



Education
Endowment
Foundation

Accelerated Reader

Evaluation Report and Executive Summary

February 2015

Independent evaluators:

Stephen Gorard

Nadia Siddiqui

Beng Huat See



Durham
University

The Education Endowment Foundation (EEF)



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;
- Encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust, as lead charity in partnership with Impetus Trust (now part of Impetus-The Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.



About the evaluators

The project was independently evaluated by a team from Durham University led by Professor Stephen Gorard.

Stephen Gorard is Professor of Education and Well-being, and Fellow of the Wolfson Research Institute at Durham University. He is a Methods Expert for the US government Institute of Education Science, member of the ESRC Grants Awarding Panel, and Academician of the Academy of Social Sciences.

His work concerns the robust evaluation of education as a lifelong process, focused on issues of equity and effectiveness. He regularly advises governments and other policy-makers, including oral and written advice to the House of Commons Education Committee every year since 2003. He is also a widely read and cited methodologist, involved in international and regional capacity-building activities, and used regularly as an adviser on the design of evaluations by central and local governments, NGOs and charities. He is currently an evaluator for the European Commission Directorate-General for Education and Culture, the Department of Work and Pensions, the Food Standards Agency, the Learning and Skills Information Service, and the Education Endowment Foundation. He is author of nearly 1,000 books and papers.

Contact details:

Professor Stephen Gorard

School of Education
Durham University
Leazes Road
DH1 1TA

p: 0191 334 8419

e: s.a.c.gorard@durham.ac.uk

Contents

Executive summary.....	4
Introduction	6
Methods	9
Impact evaluation results	15
Participants.....	16
Process evaluation results.....	22
Outcomes.....	25
Conclusions and implications	29
Future research and publications.....	31
References.....	32
Appendices.....	34

Executive summary

The project

Accelerated Reader (AR) is a whole-group reading management and monitoring programme that aims to foster the habit of independent reading among primary and early secondary age pupils. The internet-based software initially screens pupils according to their reading levels, and suggests books that match their reading age and reading interest. Pupils take computerised quizzes on the books they have read and earn AR points related to difficulty.

The evaluation of Accelerated Reader involved 349 pupils in Year 7 who had not achieved secure National Curriculum Level 4 in their primary Key Stage 2 (KS2) results for English (across four secondary schools). 166 pupils were randomly allocated to receive the intervention for 22 weeks in phase one during their first year at secondary school, with the other 183 acting as a control and then receiving the intervention in phase two. 10 pupils left their existing schools and were unable to provide post-test data despite follow-up. The intervention was organised either by taking pupils out of regular classes or making pupils stay after regular school time. The pattern varied between schools.

Key conclusions

1. Accelerated Reader appears to be effective for weaker readers as a catch-up intervention at the start of secondary school.
2. A well-stocked library with a wide collection of books banded according to the Accelerated Reader readability formula, and easy access to computers with internet connection, are the main requirements for successful implementation.
3. Pupils at very low levels of reading may not be independent readers and would need initial support from teacher to start reading books.
4. Schools can lead robust evaluations of their own planned interventions, under favourable circumstances, and with some advice and oversight from expert evaluators.

This was a different kind of study to the trials usually funded by the EEF. In 2013, four secondary schools applied independently to EEF for funding to set up a programme for AR, and simultaneously evaluate its impact in their own schools. Each application was deemed too small in scale to run a successful evaluation of the programme, but if the schools involved were to co-operate then the scale would be sufficient for an 'aggregated' efficacy trial. Efficacy trials seek to test evaluations in the best possible conditions to see if they hold promise.

Grants were given directly to schools to run the trial, including randomisation and testing, as well as implementing the intervention. Each school ran a small trial of AR in isolation, and made all of the relevant evaluation decisions such as randomly allocating pupils to the control or intervention group.

Researchers at Durham University were assigned as independent evaluators for this trial. Their roles were to advise the school leads on the process of conducting research, randomisation and testing, and to aggregate the eventual results from all schools. Online testing was done by an external company (GL), although delivered by the schools themselves, and the independent evaluators were provided with access to the results. The schools managed the administration of the online tests.

What impact did it have?

The intervention group exposed to Accelerated Reader recorded higher literacy scores than the control group, using the GL Assessment New Group Reading Test. The overall effect size of +0.24 is the equivalent of approximately 3 months of additional progress in reading age after 22 weeks. The evaluation also indicates a positive impact for FSM-eligible pupils although as the numbers of pupils in the sub-groups are smaller, these findings are less secure than the overall finding. This is independent of the precise age, sex, first language, ethnicity, and special education needs of the pupils. If replicated, the intervention has a good chance of delivering similar results in similar conditions.

The evaluation team observed the implementation of the project and considered it to be generally well conducted, and attractive to teachers and pupils.

On the basis of this aggregated trial, it appears that schools are able to lead evaluations of their own practice given the two conditions that applied in the situation described here. It represents a more 'real world' approach to evaluation, akin to schools buying an intervention and implementing it themselves with no external support. However, the findings here suggest that it is necessary for schools to be trained and supported by independent experts, and it is helpful that the developers of the intervention are not involved at all.

How secure is this finding?

The evaluation was set up as a 'waiting list' efficacy trial with the control group receiving AR after the intervention group and after both groups had completed the post-intervention test. The control group initially carried on with 'business as usual' and were not given access to the programme.

Four secondary schools participated and 349 pupils were involved in the study (with 10 dropping out). The schools participated from different regions of the country. This participation was based on school leaders' initiatives to implement the intervention and their willingness to become a part of this research project. However, the trial itself does not demonstrate that the findings hold at a similar scale in all types of schools or for other age groups.

A number of statistical techniques were applied to check whether the results are due to chance or bias caused by dropout. The security rating is considered to be three padlocks.

How much does it cost?

The direct cost to schools is for the annual licence to use Accelerated Reader for each individual pupil. A minimum subscription rate for 50 pupils is £450, or £9 per pupil. One day teacher's training cost is included in the subscription licence and the schools also have year-long access to a free hotline telephone service.

Group	No. of pupils	Effect size	Estimated months' progress	Evidence strength*	Cost**
Intervention vs Control (all pupils)	339	+0.24	+3 months	🔒🔒🔒🔒🔒	£
Intervention vs Control (FSM)	115	+0.38	+5 months		£

*For more information about evidence ratings, see Appendix 1 in the main evaluation report. Evidence ratings are not provided for sub-group analyses, which will always be less secure than overall findings.
 **For more information about cost ratings, see Appendix 2 in the main evaluation report.

