was sufficient to equivalise the older cohort of treatment and control groups at the end of KS2. We therefore infer that compared to 'business as usual', there was not enough activity that was 'unique' to AR to produce measurable differences in reading outcomes. This should not be viewed as a failure of AR but as a success story for the teaching profession.

IPE data suggests that there were low contamination levels in control schools

The advantage of implementing an intervention that used an online platform was that Renaissance Learning could easily track instances of contamination (that is, when control group pupils were using AR).

During the first year of the trial (the 2016/2017 school year), there was only one instance of contamination identified. In the second year of the trial (the 2017/2018), Renaissance Learning identified three instances of potential contamination which were quickly resolved. In all of these cases, Renaissance Learning was made aware of the contamination after a small number of pupils in the control school took the STAR Reading test (though none of them had accessed other aspects of AR such as the quizzes). Staff at Renaissance Learning then got in touch with the schools in question and clarified the design of the trial and reminded staff that pupils in Year 5 and Year 6 (appropriate to the second year of the trial) were not meant to access AR. In the endline survey for control schools, we asked respondents which year groups they had used AR with in the 2017/2018 school year. Responses from 42 participants suggested that the waitlist design had mostly been implemented as planned, though two respondents stated that AR had been used with Year 2 pupils (with no implications for this trial regarding spillover—see below).

Overall, therefore, using the above, the level of contamination between intervention and control schools in this trial was minimal. This was due to tight controls and monitoring by the delivery team who notified the evaluation team whenever such instances occurred and also contacted schools to ensure that any potentially contaminating activity would be stopped.

Potential spillover effects were also small

This does not entirely remove the possibility of slight spillover effects from the second year of the trial onwards, where control schools (originally on the waitlist) had access to the intervention with young year groups only. Spillover effects that could have been observed might have related to the restructuring of the libraries in control schools or, potentially, to more IT resource becoming available to the control pupils by virtue of it being used originally for the waitlist groups. As stated above, the IPE found no evidence that direct contamination was an issue.

Again, the same robust monitoring processes as above were applied. The IPE has not collected evidence on this aspect over what has been reported above, but all other pieces of evidence point to limited spillover effects.

[21] 'Reading Recovery' has been recognized by the Early Intervention Foundation as being an effective reading intervention: **https://www.eif.org.uk/files/pdf/realising-the-potential-of-early-intervention.pdf**

Conclusion

Key conclusions
1. Children who started Accelerated Reader in Year 5 made, on average, no additional progress in reading compared to children in the comparison schools. This result has a very high security rating.
2. Children eligible for free school meals who started Accelerated Reader in Year 5 made, on average, no additional progress in reading compared to FSM children in the comparison schools. However, this result has high statistical uncertainty.
3. Data during the first year of implementation indicated that AR was implemented as intended in intervention schools. During the additional second year of the trial one third of pupils were no longer accessing AR. Findings suggest that higher fidelity was not associated with better outcomes.
4. The implementation and process evaluation suggests that most of the 'business as usual' comparison schools had similar amounts of dedicated reading time and some used other evidence-based reading schemes and reading interventions.
5. AR was very well received by the vast majority of teachers, teaching assistants, and librarians who perceived positive impacts on pupil reading ability, reading stamina, and attitudes.

Interpretation

This evaluation is the largest independent evaluation of Accelerated Reader ever undertaken using a randomised controlled trial. The evaluation aimed to test the effectiveness of AR at scale, in a large number of schools, and under normal school conditions. Additionally, this new RCT focused on younger children in primary school, Year 4 and Year 5, and was powered to detect an impact on disadvantaged children—those eligible for free school meals. Table **2** above lists the key differences between this, and the previous efficacy trial (Gorard et al., 2015).

Through the implementation process evaluation of this trial it became evident that, overall, AR was well received by trial participants. The training was perceived to be useful and some participants reported that their personal experience suggested that AR had had a positive impact on pupils' reading ability and attitudes towards reading. Our process and implementation evaluation activities, supported by routine data collected by Renaissance Learning, found that implementation fidelity was high. Specifically, our data showed that AR was implemented as intended in the vast majority of schools, with teachers and those implementing AR being clear about AR's 'non-negotiables' (reading time, regular quizzing, ensuring an appropriate book stock, and regularly using AR data).

Yet we did not find support for a causal effect of AR on pupil reading outcomes at KS2 compared to pupils in business-as-usual schools. In short, pupils in control schools did just as well as pupils in treatment schools. Similarly, we found no evidence that AR having a differentially positive impact on FSM pupils' reading outcomes at KS2 compared to FSM pupils in control schools. Furthermore, our quantitative assessment of fidelity also showed no difference between reading outcomes of pupils from treatment and control schools at differing levels of fidelity. This latter result is likely due to the overall high implementation fidelity: the actually variation in fidelity was small as many schools implemented AR as expected.

Given the high levels of implementation fidelity, it is our interpretation that business as usual practices in schools acted to equivalise treatment and control pupils. As we have set out, we do not have complete information on the exact nature of 'business as usual' in control schools but what we do know is that roughly half of control schools were using reading schemes or other reading interventions. While these reading interventions and schemes varied in their intensity and focus, this suggests that business-as-usual is not represented by the absence of an intervention but, in many cases, it consists of *other* interventions, at least some of which have evidence of effectiveness or are being evaluated as promising. As such, the overall, and subgroup, null effect identified by the analysis is perhaps unsurprising. This mirrors previous research that suggests a very high incidence of such a finding in education trials (Lortie-Forgues and Inglis, 2019).

That said, we do not think that the full explanation of the null effect results from the use of other interventions. Not all headteachers in control schools reported using reading schemes or interventions; the lack of 100% coverage of these leaves a 'gap' to be filled. Rather than write this gap off by assuming that heads *not* responding to surveys already use highly effective reading interventions, or heads who did respond did not do so accurately (both possible but not *probable*) our interpretation is that *teachers' routine practice in control primary schools was sufficient to match any effects of AR at KS2 in the older cohort of pupils*.

We view our results as suggesting that much greater attention should be paid to what 'business as usual' for reading development consists of in primary schools in Years 5 and 6 (and perhaps Year 4 as well depending on the follow-up results). We take this view because if a highly professionalised, well-implemented, and widely used intervention such as AR does not have an impact on reading outcomes then other, similar, interventions aimed at these year groups also may not. We also infer that less well-developed interventions may struggle to compete with routine practice, which in turn suggests that the EEF and other funders could afford to be more 'choosy' about which reading interventions to evaluate. Following the line of argument made in Lortie-Forgues and Inglis (2019), reading interventions with weak theoretical or empirical underpinnings may be doomed to fail against routine practice. This is particularly pertinent when we recall that the actual comparison being made in this trial was not 'treatment' versus 'control' but 'business as usual + treatment' versus 'business as usual'. In short, if routine practice is effective, we are (at best) searching for marginal gains, which we did not find here. In the face of these null results, we believe the focus should be on lauding and better understanding routine teaching practice—and *teachers*—rather than concentrating on the apparent 'failure' of one approach or another (least of all because we have to acknowledge that our results may not replicate in another sample, location, or time).

The per-pupil-per-year cost to schools as calculated in this trial (over three years of AR) is £48. This includes the costs for annual licenses, start-up costs (one-off training and the purchasing of books), and running costs. Compared to two other EEF-evaluated reading interventions, AR's per-pupil cost is higher. For example, Abracadabra is estimated to cost £9 per pupil over three years while Catch-Up Literacy costs £30 per pupil.[22] However, it should be noted that the cost estimate in this evaluation is for schools implementing AR in two year groups. The per-pupil-per-year cost of AR would reduce if a school obtained an AR license for more pupils (that is, additional year groups). Furthermore, the previous EEF-funded AR evaluation estimated a considerably lower cost of AR (£9 per pupil for 50 pupils; Gorard, et al., 2015).[23] However, this estimate only considered the license fee and the one day of teacher training; no costs for purchasing books were included, which in this former evaluation were considered to be part of marginal costs. Furthermore, the earlier AR trial was conducted in secondary schools—and it is worth noting that the cost of AR licenses varies for primary and secondary schools. Despite the higher per-pupil costs of AR estimated in this trial (compared to the previous trial), under EEF cost ratings rule this is still a very low cost intervention.

Limitations

We acknowledge several limitations of the trial. First, the trial underwent a redesign as a result of the issue with the initial Year 4 outcome measure. The redesign was approved by the evaluators, the EEF, and the developers. Our concern about the redesign was lessened because during the project inception meetings the developers referred to the fact that visible effects of AR might take time to emerge as the intervention requires schools, teachers, and pupils to change their behaviour, not least in relation to reading time and choice of books.