Introduction

Intervention

Accelerated Reader (AR) is a widely used web-based intervention produced by Renaissance Learning Company. It monitors and manages pupils' reading practices and encourages them in independent reading. AR identifies pupils' initial reading levels through the Standardised Test for Assessment of Reading (STAR) test which is a baseline assessment. The scores gained also suggest a suitable range of books which can be easily read by each pupil. Teachers and pupils set reading goals to be achieved in independent reading time. After completing the reading targets the pupils take an AR quiz based on the selected book. AR uses an Advantaged/TASA Open Standard (ATOS) readability formula that measures the complexity of the text. The range of books suggested by AR have ATOS book levels and an interest level—a qualitative measure of text complexity (the intervention is explained in more detail below). AR has a web-based repository of 150,000 quizzes based on fiction and non-fiction books. This technology-enhanced reading programme works on the principles of motivation through gaining reading points and rewards for achieving reading goals. AR is designed to encourage regular book reading that increases what AR refers to as each pupil's zone of proximal development (ZPD) in reading through stretching their reading abilities after every quiz. AR quizzes allow teachers to monitor pupils' performance and give feedback with appropriate rewards. AR is used as a whole-class reading intervention to motivate pupils to read in an atmosphere of competition and ultimately develop a lifelong habit of book reading.

Background

In the UK, over 2,000 schools are using AR on a regular basis, which means that well over 400,000 students are reading what is recommended in AR or what AR supports through quizzes (Topping 2014). The intervention has evolved over a period of 34 years and many schools across the world are using it. The AR software and website are user friendly and are regularly updated for improvement in quality and access. However, the implementation of AR at such a large scale is not solely based on the existing evidence, which is represented as largely supporting the effectiveness of AR in pupil's attainment. It is possible that increasing implementation is also due to its regular enhancement and adaptation to make use of new technologies available to schools.

There have been several small and weak studies reporting success for AR. For example, Scott (1999) involved only 28 pupils (from a larger number approached), had unbalanced groups at the outset, and the report is unclear how the cases were allocated to AR or not. There are also studies showing no effect or even negative impact. Mathis (1996) compared the progress of 37 AR pupils (from a larger number approached) over one year with the whole year cohort, using the Stanford Achievement Test. There was a large negative effect size for AR pupils on reading comprehension.

AR is one of 24 effective reading interventions listed by the What Works Clearinghouse (WWC) (IES 2008). According to the findings of their systematic review, AR has no visible effect on reading fluency, a mixed effect on comprehension and a possible positive effect on reading achievement (WWC 2008). These results are based on only two studies that fulfilled WWC standards for systematic reviews.

In one study, 45 teachers (with 572 K-3 grade students, aged 11-14, in 11 schools) were randomised to teach using AR or another commercially available reading programme (Ross, Nunnery and Goldfeder 2004, 2006). The results were reported after one year. The authors reported what they termed a 'significant' impact on reading comprehension using the STAR reading test, but WWC recalculated and reported that they found it was not statistically significant, although the effect size was over 0.25. Similarly, there was no significant effect on general reading achievement based on the STAR Early Literacy test for each year group, but the overall effect size was over 0.25. Baseline

equivalence in terms of reading ability was not established before the intervention. The study is relatively small, and is not based on individual randomisation. Also, the STAR tests are part of the AR programme rather than being independent assessments (Krashen 2007). The second study involved only 32 grade 3 students attending one school in the Pacific Northwest (Bullock 2005). They were individually randomised to receive 90 minutes of AR reading or not per week for 10 weeks. At the end there was no difference in terms of oral reading fluency. As above, the author reports no 'significant' effect on reading comprehension using the STAR reading test, but the effect size is greater than 0.25.

Brooks (2007) conducted a meta-synthesis of studies involving reading interventions for pupils with reading difficulties. The review reported on 14 reading interventions used at secondary school level, including AR. The meta-synthesis for AR found 47 studies conducted mostly in the US, but only two were selected for inclusion (Vollands et al. 1996, 1999). According to the results, AR produced positive effect sizes of 0.55 and 0.21 respectively. However, the sample sizes were only 27 and 12 pupils, which is too small to draw conclusions on the effectiveness of the intervention. It is not clear how the groups had been created, nor whether baseline equivalence was established between the treatment and control groups.

A more recent study was conducted with 108 primary age pupils from two schools in the US (Nicholas 2013). Pupils were randomly allocated to AR in one (treatment) school, and to a literacy plan in the other (control) school. After one year, there was no difference (or rather a small negative effect of -0.02) between the two groups in terms of the Standards of Learning (SOL) test.

There have been some larger studies, all with deficiencies of design (and mostly abusing the concept of 'significance' with non-random samples and non-random allocation to groups). For example, Paul et al. (1996) had a large sample of 6,000 schools in which 58 percent were non-AR comparison schools in similar geographic locations. According to official records, the schools having access to AR had better pupil attendance records and reading performance scores compared to the schools not using AR. A similar study, based on schools that had already adopted AR or not was conducted by Peak and Dewalt (1993), who reported greater success for the AR group at both primary and secondary levels. Pavonetti et al. (2000) developed a test to measure the quantity of books read, called the Title Recognition Test (TRT). Pupils were asked to mark the book titles they had read and in order to check if they were guessing rather than giving true responses, some foils for book names were added in the list (25 titles were actual books and 16 were foils for book names). AR claims that pupils' quantity of book reading increases if they use AR in schools. This claim was assessed using a school-level matched comparison design, with 10 secondary schools. There was no difference in the quantity of reading between the pupils using AR and those not using AR (reported mean difference was -.008). Goodman (1999) involved 282 pupils in one US secondary school with no comparator, and claimed a positive gain based on pre- and post-test only.

Therefore, there is a considerable research base on AR, making it one of the most researched interventions in which reading is practised through online resources. Prior research has mainly been carried out in the context of US schools, not the UK. Also, the evidence on the effectiveness of AR on attainment is mixed, with much of the research small, high attrition, using AR-led measurements, or based on weak research designs. There is therefore a role for a UK-based trial based on a larger sample, true random allocation, baseline equivalence between the two groups, an independent test of attainment, and complete process evaluation. This trial is described in the rest of the report.

Objectives

The purpose of the impact evaluation, designed as an RCT, was to assess the strength of any benefit from using AR for 20 weeks on the literacy performance of Year 7 pupils who had not achieved secure Level 4 in KS2 English at their prior primary schools. In addition, the trial should lead to an estimate of impact for children with indicators of possible disadvantage, such as being eligible for free school meals.

The parallel process evaluation, based on observation and interviews, looked at how closely schools and teachers implemented the intervention, and yielded formative suggestions both for implementation and for the extent to which the intervention could be implemented more widely. It also looked at what happened in the control schools, and how fairly the post-testing was conducted.

In addition, and specific to this aggregated trial, the project was intended to help assess to what extent groups of schools are able to implement RCTs without close control by independent and expert evaluators, and at which stages they would need most assistance.

Project team

AR is a commercially available reading programme delivered by Learning Renaissance Company Inc. Neither the developers nor the AR authorised company had any involvement in this intervention and its evaluation. The intervention was conducted independently by volunteer schools, and the evaluation process was supervised by the independent evaluation team. A big thank you goes to all the schools and teachers involved in this research, who ran the project within their own schools and contributed a great deal to the evaluation. Thanks goes to Epping St. John's Church of England VC school, who indicated they were happy to be named in the report. Three other schools based in Liverpool, Thornaby and East Sussex were also involved in the trial.

Methods

Trial Design

This evaluation is a two-arm, school-led, post-test-only design, randomised controlled trial in which four schools participated. The target group of 349 pupils across four schools was selected on the basis of their prior KS2 scores, and pupils at Level 4c and below in English KS2 were selected with an intention to treat. The AR intervention itself was carried out from September 2013 to February 2014 (nearly 20 weeks).

After selection of the target group, the schools themselves randomised the target group into two arms and began the intervention in September 2013. Baseline equivalence was established on the basis of prior KS2 scores. The evaluators advised the school leads on the process of randomisation, and checked that the two arms of the trial were balanced in terms of KS2 scores in English. The treatment group received AR first and the pupils in the control group continued their regular school lessons and would receive AR in a second phase after the post-test for the first phase.

This is a waiting list design. This design is ethical as it does not deprive anyone of the treatment (given that the intervention was to take place anyway). It is less likely to demoralise pupils once they know the results of allocation than had there not been a wait list. The post-test design is cost-effective in terms of cost of the evaluation as only one test is involved after the intervention.

The developers of AR were not involved which gave flexibility to the school leaders to manage AR according to the available space, time and need of the schools. School leaders being in charge of the intervention and trial helped reduce the attrition as they took ownership of the project supported by the evaluation team.

As this is a school-led trial, the school leaders sought parental opt-out consent for their children's involvement in the intervention and evaluation. The schools used their usual communication strategies such as letter or email to parents/guardians. This procedure was followed by the school leaders before randomisation in September 2013.

Ethical approval for the role of the independent evaluators came through the Durham University Ethics Committee.

Testing

The New Group Reading Test (NGRT) is the third edition of Group Reading Test (GRT) developed by GL Assessment and the National Foundation for Education Research. The age appropriate level NGRT 3A was selected which is suitable for the age group 10-13 (Years 5 to Year 8) (<http://www.gl-assessment.co.uk/products/new-group-reading-test/test-detail>). The test items are 20 sentence completions, and four short passages for context comprehension and reading comprehension. The areas of assessment are as follows:

- Vocabulary
- Grammatical knowledge
- Inference skills
- Ability to recognise
- Authorial intent
- Deduction skills

The test has an online set-up and it has no time limits for completion of the test. However in practice, test completion was no more than 45 minutes (one school lesson). The test difficulty level adapts according to pupil's initial responses. The test was agreed as suitable by all parties.

Eligibility

The four schools themselves were self-selected since they were the applicants for funding from the EEF for the intervention. Two schools had already used AR but wanted to build more capacity and resources. All four secondary schools independently proposed using AR, which would have led to four very small trials. This was considered an opportunity by the EEF to run a school-led trial on four different sites and appoint an independent evaluation team to aggregate the results.

The intervention was targeted at all pupils in Year 7 who had arrived in secondary school who achieved below secure Level 4 in KS2 English. Operationally this meant achieving Level 4c or below. Parental opt-out consent for their child to participate in the trial was handled by the schools. The evaluators were not involved in this stage. The opt-out consent letters were sent only for those pupils who were selected in the target group. No pupil was excluded from the programme once identified for randomisation.

Intervention

AR is a networked computer-based management programme intended to encourage pupils in independent book reading. The AR programme allows teachers to monitor pupils' reading levels and progress, and based on this information the teacher's role is to support pupils in making an appropriate selection of books for reading and to motivate them in achieving advanced reading levels. AR has the following main features:

Standardised Test for Assessment of Reading (STAR)

- ATOS readability formula and book selection
- Independent reading
- AR quiz
- TOPS report and AR points

Standardised Test for Assessment of Reading (STAR)

STAR is a 20-minute screening test that determines each pupil's 'optimal' level of reading comprehension. The test includes vocabulary-in-context and other skill-based items. It is an adaptable assessment system that changes the question choices and level of challenges according to the pupil's prior responses. The STAR scores pupil's reading ability and generates a diagnostic report that includes percentile rank, National Curriculum Level in reading, reading age, estimated oral reading fluency and Zone of Proximal Development (ZPD: maximum ability to read and understand a book of a certain difficulty level). The diagnostic report also gives recommendations to the teacher on how to support the pupil for further improvement in reading. STAR can be conducted repeatedly and periodically to monitor pupil's progress. It is recommended on the Renaissance Learning Inc. website that teachers should conduct STAR three to five times in a year to follow pupil's gradual progress.

Advantaged/TASA Open Standard (ATOS) readability formula and book selection

ATOS is a measure of text difficulty created by Renaissance Learning Inc. The readability of a book is calculated taking into account the word count, average sentence length, average word length and word difficulty. There are over 160,000 books (fiction and non-fiction) available in the AR programme, allotted to bands on the basis of the ATOS readability formula. ATOS indicates the level of challenge in any book to be matched with a pupil's reading (ZPD) and their areas and levels of interest. ATOS

measures only the readability level of books and does not take into account the literary merit, thematic construction, quality and complexity of ideas or maturity of the content. Areas of interest based on pupil age are suggested along with ATOS book levels in order to help make an appropriate book selection.

Independent reading

Once an appropriate book selection has been made, pupils are given time in school to read independently. AR recommends teachers motivate pupil to read regularly, and finish reading the selected book promptly. AR suggests 30 to 60 minutes of independent reading time every day.

AR quiz

There are around 156,000 quizzes in AR. These reading practice quizzes assess pupils' comprehension of the specific books they select to read. The format is generally multiple choice items that ask factual and inferential questions from the book. The quizzes are computer based and can be taken on laptop and tablets. Each pupil gets an individual login and password to have access to AR and complete the quiz. It is recommended that pupils take the AR quiz within 48 hours of finishing the book.