Second, the measure used in the fidelity analysis is unusual in that it is a continuous measure without a cut-off point below which a school is considered as having failed to implement the intervention as intended. The derived continuous measure, however, makes use of extensive administrative data supplied by the developer (from automated online systems) and has been agreed on by the EEF and the evaluators, and has been treated analytically in a robust manner that allows for a conclusion as to the relationship between incremental increases in fidelity and the pupil reading primary outcome.

Third, the implementation and process evaluation suffered from limited engagement. Ideally, we would have liked to obtain responses from all control schools as to their practices in relation to the use of other reading schemes or interventions. This would allow for more definitive conclusions as to what business as usual consisted of in control schools. Related, we suggest that future trial process evaluations focus on control conditions to a much greater extent than they do presently, particularly in light of the small average effect sizes in education trials. That said, this trial is special from the perspective of the substantial amount of administrative data on treatment schools, which allows for detailed analyses of fidelity, but even in the absence of this data, more information on control schools would provide further support to conclusions from future trials.

Finally, while the trial was well powered to provide a precise estimate of the impact of the intervention on pupil outcomes, the use of the KS2 reading outcome measure alone—without a secondary outcome measure—may not capture all the potential effects of the intervention. More specifically, had the original Year 4 outcome measure provided usable data, we would have been able to examine the effect of the intervention on an outcome that included word comprehension. Secondary outcomes, such as attitude to reading, reading habit formation, and enjoyment of reading, were also not included in the trial, partly due to the scale of the trial and budgetary constraints on the evaluation. However, such outcomes formed part of the assumed outcomes of the intervention, alongside reading ability. Indeed, the original trial design, whereby the KS2 reading score was going to be used as an outcome for the Year 5 cohort (as presented in this report), suggests that the KS2 reading outcome is indeed an outcome relevant to the intervention. More broadly, though,

[22] https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/abracadabra-abra-pilot/ and https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/catch-up-literacy/
[23]https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/accelerated-reader/

this raises issues in relation to the continued use of high stakes testing as outcome measures as opposed to other outcomes that are closer to the intended aims of the intervention being tested. Evaluators and the EEF need to find a better balance between 'treatment inherent' measures that we expect the intervention to 'move' because they are the direct focus on the intervention and more general measures that speak to the external validity of the trial.[24] This may be achieved, in part, by a greater interest on measuring treatment mechanisms as part of trials, or indeed through running pure mechanism experiments (Ludwig et al., 2011).

Strengths of trial

This trial is an independent evaluation of AR involving the developers at relevant phases (such as the theory of change) but not in the randomisation, outcome data collection, or analysis—all aspects critical to minimising trial bias. The treatment and control groups were balanced at baseline, attrition from the trial was minimal, and the proportion of missing data for the analytical sample was very small (small enough that no missing data imputation was necessary).

The trial is well powered, both for the overall analysis and, unusually for the subgroup analysis, with minimum detectable effect sizes at analysis stage actually better than at the design stage because the between-school variation was smaller than planned for. The large sample sizes in terms of both pupils and schools allow for high levels of precision in estimating treatment effects on pupil outcomes.

As a highly professional developer, RL records (and has shared) very good data on fidelity through its online system (again, minimising the risk of bias); this has been used in the fidelity analyses reported previously.

Finally, on the issue of generalisability, we have provided in Table **11** a descriptive analysis of the analytical sample at the pupil level. We believe this provides schools useful information on the rate of Ever-6 FSM eligibility in the trial analysis (37%), as well as information on gender and pupils' ethnicities that allows them to ascertain the extent to which they serve similar pupils. This does not amount to generalisation, but to contextualised extrapolation.

Future research and publications

As part of the one-year trial extension, an addendum report will be published at the end of 2019 to report on KS2 outcome data for pupils that were in the Year 4 cohort at the beginning of the trial. This report will provide outcome analyses for the younger cohort of pupils and will assess whether AR has effects that persist over a longer period of time (under the updated trial design, the outcome assessment on KS2 reading will come three years after the start of the AR implementation). The addendum report will also present additional fidelity analyses that make use of the administrative data collected by RL that capture fidelity at both pupil and teacher level, in both the first and second year of the trial.

Finally, it would be interesting to assess whether AR may have positive longer-term effects for pupils that participated in this study, despite AR seemingly not having immediate effects based on this trial's results. Future research can focus on KS4 outcomes for English but also more broadly on outcomes for other subjects such as maths or sciences. Evidence shows that strong literacy skills are crucial for later performance (Cunningham and Stanovich, 1997) and attainment in other subjects (Hall et al., 2015). It would therefore be interesting to explore (1) if AR has longer term positive effects on pupil attainment and (2) what, if any, spillover effects AR may have on outcomes in subjects other than English or literacy in the medium to long term.

The findings above, in particular those pertaining to control schools' use of other reading interventions, provide good reason to undertake another study looking at how AR's effectiveness may compare to that of other reading interventions. Therefore, we recommend that the EEF consider funding a superiority or non-inferiority trial between AR and other reading interventions (Lesaffre, 2008). This relates to the plausible explanation for the null results in this trial whereby they are linked to control schools already using other reading interventions to improve pupil reading outcomes. The implementation and process evaluation suggests that this is widespread. If this were to be considered common practice across all U.K. schools, it would therefore be informative to compare the relative effectiveness of different interventions in an attempt to aid schools in making an evidence-based decision as to which interventions to invest in.

Furthermore, since this is the first large scale RCT study involving 12,000 pupils in an U.K. setting, the evaluation team will publish an academic journal article reporting on the Year 4 and the Year 5 analysis together, and include the fidelity analysis as undertaken in this report. Pre-specified analysis will form the key results to be reported for the purposes of this publication, including the subgroup analysis of pupils eligible for free school meals. One potential analysis for this publication is to explore whether, for the Year 4 cohort, there are any spillover effects into other outcomes (for example, maths). Further, given that we will eventually be able to request data from KS1 through to KS4, it may be possible to assess the trajectories of pupils in the trial across their schooling career to assess whether there are so-called 'sleeper' effects (Barnett, 2011; White et al., 2007) of AR on later outcomes.

[24] Though we note that narrower, more specific outcome measures may be related with less educational value from the perspective of later outcomes.

References

Allen, R., Jerrim, J., Parameshwaran, M. and Thompson, D. (2018) 'Properties of Commercial Tests in the EEF Database', EEF Research Paper 001, London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Support/EEF_Research_Papers/Research_Paper_1_-_Properties_of_commercial_tests.pdf

Barnett, W. S. (2011) 'Effectiveness of Early Education Intervention', *Science*, 333 (6045), pp. 975–978.

Bullock, J. C. (2005) 'Effects of the Accelerated Reader on Reading Performance of Third, Fourth, and Fifth-Grade Students in One Western Oregon Elementary School', University of Oregon.

Cunningham, A. E. and Stanovich, K. E. (1997) Early Reading Acquisition and Its Relation to Reading Experience and Ability 10 Years Later', *Developmental Psychology*, 33 (6), p. 934.

Department for Education (2013) 'The National Curriculum in England: Key Stages 1 and 2 Framework Document' (September 2013): https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/425601/PRIMARY_national_curriculum.pdf

Department for Education statistics (2015) 'National Statistics: Schools, Pupils and Their Characteristics: January 2015': https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2015

Dong, N. and Maynard, R. A. (2013) 'PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Sample Size Requirements for Experimental and Quasi-Experimental Designs', *Journal of Research on Educational Effectiveness*, 6 (1), pp. 24–67.

Doyle, O., Harmon, C. P., Heckman, J. J. and Tremblay, R. E. (2009) 'Investing in Early Human Development: Timing and Economic Efficiency', *Economics and Human Biology*, 7 (1), pp. 1–6.

Dunatchik, A., Wishart, R., Cartagena-Farias, J. and Smith, N. (2018) 'Regional Differences in Attainment in the Early Years' Research Report (August 2018: https://www.gov.uk/government/publications/regional-differences-in-attainment-in-the-early-years

EEF (2016a) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Guidance_Final.pdf

EEF (2016b) 'Cost Evaluation Guidance', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/EEF_guidance_to_evaluators_on_cost_evaluation_2016_revision_FINAL.pdf

EEF (2018) 'Statistical Analysis Guidance for EEF Evaluations', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

EEF (2019a) 'Catch Up Literacy' (re-grant), London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/catch-up-literacy-effectiveness-trial#closeSignup

EEF (2019b) 'Read Write Inc. Phonics', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/read-write-inc-phonics/

Gorard, S., Siddiqui, N. and Huat See, B. (2015) 'Accelerated Reader: Evaluation Report and Executive Summary', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/uploads/pdf/Accelerated_Reader_(Final).pdf

Hall, S. S., Kowalski, R., Paterson, K. B., Basran, J., Filik, R. and Maltby, J. (2015) 'Local Text Cohesion, Reading Ability and Individual Science Aspirations: Key Factors Influencing Comprehension in Science Classes', *British Educational Research Journal*, 41 (1), pp. 122–142.