TOPS report and AR points

As soon as the quiz is completed the AR generates a TOPS report (The Opportunity to Praise Students) showing the results. The performance is intended to be monitored by a teacher, and if the pupil scored lower than 60% repeatedly then the teacher needs to both recheck the book selection and make different choices according to the ZPD. Pupils achieve AR points every time a quiz is passed, calculated on the basis of the ATOS readability level and word count of the book. The formula introduced by the developers of AR intervention (Paul et al. 1996) is:

- $AR\ Points = (10 + ATOS\ readability\ level) \times (words\ in\ book / 100,000)$

The teacher can set an AR point goals for each pupil or for the whole-class group. The computer programme flags issues if pupils are not attaining the set targets or just selecting books to attain points rather than increasing the ZPD levels. Teachers are recommended to be innovative in giving rewards on achieving the targets such as certificates of achievement, gift vouchers, club membership, and announcements in the school assemblies.

As described above, AR is a developed intervention with clear methods of implementation and a process of regular assessment. The Renaissance Learning Inc. website is an important resource for teachers to consult and implement this intervention. There are various applications that Renaissance Learning Inc. regularly add in the AR programme and recommend teachers to integrate into AR. There are also staff training workshops offered by Renaissance Learning Inc.

As observed in the schools participating in the trial, the ones which have been using AR previously were more practised in using the full range of features in AR. However, the two schools new to the AR intervention adopted the basics and slowly began using extra features.

The intervention relies on technology (computers or tablets), the internet and a full library. The schools bought more books with funding for this trial. Two of the schools purchased tablets to make the AR quiz time a more fun activity for pupils, with enough for each child in a small group to use one each. In one of the schools pupils were doing their independent reading every day for 40 minutes after school time. In two of the schools pupils were taken out of French and other language lessons, while in the other they were mainly taken from English. The library space was mainly used for the independent reading time and for taking the AR quizzes. Pupils were also encouraged to take the books and read at home.

Control group activity: Pupils in the control group continued the usual school activities. There was no chance of contamination because pupils in the control group had no access to the AR programme. However, the material purchased for the intervention was made available to everybody. The books, computers, iPads and tablets were not restricted to treatment pupils only. The pupils in the control group were assigned into a waiting list and the schools started AR with them any time after 20 weeks of the trial.

Sample size

This is a school-run trial in which the sample was determined by individual schools. Initially the schools proposed a rough estimate of 450 pupils across all four schools which would traditionally be considered sufficient to detect a minimum effect size of around 0.27 (Lehr's approximation). The final figure was dependent on the KS2 results of the incoming cohort, and the eventual intake of pupils by these schools. The eventual sample included 349 pupils from four secondary schools located in different regions of the country. The selection was based on KS2 scores and pupils selected in the target group were at Level 4c and below in KS2 English. The schools selected their target groups on arrival in Year 7. Three schools conducted individual pupil randomisation. One school randomised into treatment and control group by classes. The school that randomised classes had 119 pupils identified in the target group and they were already spread across different class groups (i.e. the usual classes for that school). The school claimed that it was not practically possible to individually randomise the pupils and conduct the intervention. The evaluators ran a separate group analysis for this school and found that the groups were well balanced in terms of KS2 scores before the intervention began.

No school dropped out from the trial and the pupil attrition rate was considerably lower than usual in large scale trials. By the final analysis, 8 pupils had dropped out or could not be included in the analysis (having moved abroad, for example). The total attrition rate is just over 2% of the total sample identified as a target group. The evaluation team along with the school leads tried to approach all of the missing pupils to complete the post-test. There were 6 pupils who completed the test as a result of following them to different schools.

Randomisation

This was an aggregated trial, really consisting of four separate one-school evaluations. In a meeting before the trial began in September 2013, school leaders were given a tailored workshop on conducting an RCT and the process of randomisation, led by the lead evaluator. In this workshop, methods of randomisation were suggested by the evaluators and discussed with the school leads. The emphasis was laid on keeping the process fair and unbiased by completing the randomisation in one single phase. The timing and methods were also co-ordinated as much as the individual circumstances of each school allowed. The schools sent their target groups and the background characteristics of eligible pupils to the evaluators as a record before the schools randomised the pupils into treatment and control groups. The schools used simple methods such as taking the enclosed name from a pot which contained the enclosed name chits for the pupils. One school used a computer pseudo-random number generator and another used a shuffled pack of cards as demonstrated in the workshop. In each such method, the number of eligible cases and the number in each group were pre-determined. For example, schools were shown that there could be one card for each pupil, and red cards might mean treatment and black cards control. The cards were mechanically shuffled and then laid out in sequence next to the list of pupil names. The schools then sent the list of allocated groups to the evaluators to record, and started the intervention in September 2013. The evaluators did not observe the randomisation.

The process of randomization appeared fair in this trial, despite taking place in each school separately. The evaluators also ran a simple analysis of the groups to check if the allocation was balanced in terms of KS2 scores in English.

Analysis

The total number of cases (349) was determined by the number of schools funded by EEF, and the number of pupils identified as eligible in each school. This is regarded as sufficient for an efficacy trial.

The analysis used here is 'intention to treat', meaning that all pupils originally randomly allocated were tested and their outcomes analysed, regardless of the time actually spent on the intervention. The impact of the trial is represented by the effect size (Hedges' g) for the post-test only results of New Group Reading Test (NGRT 3-A). These were calculated both for the overall reading score and the standardised age score (SAS), yielding equivalent results in terms of effect sizes. Further analyses include finding the equivalent 'effect' sizes for sub-groups such as FSM pupils, boys and girls. These analyses are not dealing with true effect sizes from a trial since the sub-groups were not randomised as such.

A simple multivariate regression model was also created using the overall reading score as the 'dependent' variable. The possible predictors added to the model included the prior KS2 attainment in English, the pupil age in months, the known pupil characteristics such as sex, FSM, SEN, ethnicity and EAL, and a flag variable representing their treatment group, which entered into the model only after all other variables had been accounted for.

Readers may wish to note that significance tests and confidence intervals are not presented in this report. These do not work as intended (Carver 1978), are almost always misinterpreted (Watts 1991), and can lead to serious mistakes (Falk and Greenbaum 1995). Above all, they take no account of sample quality or attrition (Lipsey et al. 2012), being predicated on complete random samples of a kind never encountered in real-life research (Berk and Freedman 2001). This kind of explanation should no longer be necessary in a report; rather those who still use such approaches must explain what their cited probabilities could possibly be probabilities of.