Hedges, L. V. (2007) 'Effect Sizes in Cluster-Randomized Designs', *Journal of Educational and Behavioral Statistics*, 32 (4), pp. 341–370.

Huang, S. (2012) 'A Mixed Method Study of the Effectiveness of the Accelerated Reader Program on Middle School Students' Reading Achievement and Motivation', *Reading Horizons*, 51 (3), pp. 229–230.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook', London: Education Endowment Foundation.

Hutchison, D. and Styles, B. (2010) 'A Guide to Running Randomised Controlled Trials for Educational Researchers', Slough: NFER.

Institute for Effective Education (2019) 'Interference Training: Intervention Details': https://www.evidence4impact.org.uk/interventions/1060?page=9#

Kahan, B. C. and Morris, T. P. (2012) 'Reporting and Analysis of Trials Using Stratified Randomisation in Leading Medical Journals: Review and Reanalysis', *BMJ*, 345: http://www.bmj.com/content/bmj/345/bmj.e5840.full.pdf

Lesaffre, E. (2008) 'Superiority, Equivalence, and Non-Inferiority Trials', *Bulletin of the NYU Hospital for Joint Diseases*, 66 (2).

Lortie-Forgues, H. and Inglis, M. (2019) 'Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned?', *Educational Researcher*, p.0013189X19832850: http://eprints.whiterose.ac.uk/141754/

Ludwig, J., Kling, J. R. and Mullainathan, S. (2011) 'Mechanism Experiments and Policy Evaluations', NBER Working Paper No. 17062: https://www.nber.org/papers/w17062

NGRT (2015) 'New Group Reading Test': http://www.gl-assessment.co.uk/products/new-group-reading-test

Nichols, J. S. (2013) 'Accelerated Reader and Its Effect on Fifth-Grade Students' Reading Comprehension', doctoral dissertation, Liberty University.

Nunnery, J. A., Ross, S. M. and McDonald, A. (2006) 'A Randomized Experimental Evaluation of the Impact of Accelerated Reader/Reading Renaissance Implementation on Reading Achievement in Grades 3 to 6', *Journal of Education for Students Placed at Risk*, 11 (1), pp. 1–18.

Oakley, A., Strange, V., Bonell, C., Allen, E. and Stephenson, J. (2006) 'Process Evaluation in Randomised Controlled Trials of Complex Interventions', *BMJ (clinical research edn)*, 332 (7538), pp. 413–416.

Ross, S. M., Nunnery, J. and Goldfeder, E. (2004) 'A Randomized Experiment on the Effects of Accelerated Reader/Reading Renaissance in an Urban School District: Final Evaluation Report', Memphis, TN: Center for Research in Educational Policy, The University of Memphis.

Shannon, L. C., Styers, M. K., Wilkerson, S. B. and Peery, E. (2015) 'Computer-Assisted Learning in Elementary Reading: A Randomized Control Trial', *Computers in the Schools*, 32 (1), pp. 20–34.

Slavin, R. E., Lake, C., Davis, S. and Madden, N. A. (2011) 'Effective Programs for Struggling Readers: A Best-Evidence Synthesis', *Educational Research Review*, 6 (1), pp. 126.

Sutherland, A. and Ilie, S. (2018) 'Accelerated Reader Statistical Analysis Plan', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Projects/Accelerated_Reader_SAP_2018.05.02__FINAL.pdf

Sutherland, A., Sim, M., Ilie, S. and Krapels, J. (2018) 'Evaluation Protocol' (March 2018), London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Accelerated_Reader_Protocol_Update_2017.12.18.pdf

Treadaway, M. (2013) 'An Analysis of Key Stage 2 Reliability and Validity', FFT Research Paper No.2, FFT Education Ltd: http://www.fft.org.uk/FFT/media/fft/News/KS2_Reliability_-_Validity_MT_FINAL.pdf

Vygotsky, L. (1978) 'Interaction Between Learning and Development', in *Mind and Society*, Cambridge, MA: Harvard University Press. Reprinted in Gauvain, M. and Cole, M. (eds) (1997) *Readings on the Development of Children* (2nd edn), New York: W. H. Freeman and Company.

White, H. R., Mun, E. Y. and Morgan, T. J. (2007) 'Long-Term Effects of Brief Substance Use Interventions for Mandated College Students: Sleeper Effects of an In-Person Personal Feedback Intervention', *Alcoholism: Clinical and Experimental Research*, 31 (8), pp. 1380–1391.

WWC (2008) WWC Intervention Report: Accelerated Reader. Available from https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_acceleratedreader_061416.pdf, last accessed 5th July 2021.

WWC (2009) 'Intervention Report: Lexia Reading' (June 2009), What Works Clearing House, Institute of Education Sciences, U.S. Department of Education: https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_lexia_063009.pdf

Appendix A: EEF cost rating

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

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

Appendix B: Security classification of trial findings

OUTCOME: *KS2 reading scores*

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

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Low	Appropriate randomisation done by independent evaluators. Primary outcomes assessments (KS2 R&W) highly reliable. Groups (overall and FSM) fairly balanced according to the KS1 (APS Scores).
Threat 2: Concurrent Interventions	Moderate	The control schools invested in buying resources for the library and using a number of other literacy interventions which suggests some level of literacy activity might have impacted on pupils in control conditions. The evaluators argue this explains the lack of difference in outcomes between treatment arms.
Threat 3: Experimental effects	Low	The study is a school level randomisation so not a threat due to experimental conditions implemented in treatment schools. Evaluators report control schools which intended to use the STAR Reading test (part of the AR approach), but these were immediately stopped.
Threat 4: Implementation fidelity	Low	Fidelity was assessed using RL data. The results suggest high level of intervention implementation. Relevant analyses were conducted which showed consistent results. However, some schools discontinued the implementation. This meant that out of the 86 schools implementing AR, 69 schools (80%) implemented AR for two years, 10 schools (12%) implemented AR for one year and 7 schools (8%) stopped implementation during the first year of implementation.
Threat 5: Missing Data	Low	Missing data was minimal and similar across treatment arms.
Threat 6: Measurement of Outcomes	Low	KS1 and KS2 are highly reliable outcome measures.
Threat 7: Selective reporting	Low	Analyses presented according to the statistical analysis plan.

- **Initial padlock score:** [5] Padlocks – Well conducted randomised controlled trial with low attrition (6%) and MDES of 0.16 at randomisation for the whole sample and 0.18 for FSM pupils.
- **Reason for adjustment for threats to validity**: [0] Padlocks
- **Final padlock score:** initial score adjusted for threats to validity = [5] Padlocks

Appendix C: NFER school invitation letter

RPO/EEAR/42613/2 NFER No: «NFER_No»

The Headteacher
«Description»
«Address1»
«Address2»
«Address3»
«Town»
«County»
«Post_Code»

February 2016

Dear Headteacher

Evaluation of the Accelerated Reader Programme

The National Foundation for Educational Research (NFER) has been asked by the Education Endowment Foundation (EEF) to carry out, in collaboration with RAND Europe and University of Cambridge, an evaluation of the effectiveness of the Accelerated Reader programme that is made available to schools by Renaissance Learning. We would like to invite your school to take part in a randomised control trial which is designed to establish the extent to which the Accelerated Reader programme helps to improve children's reading skills. A number of considerable benefits are being made available to schools that participate in the trial and we believe that the children involved will enjoy the experience that the reading programme has to offer. Over 4,000 schools in the UK are already using and enjoying Accelerated Reader, this trial would give your school the opportunity to experience these benefits free of charge for one year; the average cost normally associated with the first year purchase of the programme for one year group is around £2,000 (which includes a one-time first year training cost of £998).

In the case of LA maintained schools we have notified your authority of our intention to contact you about this research.

What is Accelerated Reader?

Accelerated Reader is a programme that encourages a love of reading in children. Delivered by Renaissance Learning, the Accelerated Reader programme aims to help schools foster the habit of independent reading among primary and secondary pupils of all abilities. Through a combination of Accelerated Reader software, dedicated reading time and ongoing professional development for staff, the programme helps teachers to screen pupils to find their reading abilities and then easily match pupils to popular books that you have in your school at the appropriate level for each child. The children's understanding and enjoyment of the books they have read are measured through a series of mini-quizzes taken online and reports are produced to enable teachers to identify what kind of support is needed for each child to progress further.

What we would like your school to do for this evaluation

All schools that agree to participate in the evaluation will be randomly assigned to one of two groups. Half of the schools will be assigned to be in the intervention group and the other half to a control group and undertake activities as follows:

- Intervention schools commit to using the Accelerated Reader programme with years 4 and 5 children during the academic year 2016-17, starting this September;

- Control group schools are asked to maintain a 'business as usual' approach to the teaching of reading to their children until September 2017. Accelerated Reader is then used with children in years 3 and 4, during the academic year 2017-18.