Process evaluation methods

The light-touch process evaluation was conducted in co-operation with the school leaders who were mainly in charge of the intervention programme in their respective schools and at the same time conducting the trial for evaluation purposes. The evaluation team participated in the process evaluation as a second party to ensure that the school leads were following the protocol of the RCT. The information on the process reported here is twofold: the process of intervention as observed in the actual school settings and the management of the trial by the school leaders.

Two meetings were conducted with the school leaders in which the protocol of conducting the RCT with advice from evaluators was discussed. The first meeting was in a workshop format in which the lead evaluator presented the process of a RCT, and the required preparation for conducting the intervention along with the trial. The second meeting was conducted before the testing phase in which the evaluators explained the standard criteria for conducting the test and discussed with the school leaders how to interpret the results of the tests.

The project leaders arranged their own internal AR training for their staff members, implemented and monitored the intervention and collected the formal records and the views of pupils. The evaluation team members observed this training in one of the four schools. The evaluators visited every school, and also observed the intervention in progress by sitting in a range of lessons in two schools at two time intervals: one in the beginning when the intervention started and the second when the intervention was nearing its completion phase.

Renaissance Learning Inc. website materials on AR and documents relating to the intervention were read to help understand the aims and methods of AR. In the schools, evaluation team members conducted face-to-face interviews with staff, pupils and project leaders. These interviews were usually conducted without a formal schedule, and arose as the situation allowed. The observations of staff training and implementation of the programme in action were as simple and integrated and non-intrusive as possible. The schedule of visits was agreed with the school leaders and some interviews were arranged at that time. The interviews and field notes were part-transcribed and shared between the evaluation team. All schools' staff members' and pupils' names have been anonymised.

The process evaluation was useful in assessing fidelity to treatment. The perceptions of school leaders, staff members and pupils provided indications of any resentment or resistance to the programme. They are also useful in identifying potential issues or barriers which could be addressed for any future scaling up.

Impact evaluation results

Timeline

July – August 2013

Schools identified pupils for the target groups as soon as they received the KS2 data from the primary schools. A meeting was held in which the evaluators presented a workshop on conducting an RCT. In this time period school leads also made preparation for the intervention such as adding resources in the libraries with books and other technological resources. The list of pupils in the target group was shared with the evaluators, along with the background characteristics of the pupils.

September 2013

School leaders conducted staff training, sought opt-out consent from parents, and randomised the target groups. The schools started the intervention as soon as possible. The evaluators conducted a simple analysis of the randomised groups to see if there was any wide variation in group allocation in terms of KS2 results.

October 2013 – January 2014

The intervention continued in these months except during the Christmas break in December 2013. The evaluators conducted visits in the participant schools for the process evaluation.

February 2014

The second meeting was conducted with the school leaders and the post-tests were conducted in the last week of February 2014. The test completion rate was nearly 77% in the first week of the specified time for the test window. By the second week 95% of the pupils had completed the test. Fifteen pupils were followed up as missing.

March – May 2014

Some schools completed the testing in the first week of March 2014. However, there were some pupils who needed following to complete the test in other schools. Their new schools were approached with the help of school leaders. The context of the project was explained to the schools and it was requested that they conduct the NGRT for the pupil. The communication with the new schools was all done via email and telephone. Generally these schools were found very cooperative in conducting the stand-alone test.

June 2014

Pupil background data provided by schools was updated and outcome measures were analysed. A synthesis with the process evaluation took place simultaneously as test results were analysed.

Participants

School characteristics

Four individual secondary schools proposed to conduct the intervention and evaluation of AR. There was no developer involved in any of these four proposals so it was decided that the schools should run the trial with advice from evaluators, and aggregate their results for the final analysis. The schools involved were located at four different locations in Essex, Sussex, Thornaby and Liverpool (Table 1). In terms of their most recent OFSTED inspection one school received 'Good' while the other three had 'Required Improvement' grades.

Table 1. Summary characteristics of the participating schools

Category	Phase	No. of pupils	% SEN	% FSM	% EAL	Average KS2 point score 2013	OFSTED Effectiveness 2012–2013
Academy sponsor- led	Secondary	445	17.5	43.6	7.9	27.1	Requires improvement
Foundation school	Secondary	1183	19.4	60.3	8.7	26.8	Requires improvement
Voluntary controlled school	Secondary	749	6.9	15.6	2.9	29.0	Good
Academy sponsor- led	Secondary	584	8.4	18.2	0.9	26.1	Requires improvement

All schools were urban, mixed, secondary stage schools. The proportion of disadvantaged pupils was high in all four schools. The selection of 349 eligible pupils was made from scrutiny of nearly 2,500 pupils enrolled across the four secondary schools. In the final analysis, 8 pupils dropped out and could not be followed for the post-test due to reasons such as: left country, home schooled, did not attend the school after the first initial days, and did not provide details of the destination school. According to the drop-out figures in the groups, 2 pupils were in the control group and 6 pupils were in the treatment group. The attrition is 2.3% of the total sample selected for the study. The average KS2 scores of those who dropped out in treatment and control groups was about the same, and neither unusually high or low, given the eligibility criteria.