Measuring the outcomes of the trial

An important aspect of the evaluation is to take a measurement of the performance in reading of the children in the intervention group after one year of involvement with Accelerated Reader and compare this to the performance of children in the control group that have not used Accelerated Reader. Further details with regards to the measurement of the outcome of the trial can be found in the information sheet enclosed.

What will be the benefit to participating schools?

Intervention schools will receive the Accelerated Reader programme free of charge for year 4 and 5 children during 2016-17 and control group schools will receive the programme free of charge for their years 3 and 4 children during 2017-18. Training and assistance will be provided by Renaissance Learning to help implement the programme in your school.

All marking of the reading tests taken by your pupils in Year 4 in June 2017 will be carried out by NFER and age-standardised scores returned to you.

How to let us know if you can take part

If you would like to sign up to take part in this evaluation exercise we would be delighted to hear from you. Please complete the enclosed reply form and return it to NFER in the pre-paid envelope provided. We very much hope that you can help us with this evaluation. If you have any queries about this research please do not hesitate to contact me or my colleague David Hereward on 01753 637352 or by email to **EEAR@NFER.ac.uk**.

Yours sincerely

Kathryn Hurd
Head of Survey Operations
National Foundation for Educational Research

Appendix D: NFER school reply form

National Foundation for Educational Research
RPO, The Mere, Upton Park
Slough, Berkshire. SL1 2DQ
Telephone: 01753 637007
Fax: 01753 790114

RPO/«Sample_No»/2a

NFER No: «NFER_No»

Evaluation of the Accelerated Reader programme

Please complete part 1, then part 2 or 3 as appropriate:

Part 1:

	Are your details correct?	Amendments
School name	«Description»	
Headteacher	«Head_Title» «Head_Forename» «Head_Surname»	
Telephone	«Phone»	
Fax	«Fax»	
Email	«Email»	

Part 2:

2a) Please complete the following:

My school **can** take part in this evaluation

Headteacher/SMT signature:
.....

Name of contact in the school:
Mr/Mrs/Miss/Ms/Dr.....
.....

Contact phone number: ..... **Contact job title:**
.....

Contact email address:
.....
....

«Contact_ID» For NFER use: 2a/contact/reg EEAR

Please turn over

2b) In the table below please enter the number of children that you estimate will be starting in year 4 and in year 5 in September 2016

Class name	No. of pupils in year group starting in September 2016
Year 4	
Year 5	

Part 3:

My school is unable to take part in this evaluation *Please tick* ☐

We would be grateful if you could let us know the reasons why your school cannot take part in this project. This will help us understand the factors schools take into consideration when deciding whether or not to participate in our projects.

……

……

……

Thank you for completing this form – Please see the final sheet with some additional questions relating to this project, please return the questions along with the reply form.

Please return this to the NFER as soon as possible using the pre-paid envelope provided, via email to EEAR@NFER.ac.uk or by fax to 01753 790114.

«Contact_ID» For NFER use: 2a/contact/reg EEAR

Part 4: If we could please ask you to respond to a couple of questions below which relate to the project – please tick the relevant response:

Questions about books in your school:

1. How many books, very approximately, do you have in your school?

☐ 500 or fewer ☐ 501 – 1000 ☐ 1001 – 2499
☐ 2,500 – 5,000 ☐ 5000 – 7499 ☐ more than 7,500

2. Did your school buy any books last year?
☐ Yes ☐ No

a. If yes, how many?

☐ 0-25 ☐ 26-50 ☐ 51-75 ☐ 76-100 ☐ 101-125
☐ 126-150 ☐ 151-175 ☐ 176-200 ☐ 200+

Questions about pupils accessing books in the school:

1. How are most books organised in your school?
☐ Most books are in a central area within the school i.e. a library/library space
☐ Most books are within each classroom
☐ About equal number of books in both a central area and within classrooms

2. Do you currently have dedicated reading time in your timetable for pupils in:
Year 4? ☐ Yes ☐ No

Year 5? ☐ Yes ☐ No

a. If yes, how much time per week?
Year 4
☐ Less than 1 hour ☐ between 1 and 2 hours ☐ More than 2 hours

Year 5
☐ Less than 1 hour ☐ between 1 and 2 hours ☐ More than 2 hours

Questions about current literacy-focused activities in the school:

In your school, do you have a:	
Reading scheme? *Definition:* A school reading programme is characterised by the provision of reading material and the involvement of parents to stimulate the reading of children. They are non-competitive and rest on the principle that children learn at their own pace. Some examples: Oxford Reading Tree (ORT), Collins Big Cat and Rigby Star, Lighthouse, Storyworlds and New Reading 360	Reading intervention? *Definition:* A programme seeking the improvement of reading skills with clear-cut stages and attainable goals. They may focus on phonics, comprehension, vocabulary among other aspects. Some examples: Accelerated Reader, Reading Recovery, Peer Assisted Learning Strategies (PALS) for reading, Every Child a Reader (ECaR), Catch Up Programme and Success for All.
☐ Yes ☐ No	☐ Yes ☐ No
If yes, which scheme(s) or intervention(s) do you have? Please list by name or short description below.	
Reading scheme:	Reading intervention:

«Contact_ID» For NFER use: 2a/contact/reg EEAR

Appendix E: Parental consent form

RPO/EEAR/42613/p

April 2016

Dear Parent /Guardian

Evaluation of the Accelerated Reader Programme

Your school has elected to take part in an evaluation of the Accelerated Reader Programme. The school will be randomly selected to experience the programme either for the academic year 2016 – 2017 for years 4 and 5 pupils, or for the academic year 2017 – 2018 for years 3 and 4 pupils.

Accelerated Reader is a programme that encourages a love of reading in children. Developed by Renaissance Learning, the Accelerated Reader programme aims to help schools foster the habit of independent reading among primary and secondary pupils of all abilities. Through a combination of Accelerated Reader software, dedicated reading time and ongoing professional development for staff, the programme helps teachers to establish pupils' reading abilities and then match pupils to popular books that are at the appropriate level for each child. The children's understanding and enjoyment of the books they have read are measured using mini-quizzes which help teachers to identify what kind of support is needed for each child to progress further.

After one year of using the programme the pupils in year 4 in all schools involved in the evaluation will also sit one short 30 minute English comprehension test. For the purpose of this evaluation we will collect a few details for each pupil starting year 4 and year 5 in September 2016. These details will be used to help with the analysis of the results of the quizzes and tests taken by the children and include pupil name, class name, Unique Pupil Number, Sex, and Free School Meals status.

Pupils' test responses and any other pupil data will be treated with the strictest confidence. Named data will be matched with the National Pupil Database and shared with RAND, the University of Cambridge and Renaissance Learning, the Department for Education, Education Endowment Foundation (EEF), EEF's data contractor FFT Education and in an anonymised form to the UK Data Archive. Please note that no individual children will be identified in any reports arising from the research.

Whilst we expect that your child will enjoy being part of the programme, your child's information can be withdrawn from the evaluation at any time. If you would prefer your child's information NOT to be included in the evaluation, please inform your child's teacher. There is a reply slip provided overleaf for returning to the school should you opt not to include your child's information in the evaluation. Please note that if you are happy for your child's information to be included, you **DO NOT** need to return the reply slip.

If you have any queries or would like more information about this research, please do not hesitate to contact me or my colleagues Rebecca Menys or David Hereward on 01753 637225.

Yours sincerely

Kathryn Hurd
Head of Survey Operations

Evaluation of the Accelerated Reader Programme

IF YOU APPROVE OF YOUR SON'S/DAUGHTER'S INFORMATION BEING INCLUDED IN THE EVALUATION IT IS **NOT NECESSARY** FOR YOU TO RETURN THIS FORM TO THE SCHOOL

I have been informed of the nature of the evaluation and **I do not wish for my *son's/daughter's information to be included.***

Print name of child: _____

Print name of school: _____

Parent/Guardian's signature: _____ Date: _____

Please would you briefly outline your reasons for not wishing your son/daughter to participate in this research:

_____

_____

_____

Please return this form to your son's/daughter's school.

RPO/EEAR/42613 **PCNS**

Appendix F: Breakdown of survey responses by role

In this appendix we present the breakdown of respondents by their reported main role in the first, second and endline intervention staff surveys. Note that when added up, the number of respondents in the tables do not match the total number of respondents in each survey, as a small subset of respondents reported multiple main roles.