Pupil characteristics

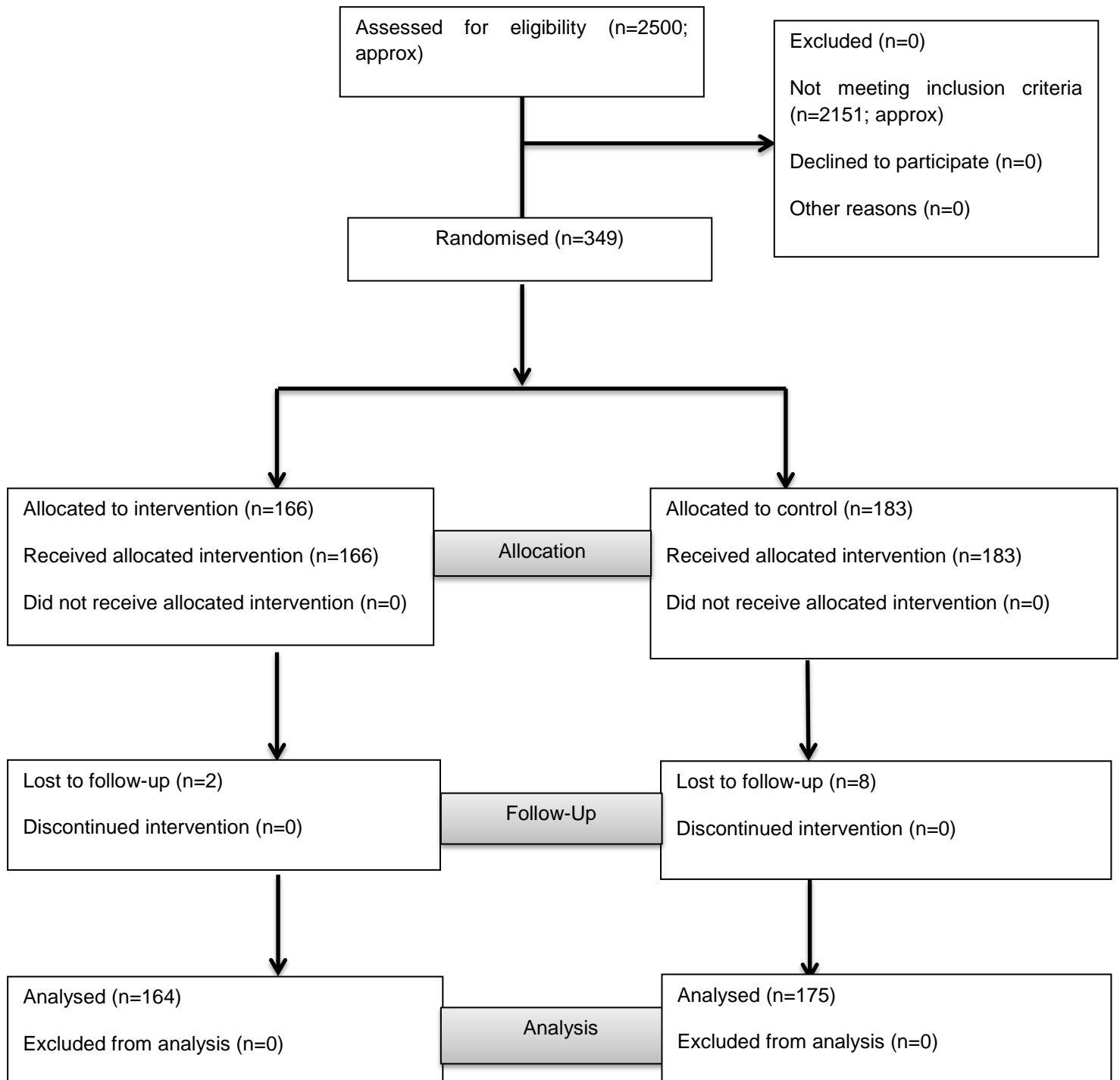
The pupil characteristics in the achieved sample are provided below (Table 2). No data are missing. In order to keep the pupil selection and group allocation procedures unbiased the following characteristics were assessed only after the randomisation procedure. The characteristics of the sample indicate that a majority of the pupils are in some major categories of the disadvantaged groups.

Table 2. Number of pupils in each group with the listed background characteristics

	AR group	Control	Percentage of total sample
Male	100	83	52
Female	83	83	48
FSM	62	61	35
Non-FSM	121	105	65
SEN	46	36	23
Non-SEN	137	130	77
EAL	7	6	4
Non-EAL	176	160	96
Non-White	29	13	12
White British	154	153	88

Participant flow diagram

All schools remained in the study throughout. All pupils who left their original schools were followed up, and all but 8 completed the post-test. The attrition rate is very low at 2%. The randomisation took place in schools for all eligible pupils, and the number in the treatment group represents the total number that schools judged that they could cope with in the first phase.



Outcomes and analysis

In terms of their prior KS2 English points, the randomisation was successful in creating balanced groups at the outset (Table 3). This suggests that a post-test only analysis is appropriate with no further consideration of prior attainment (Gorard 2013), which has the added advantage of including the four pupils with missing KS2 data.

Table 3. KS2 points in English, all pupils

Arm	N	KS2 points	Standard deviation	'Effect' size
AR	174	26.67	4.21	0
Control	161	26.64	4.00	-

Note: three control and one treatment pupil from the final analysed groups are missing KS2 scores. This can occur when they move from a school outside England.

Whether considered in terms of the NGRT overall reading score or the SAS, it is clear that AR has had a modest impact on the treatment group (Tables 4 and 5). The treatment group is ahead of the control by about one quarter of a standard deviation, equivalent to around an extra three months' reading age.

Table 4. Overall reading score, all pupils

Group	N	NGRTA	Standard deviation	'Effect' size
AR	175	327.1	51.4	+0.24
Control	164	315.3	46.6	-

Table 5. Standardised age score, all pupils

Group	N	NGRTA	Standard deviation	'Effect' size
AR	175	98.0	14.1	+0.26
Control	164	94.5	13.0	-

The results in Tables 4 and 5 are relatively unlikely to be due to either chance or bias caused by the dropout of 10 cases. For example, taking the mean post-test score for the treatment group and adding one standard deviation creates a score that would be counterfactual to the control group. It would take over 32 such scores added to the existing control group to eliminate the effect size reported in Table 4.

Although the next results do not have the force of a trial, because FSM-eligible were not randomised to groups as such, the evidence is that AR is at least as effective as overall for those pupils eligible for FSM (Tables 6 and 7). Using AR does not widen the gap between pupils designated as poor and the rest, and may well act to reduce it. It must be recalled that this intervention was applied selectively, and the results are based on only pupils originally attaining Level 4c or below at KS2 English. Therefore, there is strong evidence overall that AR improves reading for new Year 7 pupils with weak reading skills and habits, and that it does so especially for those eligible for FSM.

Table 6. Overall reading score, FSM-eligible pupils

Group	N	NGRTA	Standard deviation	'Effect' size
AR	56	319.9	42.4	+0.38
Control	59	303.9	41.1	-

Table 7. Standardised age score, FSM-eligible pupils

Group	N	NGRTA	Standard deviation	'Effect' size
AR	56	95.6	12.1	+0.41
Control	59	90.8	11.2	-

Interestingly, AR seems to be more effective for girls than boys, although the results are positive for both (Tables 8 to 11).

Table 8. Overall reading score, boys

Group	N	NGRTA	Standard deviation	'Effect' size
AR	95	321.0	53.9	+0.17
Control	82	312.2	50.5	-

Table 9. Standardised age score, boys

Group	N	NGRTA	Standard deviation	'Effect' size
AR	95	96.4	14.3	+0.18
Control	82	93.9	13.4	-

Table 10. Overall reading score, girls

Group	N	NGRTA	Standard deviation	'Effect' size
AR	80	334.3	47.6	+0.35
Control	82	318.4	42.5	-

Table 11. Standardised age score, girls

Group	N	NGRTA	Standard deviation	'Effect' size
AR	80	99.9	13.7	+0.37
Control	82	95.0	12.7	-

A slightly different way of looking at the same figures is provided by a simple multiple regression model, based on the linear correlation between prior attainment and post-test scores. The R value for the model 'explaining' variation in the overall reading score at post-test is 0.64 (Table 12). This model is not, in itself, any test of causation but it does provide a context for the strength and importance of the intervention in relation to pupil characteristics. The possible predictors included the age at test (to adjust for any variation within the testing window, and the more precise age of each pupil), FSM,

SEN, EAL, ethnicity and sex of each pupil. None of these yielded substantial standardised coefficients (Table 13). The coefficient for prior KS English points was 0.59, because the post-test was heavily correlated with prior reading ability. The coefficient for the flag variable (treatment or control group) was 0.13, suggesting an effect from the treatment over and above an individual's prior attainment. Note that this modelling does not have the force of a trial, and can often 'soak up' otherwise meaningless variance in fitting the best possible model. Nevertheless, the result agrees with the direction of the headline finding, and suggests an impact from the intervention even in this worst case (for the intervention) scenario.

Table 12. Variation explained (R) in two-stage regression model, using two possible outcomes

	Post-test outcome
Step 1 – background and prior attainment	0.63
Step 2 – intervention or not	0.64

For completeness, Table 13 presents the coefficients for all explanatory variables retained in the model. The largest of these by some way is the pre-test score as represented by prior attainment at KS2.

Table 13. Standardised coefficients for the regression model in Table 12

	Post-test outcome
FSM	-0.06
Sex (female)	+0.02
SEN	-0.08
EAL	-0.04
Ethnicity (White UK)	-0.05
Age at test	+0.03
KS2 points	-0.59
Step 2: Treatment (or not)	+0.13

Cost

The costs of AR include an annual subscription to the online resources that are available on the Renaissance Learning Inc. website. The schools buy an annual licence for each pupil involved. The cost varies depending upon adding extra feature to the AR programme. However, a minimum subscription rate for 50 pupils is £450 or an average of £9 per pupil per year. This gives access to the STAR reading assessment, 25,000 quizzes and features such as TOPS reports, other diagnostic reports, and book finders. One day teacher's training cost is included in the subscription licence and the schools also have access to a free hotline telephone service for instant information.

The cost for buying books and technology varies among schools. The schools in this project have used a major part of the funds in purchasing laptops, iPads, tablets and books for the libraries. The cost for extra teaching assistant staff (for book banding and making book inventories) was also covered in the project budget for individual schools.

Process evaluation results

Implementation

What are the necessary conditions for success of the intervention?

AR is a structured intervention that targets reading improvement through feedback, competition, reward and motivation strategies. The main feature of AR is providing pupils with the choice of books they are interested in reading and within their level of understanding. The schools in the trial increased the number of books in the libraries and banded the book stocks according to ATOS readability level. The schools using AR also needed to have regular subscription to the Renaissance Learning Inc. website through which pupils could take the STAR tests and quizzes, and teachers could find books or download pupils' diagnostic reports. In this subscription licence teachers had access to technical support and they could also exchange information with other AR schools.

AR training in the schools was given by professional trainers. The training included a tour of the Renaissance Learning Inc. website through which pupils and teachers could access AR resources. The training covered all aspects of STAR testing and TOPS reporting with an explanation on how to interpret the results. Teachers, TAs and literacy coordinators who were planning to implement AR intervention attended the training. In all of the four schools a total of 15 staff members received training. The majority had no previous experience of using AR.

The schools organised independent reading time according to the availability of time slots in the school timetable. In two of the schools, 40 minutes were given every weekday in which the pupil in the treatment group would do independent reading in schools. The other two schools arranged three days in a week in which pupils did independent reading for an hour each day in the school. Pupils were encouraged to take the books home for reading. The independent reading in the schools was usually done in groups and mostly the library was used as the space for reading.

AR recommends that teachers monitor and advise pupils in book selection according to their ZPD level. There is no element of teachers tutoring pupils who are at very low reading levels and struggle to read with comprehension. In fact the intervention involves no teaching at all. AR does recommend teachers let pupils read aloud in a small group with same-age or cross-age peers. Teachers who were supporting pupils at a very low reading level were using the independent reading time for loud reading and one-to-one tutoring.

The school leaders developed reward systems to encourage a culture of reading in their schools. This included setting up an after-school reading club, celebrating millionaire readers (who had read one million words), using symbols/medals/ badges for pupils who achieved the targets, a wall of fame for those who achieved 100% in quizzes, and Amazon gift vouchers to purchase books on completing the AR targets.

Barriers to delivery

AR has made improvements from the older version which involved downloading a complete programme from CD; this was insecure, restrictive and unadaptable. The internet website is more secure, adaptable, interactive and innovative. However, AR still depends on hardware and software resources. Lack of access to these resources can be a problem. Some schools in this project purchased tablets which could be useful in utilising classroom spaces rather than libraries and computer rooms. However, in one school the complete set of tablets was stolen in a burglary, which caused a slight problem in the AR sessions, but the school soon adjusted by using computers in the libraries.

The books that have an accompanying AR quiz cover a wide collection but it is necessarily limited. Teachers could request Renaissance Learning Inc. to include a book of interest if it was not available already with a quiz but this takes some time. Access to quizzes is limited. Not all quizzes are available to pupils if the school has just the basic subscription to the Renaissance Learning Inc website (and schools received funding from EEF for only the basic subscription). This means teachers and pupils had to select books from those for which the pupils had access to complete the quiz.

There is a possibility of 'cheating' for the AR points because the material on the books for reading tasks is widely available on the internet. Pupils could choose short ways to read summaries of the books or achieve basic information about the book through internet discussions and book clubs. It is not clear that the AR quizzes are challenging enough to assess if pupils have read independently and they had complete comprehension of the book. However, there is no direct suggestion that anything like this happened in this trial.

Pupils with low reading levels seemed to need direct instructions for reading comprehension, but the AR intervention did not suggest any instructional strategy. Some pupils could also select books that were too easy or at least insufficiently challenging for them.

One of the school leaders explained that AR needed a lot of management and organisation especially if the group had pupils with very low reading levels. AR requires a lot of support from trained staff members to manage small group work, advise pupils on book selection and monitor pupil performance closely. It was observed during an AR session in progress that some pupils needed one-to-one teaching instruction before even coming to the level of independent book reading. In such cases, extra support staff are required who could manage banding books to their readability level, make book inventories, and manage quizzes and pupils' diagnostic reports.