Table 26: Breakdown of respondents in intervention staff survey 1 (n=147) by main role

Main role	Number of respondents
Head teacher	**37**
Head of literacy	**40**
Year 4 teacher	**41**
Year 5 teacher	**43**
Librarian	**3**
Teaching assistant	**3**

Table 27: Breakdown of respondents in intervention staff survey 2 (n=169) by main role

Main role	Number of respondents
Head teacher	**26**
Head of literacy	**28**
Year 4 teacher	**61**
Year 5 teacher	**53**
Librarian	**5**
Teaching assistant	**9**

Table 28: Breakdown of respondents in intervention staff endline survey (n=87) by main role

Main role	Number of respondents
Head teacher	**16**
Head of literacy	**24**
Year 4 teacher	**23**
Year 5 teacher	**24**
Librarian	**2**
Teaching assistant	**8**

Appendix G: Addendum Report

Accelerated Reader

Evaluation Report

July 2021

Authors:

Sonia Ilie & Miriam Broeks

Contents

Executive summary

This addendum report should be read in conjunction with the main report for this project, 'Accelerated Reader Evaluation Report' (Sutherland et al, 2021). The main report presents the findings of the impact of Accelerated Reader (AR) on a cohort of 5,759 pupils who were offered the AR programme in Year 5 and when they were in Year 6 (ages 9-11 years). The randomised control trial compared longer-term outcomes of AR on reading achievement at the end of Key Stage 2 in 88 schools who received the programme and 92 schools who continued with 'business as usual' teaching. The main report also presents the implementation and process evaluation that was undertaken during the trial.

This addendum report presents the findings from the cohort of Year 4 pupils where the above noted treatment schools were offered AR for two years, when the pupils were in Year 4 and Year 5 (ages 8-10 years) and assesses the longer-term impact of the programme using pupils Key Stage 2 SATs reading score. The trial initially planned to assess the impact of one year of AR on children in Year 4 and to also look at the longer-term impact of the programme by collecting Key Stage 2 SATs reading scores for pupils who started the programme in Year 5. Due to problems with the original outcome measure delivered to pupils at the end of Year 4 the length of the trial was extended, the primary outcome measure was changed as well as the main cohort of interest. The primary outcome became Key Stage 2 SATs reading scores for the pupils who started the programme in Year 5, results of which are reported in the main trial report above.

The project

Accelerated Reader (AR), developed by Renaissance Learning, is a digital whole-class reading management and monitoring programme that aims to foster independent reading among primary and secondary pupils. AR is premised on pupils practicing reading regularly in school, reading books at an appropriate level of difficulty, testing pupil comprehension, and providing regular feedback to teachers.

The evaluation of AR involved 181 primary schools with 6,311 pupils in the Year 4 cohort across 89 treatment and 92 control schools. A total of 93.5%, or 5,901 pupils, were retained in the trial for the final analysis with Key Stage 2 reading attainment data obtained from the National Pupil Database. Schools were randomly allocated to either receive the intervention, or to a wait-list business-as-usual control group: trial cohorts in control schools continued as usual, but control schools were able to use AR with non-trial younger pupil cohorts from the second year of the trial onwards.

Summary of Results

Children who started Accelerated Reader in Year 4, on average, made no additional progress in reading compared to children in the comparison schools. This is our best estimate of impact, which has a very high security rating. As with any study, there is always some uncertainty around the result. The possible impacts found here range from small negative effects of one month's less progress to small positive effects of up to two month's progress. Children eligible for free school meals (FSM) who started AR in Year 4, on average, made one month of additional progress in reading compared to children eligible for FSM in the comparison schools. The impacts found for this group range from small negative effects of one month's less progress to small positive effects of up to two month's progress. Due to the statistical uncertainty around this finding the evaluation team do not consider the FSM result to constitute evidence of promise.

The results from this second cohort are in line with results from the Year 5 cohort where pupils in schools who were offered AR for two years made no additional progress compared to pupils in the comparison schools. The additional analysis for pupils eligible for FSM in the Year 4 cohort highlight on average a small positive impact of the programme. However, due to the statistical uncertainty around the FSM result these findings should be interpreted with caution.

Results from the process evaluation, outlined fully in the main report, suggest that some of the control schools used other reading interventions and reading programmes. This applies to the Year 4 cohort as well. This may mean there was not enough 'unique' activity in AR schools to produce measurable differences compared to comparison schools. As discussed in the main report, this may partially explain the null result for both the Year 4 and Year 5 impact analysis.

Comprehensive monitoring data provided by Renaissance Learning indicated that AR was implemented as intended in intervention schools, with high implementation fidelity during the first year of delivery across both cohorts and fairly-high implementation fidelity in the second year. Implementation was calculated by considering pupils' average score on quizzes, engaged reading time, whether appropriately-levelled books were selected, and teacher responsiveness to pupil reading data. Although implementation overall was described as fairly high, it is of interest that the ideal engaged reading time of between 15 – 30 minutes a day was only achieved by between 38% to 43% of pupils in year one and

between 30% to 42% of pupils in year two. This could provide another potential partial explanation for the null result in this trial.

AR was very well received by teachers, who reported perceived positive impacts on pupils' reading abilities and attitudes to reading. Full process evaluation results for both Year 4 and Year 5 cohorts are reported in the main report above.

Table 29 Summary of impact on primary outcome

Key conclusions
1. Children who started Accelerated Reader in Year 4, on average, made no additional progress in reading compared to children in the comparison schools. This result has a very high security rating.
2. Children eligible for Free School Meals (FSM) who started Accelerated Reader in Year 4, on average, made one month of additional progress in reading compared to children eligible for FSM in the comparison schools. However, this result has high statistical uncertainty.

Outcome/ Group	Effect size (95% confidence Interval)	Estimated months' progress	EEF security rating	No. of pupils	P value	EEF cost rating
Primary outcome (KS2 reading)	0.02 (-0.07, 0.11)	0	🔒🔒🔒🔒🔒	5,901	0.673	£ £ £ £ £
Primary outcome (KS2 reading) FSM pupils	0.05 (-0.07, 0.15)	1	N/A	2,123	0.537	£ £ £ £ £

Introduction

The current report presents results of the impact evaluation of Accelerated Reader for the Year 4 cohort of students participating in the trial. The analysis reported below follows the trial design and analytical approaches outlined fully in the main evaluation report (Sutherland et al., 2021) and summarised below

Headline evaluation findings: Year 5 cohort impact and full process results

The aim of the evaluation of Accelerated Reader was initially to assess the extent to which AR leads to an improvement in reading comprehension. After changes to the trial design resulting from issues with the original primary outcome measure (outlined above in the Main Report), the revised aim of the trial was to test the impact of AR on reading ability at Key Stage 2. The trial was designed as a two-arm, wait-list, cluster-randomised trial. The unit of randomisation was the school; the unit of analysis was the pupil. There were two cohorts included in the trial, pupils initially in Year 4, and pupils initially in Year 5. In effect, this means that two parallel trials were conducted at once. The evaluation was powered to detect an impact on children eligible for Free School Meals (FSM). A comprehensive process evaluation was undertaken, reported in full in the main report. The process evaluation consisted of baseline and end-line surveys in treatment and control groups; observations of training; a set of two staff surveys in treatment schools; and one head teacher survey in control schools only; staff workshops and staff interviews in treatment schools; and administrative data collected by the online system used in the implementation of AR.

The Year 5 impact results, reported in full in the main report, pointed to pupils in schools randomly allocated to receive AR showing no greater reading ability than pupils in control schools. The trial also found no effect of AR on reading ability for pupils eligible for FSM.

The process evaluation found that AR was implemented with a high degree of fidelity. The comprehensive monitoring data on the use of AR also showed very limited contamination in the (wait-list) control group. A majority of treatment schools reported using dedicated reading time, in keeping with the intervention's procedures. Staff implementing AR in treatment schools also reported consistently engaging with the reading tests and quizzes, as well as with data and reports generated by AR, a key element of AR's implementation. Some evidence further suggested that as time went on staff became more positive about the use of these data in their work, The staff surveys, workshop, and interviews suggested that AR was very well received. Staff working with both cohorts reported perceived positive impacts of AR on pupil reading ability, and on pupils' attitudes to reading.