Independent book reading time is usually not present in a school timetable. One of the schools had to seek special parental consent to keep pupils for 40 minutes after school hours. This could be tiring and inconvenient for pupils, as was observed in one school visited for such sessions. The schools which adopted AR during school time made pupils miss other lessons, such as modern languages, which resulted in the loss of learning – although the teachers claimed that most of these children would not be able to cope with another language.

Too much homework from other subjects for pupils was also reported as a barrier to independent reading at home. The pupils in the treatment group continued regular school activities and homework was also a part of school routine. Pupils were reporting to the school leaders that it was difficult to cope with reading at home and also completing their regular homework.

It was reported by one of the school leaders that reading complete books was initially exhausting for pupils at low reading levels. Some parents raised concerns over this extra reading in a given period of time but when the teacher explained the purpose of AR they supported their child in completing the reading tasks. The stress could be due to a feeling of group competition and the demands of completing the reading targets in a given time period.

Is the intervention attractive to stakeholders?

AR means fuller use of library resources, in order to achieve the literacy targets. The school leaders reported that due to AR they needed to update their libraries with new books which pupils demanded to read because of the intervention. It was also reported that books in AR were never just lying in the shelves because pupils read them for quizzes and due to pupils' interest in the library a book reading culture was gradually developing. Teachers even had to be updated on their knowledge of the new books as without this they cannot help pupils in making an appropriate selection. It was also reported that AR training was a very useful capacity building programme through which TAs can also be trained for AR implementation. In two schools TAs attended the training, which supported the

teaching staff in running the intervention. School librarians and literacy staff members achieved the most benefit from AR training.

One pupil said that taking a book home for reading was 'very nice' as he had not a single book in his home for reading. Pupils generally reported that they liked to follow the works of famous authors like David Walliams, Roald Dahl, Suzanne Collins, Michael Terry and JK Rowling. Pupils also showed their interest in reading books having illustrations, and most popular here were sagas such as written by Darren Shan. Pupils enjoyed having the opportunity to carry around books they were interested in reading. Some of them also said that they liked reading books because in the end they got a chance to do quizzes on an iPad.

One school leader reported that they received encouraging remarks from parents when their child brought a book home and tried to read it independently. The teacher informed that this activity made some parents aware of the need to keep interesting books at home and to read or at least talk about a book of interest with their child.

Outcomes

Perceived outcomes

School leaders observed pupils gradually improving in their STAR performance. They thought that pupils having exposure to the books of interest and appropriate readability had brought this improvement in pupil's reading levels. It was also perceived by the teachers that attaining AR points and setting targets is in itself a motivation for the students, so an external reward for achievement may not really have any deeper impact. This point of view was not the same among all teachers who used AR because some also believed that extrinsic reward played an important role specifically in encouraging the readers at low levels, and in the first instance.

School leaders suggested that monitoring and advising pupils helped their confidence to take up a reading challenge. Pupil with low reading levels initially believed that finishing a book was almost impossible and they would fail in the quiz. In some cases they were even reluctant to select a book for independent reading. However, teacher's advice and motivation strategies helped these pupils to achieve the reading targets. The teachers reported that the success of these pupils was very rewarding for themselves as well as for the pupils.

In terms of AR implementation during the transition from primary to secondary school one of the school leaders reported that it was not the appropriate timing for some pupils to be introduced to AR. Pupils coming from primary schools needed support from teachers to adjust in the new format of schooling. AR could be a challenge at this beginning stage of secondary school experience.

Some pupils also reported that they had been asked to select a book for reading and if they had the choice they would not have selected and read any book. Their choice was guided by the teacher and they personally had no interest in reading books. They found reading books a very dull and boring activity. Some pupils also said that reading lengthy books put them off reading so they chose smaller books for their reading targets. One of the pupils said that he liked just looking at pictures in the books rather than reading the text so whenever he selected any book he made sure it had more pictures than text. Several pupils were more attracted to books with flashy images rather than books with just plain text. It is possible that pictures and images helped pupils in understanding.

Some teachers believed that regular independent reading inculcates a habit of reading. This is only a perceived outcome. There was, and could be, no solid evidence from this trial that pupils on AR were more likely to become lifelong readers.

Some teachers believed that completing AR quizzes was as much a test of memory as of pupil's comprehension. They thought that pupils who had good memory and concentration levels scored higher than those who could not memorise details or those who had a short concentration span while doing a screen test.

AR is a quite well-known intervention in schools and some parents had already known about AR before their child's participation. Some of those who already knew about AR were very enthusiastic for their children to participate. It was also noted that some parents became aware of having books for children's interest in the home once their children participated in AR.

The AR attendance records showed that pupils were attending the sessions regularly. Although many of them had to come from regular classes and some even had to stay after school, the overall AR attendance record was very good. During observations and interviews pupils reported that they liked coming for AR because there was no teaching and they would get a chance to do a quiz using the tablets. Some of them reported that they preferred coming for AR sessions rather than attending regular classes which they found boring.

NGRT is a computerised screen-test which adapts the level of challenge according to pupil's initial responses. AR quizzes and STAR test are also screen-test and evaluators perceive that pupils were familiar with this format of computer screen-testing. In other tests for interventions screen-test is a challenge for pupils or they consider the screen-test as a game. There is a possibility that pupils on the AR intervention could have more practice of screen-tests, and therefore performed better than the control without necessarily being better at reading. If all were given a paper-pencil test the results could have been different. According to the school leaders the NGRT was a challenging test of reading comprehension for pupils. The format of the test was multiple choice questions. It is possible that some pupils were just guessing the right answers. However, in terms of appropriateness of the reading test content and the level of challenge for pupils in secondary Year 7 NGRT fulfils the criteria for a powerful assessment tool.

Fidelity

The process of AR as demonstrated in the training was generally followed in the schools. On the Renaissance Learning Inc. website there are various features recommended to help teachers make AR effective. However, teachers needed considerable time and practice to make use of all recommended applications. In terms of basic implementation the main process of AR was followed in all of the schools. The following core features of AR were observed in all schools:

- Conducting STAR to assess pupil's ZPD level
- Book selection by teacher and pupil
- Independent reading time by pupil
- AR quiz by pupil
- Test results discussed by teacher and pupil

In the case of pupils with very low reading levels teachers were seen giving them individual attention. During independent reading time teachers were seen tutoring these pupils. This is a variation from the actual process of AR, but the teachers claimed there was no other choice but to support the pupils and then let them be independent in reading. The evaluation included no observation of independent reading at home. Teachers did not log how many books were read independently by each child. A number of EEF evaluations of different literacy interventions have now suggested small positive effect sizes, and