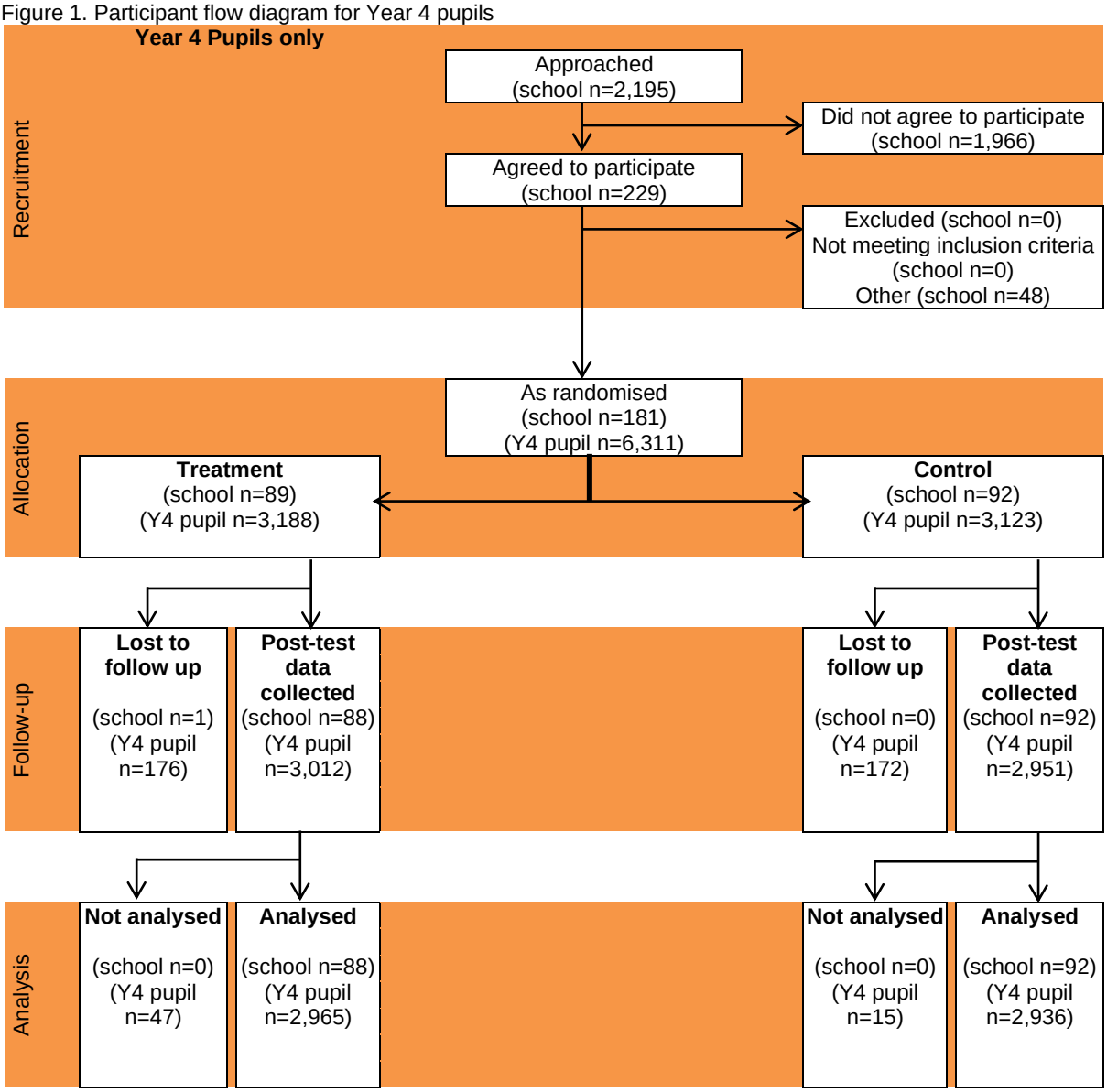
Impact evaluation: Year 4 cohort

Participants

The trial design treats Year 4 and Year 5 pupils as two separate trials. Figure 1 illustrates the full Year 4 participant flow diagram.

As Figure 1 illustrates, at randomisation stage NFER had recruited 181 schools to participate in this trial. Using randomisation procedures outlined in the main trial report, schools were randomly allocated to either control or treatment groups. Following this random allocation, the Year 4 arm of the trial consisted of a total of 6,311 pupils in 89 treatment schools (3,188 pupils) and 92 control schools (3,123 pupils).

Figure 1. Participant flow diagram for Year 4 pupils



Attrition

Of the 6,311 pupils in the Year 4 trial arm, 6,279 were matched to NPD records for the purposes of outcome data collection, and of these 6,279, data on Key Stage 2 outcomes was available for 5,963. This represents a rate of attrition from randomisation to post-test data collection of 5.6%.

At analysis stage, a further 62 pupils were not analysed due to missing data on their Key Stage 1 baseline measure. A total of 5,901 pupils were analysed: 2,965 in the treatment group and 2,936 in the control group. This represents a rate of attrition from randomisation to analysis of 6.5%. From randomisation to analysis, the retention rate is 93% (7% attrition rate) for the treatment group; and 94% (6% attrition rate) for the control group. For comparison, the randomisation-to-analysis attrition rate for the Year 5 cohort in the trial was 5.8% (for both trial arms).

Sample size calculations

At the analysis stage for the younger (Year 4) cohort, there were 88 schools in the treatment group (2,965 pupils) and 92 schools (2,936 pupils) in the control group.

For the purposes of at-analysis effect size (MDES) calculations, the intra-class correlation (ICC) and pupil-level explained variance were derived empirically from the data held and are presented in Table 2. All power and effect size calculations were carried out using PowerUp (Dong & Maynard, 2013).

Table 30. Sample size and minimum detectable effect sizes for different stages of the trial: Y4 cohort

		Protocol		Randomisation		Analysis: Y4 only	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.15	0.17	0.16	0.18	0.132	0.143
Pre-test/ post-test correlations (KS1_READWRITPOINTS and KS2READMARKN)	level 1 (pupil)	0.73	0.73	0.73	0.73	0.667[e]	0.628[e]
	level 2 (class)	NA	NA	NA	NA	NA	NA
	level 3 (school)	0.00	0.00	0.00	0.00	0.00	0.00
Intracluster correlations (ICCs)	level 2 (class)	NA	NA	NA	NA	NA	NA
	level 3 (school)	0.13	0.13	0.13	0.13	0.083	0.067
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two	Two	Two
Average cluster size		66	20	68.5[a]	22.8[b]	32.8[f]	12[f]
Number of schools	intervention	100	100	89	89	88	86
	control	100	100	92	92	92	91
	total	200	200	181	181	181	177
Number of pupils	intervention	6,600	2,000	6,274	2,032[c]	2,965	1,088
	control	6,600	2,000	6,124	2,101[c]	2,936	1,035
	total	13,200	4,000	12,398	4,133[d]	5,901	2,123

Table note: [a] Based on summing the average number of pupils per school for treatment and control arms (70.5 and 66.6 respectively) and dividing by two. [b] This is estimated by dividing the overall figure by three, given an expected rate of 33% Ever 6 FSM eligibility. [c] Based on multiplying the number of schools by the average number of pupils per school. [d] The sum of the two cells above. [e] Correlation between KS1_READWRITPOINTS (measured 2014 for Year 4 cohort) and KS2READMARKN (2019 exam, accessed 2020). [f] Average observations per group in analysis.

At design stage, the overall MDES for the primary outcome analysis was 0.15 for each respective year group. At randomisation, this stood at 0.16. At analysis stage, for the Year 4 cohort only, the overall MDES for the primary outcome analysis was 0.132.

At design stage, the MDES for the sub-group of free school meal eligible pupils for the primary outcome measure was 0.17 for each respective year group in the trial. At randomisation, this was 0.18. At analysis stage, for the Year 4 cohort only, the MDES for the FSM sub-group was 0.143. The at-analysis MDES was smaller than the at-protocol MDES primarily due to the lower intra-cluster correlation (ICC), replacing the estimated value (from previous research) with an empirically derived value (from data on the primary outcome measure).

Sub-group sample size: pupils eligible for free school meals.

As Table 2 above illustrates, separate sample size calculations were carried out for the pre-specified sub-group of pupils eligible for free school meals (FSM). The Ever-6 FSM variable was used to identify FSM-eligible pupils. To ensure maximum data coverage, this combined existing information from Spring Census data in two consecutive pre-trial years (variables: EVERFSM_6_P_SPR17 and EVERFSM_6_P_SPR16). Fewer than 10 pupils were observed with missing data on the combined FSM variable, with a total proportion of available FSM data of over 99.9%.

Therefore, a total of 2,123 FSM-eligible pupils were included in the sub-group analysis, 1,088 in the treatment group and 1,035 in the control group.

School and pupil characteristics

Full accounts of the characteristics of trial schools, and of baseline pupil characteristics are included in the main report.

The main report notes that 6,224 Year 4 pupil records were matched with the NPD for the purposes of baseline data collection. For the purposes of outcome data collection, this match was increased to 6,279.

At analysis stage, the characteristics of the Year 4 pupil cohort are illustrated in Tables 3 and 4.

Table 31. Year 4 pupil characteristics at analysis stage (categorical variables only)

Variable	Treatment group		Control group		ES
	N	Percentage	N	Percentage	
Eligible for FSM (ever-6)	1,088	36.69%	1,035	35.25%	0.03
Female	1,515	51.09%	1,462	49.79%	0.02

Table 32. Year 4 pupil characteristics at analysis stage (continuous variables only)

Variable	Treatment group		Control group		ES
	Mean (N)	Standard Deviation	Mean (N)	Standard Deviation	
Key Stage 1 score	16.11 (2,965)	3.22	16.06 (2,936)	3.17	0.01

In keeping with school and pupil characteristics at baseline for both Year groups in the trial, and with the at-analysis Year 5 characteristics illustrated in the main report, balance was continued to be observed at analysis stage for Year 4 pupils. In terms of FSM eligibility, the rates were very similar and within 2 percentage points of each other, by treatment and control group. A similar pattern was observed for the gender distribution (proportion of female pupils in each group). The Key Stage 1 baseline score was similarly balanced across the two groups, with a difference of less than 2% of a standard deviation. The balance between intervention and control group for major ethic group (White British) was slightly reduced, but still within an acceptable 10 percentage point margin (86% in the Intervention group; 78% in the Control group).

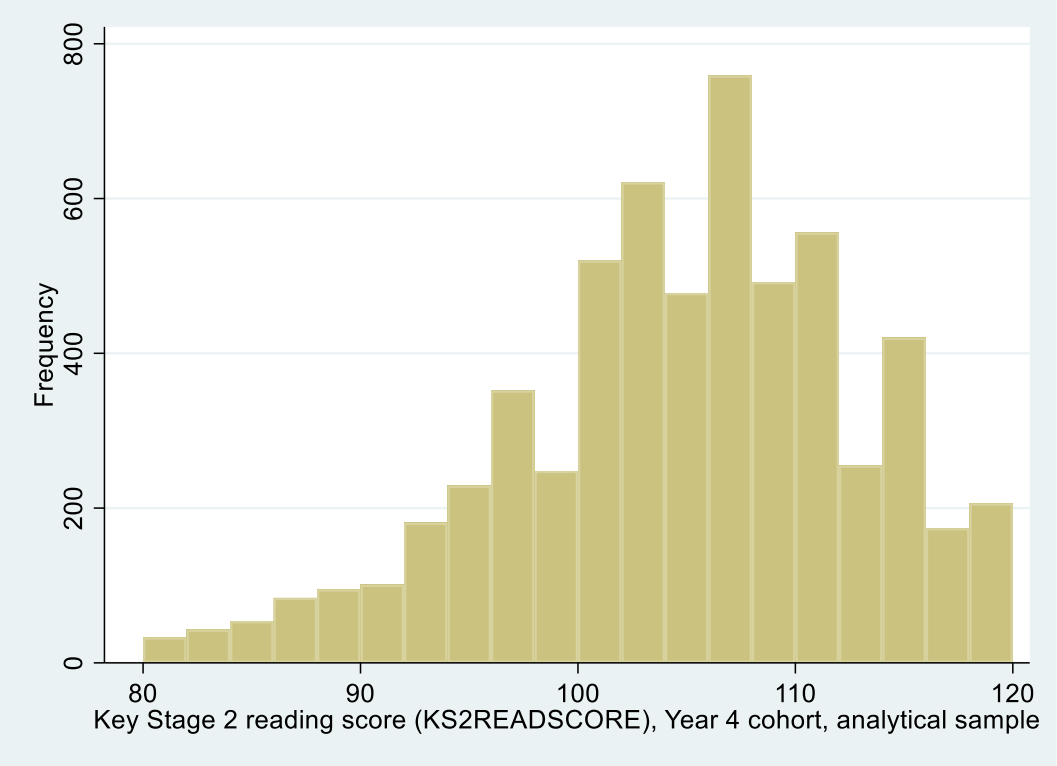
Outcomes and analysis

Primary outcome measure

The primary outcome measure for the Year 4 cohort is the raw fine grained Key Stage 2 reading score, obtained from the KS2_READSCORE variable. This variable was sourced from the National Pupil Database; it referred to the school year 2018-2019, i.e. Key Stage 2 exams sat in 2019; and accessed in 2020. In the Year 5 cohort, the KS2_READMARK variable was used, which showed a very high correlation (>0.99) with the KS2_READSCORE variable. The KS2_READMARK variable was not available in the NPD for the Year 4 cohort and therefore the KS2_READSCORE variable was used instead. This is aligned with the updated trial protocol (Sutherland et al., 2018, Evaluation Protocol).

Means, variances, and confidence intervals around the mean for the outcome measure, by treatment and control group, and for both the overall and sub-group analysis are reported in Tables 5 and 7 below. The distribution of the outcome measure for the analytical sample is illustrated in Figure 2 below.

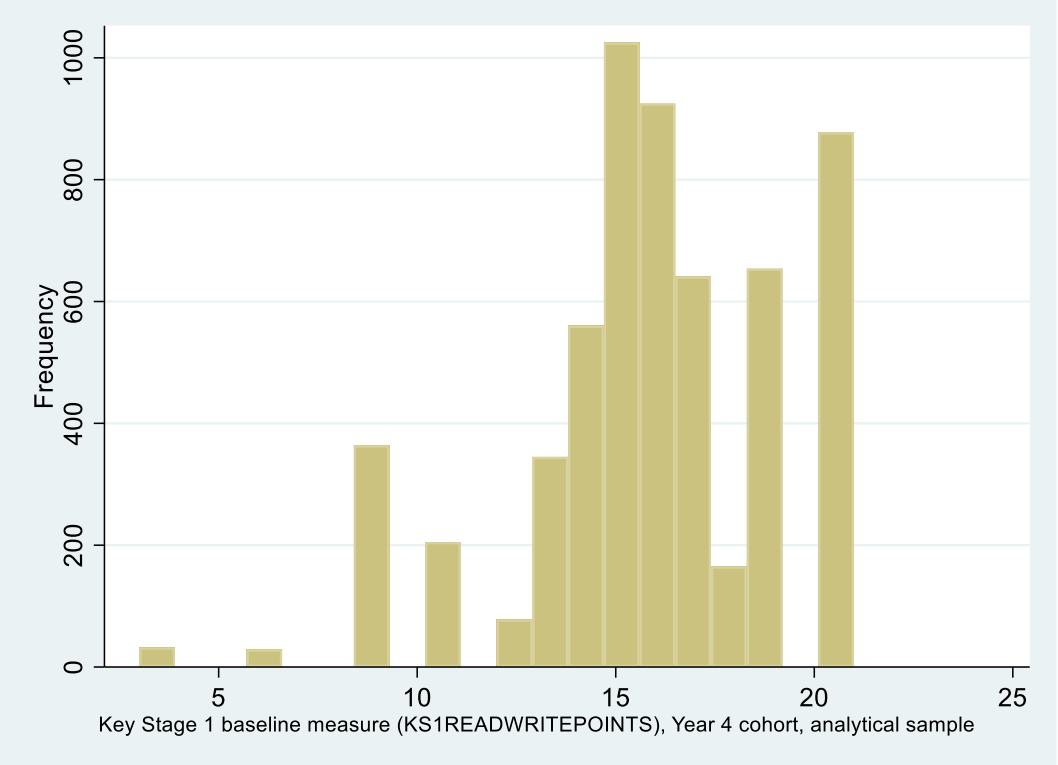
Figure 2. Distribution of Key Stage 2 reading score outcome measure, analytical sample. (smallest bin based on 36 observations)



Data on the primary outcome measure was available for 5,963 pupils (in Figure 1, the total of 'post-data collected at follow-up' pupils across treatment and control group).

Data on the baseline measure was collected through the NPD, for the Key Stage 1 assessments undertaken in 2014 for the Year 4 cohort, and first accessed 2017. Figure 3 illustrates the distribution of the baseline measure for the analytical sample.

Figure 3. Distribution of Key Stage 1 baseline measure, analytical sample. (smallest bin based on 39 observations)



Results

Primary outcome

Results of the intention to treat (ITT) analysis for the Year 4 cohort only point to an effect size (adjusted for the multi-level nature of the analytical model) of 0.02, with a significance level of 0.673. This indicates that younger cohort pupils' reading levels were similar in both treatment and business-as-usual schools at the end of KS2.

Table 33. Primary analysis (ITT)

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	N	Mean (95% CI)	N	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Primary outcome	2,965	104.46 (104.31, 104.61)	2,936	104.11 (103.96, 104.26)	5,901 (2,965, 2,936)	0.02 (-0.07, 0.11)	0.652

Table 34. Effect size estimation, primary analysis

Outcome	Unadjusted difference in means (intervention-control)	Adjusted difference in means	Intervention group		Control group		Pooled variance (Standard deviation)
			N	Variance of outcome	N	Variance of outcome	
Primary outcome	0.336	0.167	2,965	64.28	2,936	63.34	65.34 (8.08)

Sub-group analysis

The sub-group analysis focuses on pupils identified as eligible for free school meals (FSM) and uses the FSM-eligible sample subset to estimate the effect.

Table **14** and

Outcome	Raw means				Effect sizes		
	Intervention group		Control group		n in model (intervention; control)	Hedges g (95% CI)	p-value
	n (missing)	Mean (95%CI)	n (missing)	Mean (95%CI)			
Primary outcome	1,092	30.62 (30.01, 31.23)	1,0.71	30.33 (29.72; 30.94)	2,163 (1,092; 1,071)	0.020 (-0.08, 0.12)	0.685

Table 15 present the results of the analysis for the FSM group. Table 37. Subgroup analysis (ITT), Non-FSM groupTable 37 and

Table 17 present the results of the analysis for the non-FSM group.

The results of the sub-group analysis mirror those of the overall analysis for the Year 4 cohort. For the FSM sub-group, reading outcomes for pupils in treatment schools were as good as in control schools (there was no statistically significant difference). The effect size for the difference was positive, but very small, at 0.05 (Table 7). The non-FSM sub-group displayed similar outcomes, with a smaller effect size at 0.03 (Table 9).

Table 35. Subgroup analysis (ITT), FSM group

Outcome	Raw means				Effect sizes		
	Intervention group		Control group		N in model (intervention; control)	Hedges g (95% CI)	p-value
	N	Mean (95%CI)	N	Mean (95%CI)			
Primary outcome	1,088	102.32 (102.58, 102.07)	1,035	102.21 (101.95, 102.46)	2,123 (1,088; 1,035)	0.05 (-0.07, 0.15)	0.427

Table 36. Effect size estimation, FSM group

Outcome	Unadjusted difference in means (intervention-control)	Adjusted difference in means	Intervention group		Control group		Pooled variance (Standard deviation)
			N	Variance of outcome	N	Variance of outcome	
Primary outcome	0.11	0.369	1,088	70.28	1,035	66.28	68.30 (8.27)

Table 37. Subgroup analysis (ITT), Non-FSM group

Outcome	Raw means				Effect sizes		
	Intervention group		Control group		N in model (intervention; control)	Hedges g (95% CI)	p-value
	N	Mean (95%CI)	N	Mean (95%CI)			
Primary outcome	1,877	105.70 (105.52, 105.70)	1,901	105.14 (104.97, 105.32)	3,778 (1,877; 1,901)	0.03 (-0.06, 0.13)	0.482

Table 38. Effect size estimation, Non-FSM group

Outcome	Unadjusted difference in means (intervention-control)	Adjusted difference in means	Intervention group		Control group		Pooled variance (Standard deviation)
			N	Variance of outcome	N	Variance of outcome	
Primary outcome	0.56	0.265	1,877	61.39	1,901	58.76	60.13 (7.75)

Fidelity

As outlined in the main report, the main assumption behind the ITT approach taken to the analysis above means that levels of implementation fidelity are not considered in this analysis. To understand if fidelity of implementation is related to outcomes, a fidelity analysis is undertaken using a measure derived from records collected during the day-to-day implementation of AR. A full account of the fidelity measurement approach is provided in the main report.

Using the fidelity measures calculated for the purposes of the Year 5 (older) cohort analysis, a Year 4 cohort fidelity analysis was also undertaken. This used a two-stage least squares approach, with the fidelity score above regression on the treatment allocation variable, and residuals predicted from this. Control schools were assigned a fidelity score of 0 (i.e. no implementation of AR). The predicted standardised residuals previously generated were then entered into a multilevel model that matches the analytical approach for the primary outcome measure for both the Year 5 (main report) and the Year 4 cohorts (above), replacing the treatment variable with the newly-estimated fidelity measure.

This approach was undertaken separately for fidelity in each of the two years of AR implementation (2016-17 and 2017-18). Table 10 reports the results.

Table 39. Fidelity results

Fidelity year	Regression coefficient	95% CI for coefficient	P value	N
2016-17	0.11	-0.07, 0.29	0.223	5,822
2017-18	0.13	-0.02 0.28	0.053	5,822

The results from both fidelity measures above suggest a small, positive, but statistically non-significant (at the 5% level) relationship between fidelity of AR implementation and Key Stage 2 reading outcomes.

Conclusions

The trial was a two-arm, wait-list, cluster-randomised controlled trial, with schools as the unit of randomisation, and pupils as the unit of outcome analysis. Given the design of the trial, with two cohorts included in the study, this means in effect, that the evaluation team conducted two trials at once – one for Year 4 pupils and the other for Year 5 pupils. This trial was powered to detect an impact on disadvantaged children (eligible for free school meals).

Through the implementation process evaluation of this trial, it became evident that overall AR was well received by trial participants. Full process evaluation results are included in the main trial report. This report has presented only impact evaluation results for the younger (Year 4) trial cohort. These results, matching conclusions for the Year 5 results, showed no evidence of an effect of AR on pupil reading outcomes at Key Stage 2, compared to pupils allocated to business-as-usual schools.

As for the Year 5 cohort, Year 4 pupils in control schools did just as well as pupils in treatment schools. Similarly, we found no evidence that AR having a differentially positive impact on Year 4 FSM-eligible pupils' reading outcomes compared to FSM-eligible pupils in control schools.

One limitation pertaining to the analysis for Year 4 pupils pertains to the timing of the intervention. As outlined in the main report, the intervention was extended to a second year of implementation upon the re-design of the trial, with a large proportion of treatment schools continuing the delivery of AR in the second year. AR implementation then concluded, and initially Year 4 pupils then spent a further year in school prior to their sitting Key Stage 2 exams (the source of the outcome measure in the above analysis). It is possible that this time delay from intervention conclusion to outcome measure collection may have diminished the impact of AR on reading. However, that would also imply that any effects which might have been initially present would disappear in the course of a post-intervention year. Evidence from the Year 5 cohort, whose outcome measure was collected at the end of the second year of AR implementation would not necessarily support this explanation.

Our quantitative assessment of fidelity also showed no difference between reading outcomes of pupils from treatment and control schools at differing levels of fidelity. This latter result is likely due to the overall high implementation fidelity, whereby the variation in fidelity was small, as many schools implemented AR as expected. A full account of the levels of implementation fidelity are included in the main evaluation report.

Results from the process evaluation, outlined fully in the main report, suggest that some of the control schools used other (non-targeted) reading interventions. This applies to the Year 4 cohort as well. Business-as-usual, therefore, is not the absence of any reading intervention; instead, it covers the use of other reading programmes and interventions. As discussed in the main report, this may partially explain the null result for both the Year 4 and Year 5 impact analysis. We note, however, that it extremely rare for trials in education to include a completely untreated control group. The condition for participation in this trial for schools allocated to the control group was that they did not implement an intensive or targeted reading intervention, and this was adhered to, but as the process evaluation in the main report shows, control schools implemented a range of non-targeted reading interventions.

Overall, we therefore conclude that while positively received by teachers and schools allocated to the treatment group, and implemented with an overall high degree of fidelity, the effects of Accelerated Reader on pupils' reading outcomes at Key Stage 2 were not high enough to differentiate it from results in business-as-usual schools.

References

Dong, N. and Maynard, R. A. (2013). 'PowerUp!: A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs', Journal of Research on Educational Effectiveness, 6(1), 24-67.

Sutherland, A., Broeks, M., Ilie, S., Sim, M., Krapels, J., Brown, E.R. and Belanger, J. (2021). Impact Evaluation Report: Accelerated Reader (Part I).

Sutherland, A. and Ilie, S. (2018). Accelerated Reader Statistical Analysis Plan. Available: https://educationendowmentfoundation.org.uk/public/files/Projects/Accelerated_Reader_SAP_2018.05.02__FINAL.pdf

Sutherland, A., Sim, M., Ilie, S., and Krapels, J. (2018). Evaluation Protocol. March 2018, Available: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Accelerated_Reader_Protocol_Update_2017.12.18.pdf

Appendix A: Security classification of addendum findings

OUTCOME: ***Key Stage 2 Reading*** *Score (KS2_READSCORE)*

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

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%	**5**				5
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%					
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity **[None required]**		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	Good balance at baseline
Threat 2: Concurrent Interventions	Low-moderate	Concurrent interventions reported in the main report but the description of these suggests minimal threat to internal validity
Threat 3: Experimental effects	Low	No evidence of contamination
Threat 4: Implementation fidelity	Low	Only a handful of schools discontinued implementation, and fidelity was high. Small association between fidelity and pupil outcomes. Logic model outlined in full in main report
Threat 5: Missing Data	Low	Attrition was low and balanced across trial arms. Nonetheless it would have been useful to include some analysis of missingness and consequent sensitivity analyses with missing data imputed
Threat 6: Measurement of Outcomes	Low	The measures used are standard, national assessments (KS1 baseline, KS2 follow up) and are marked blinded to allocation
Threat 7: Selective reporting	Low	No concerns here

- **Initial padlock score:** 5
- **Reason for adjustment for threats to validity**: No adjustment required
- **Final padlock score:** 5

Education Endowment Foundation

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn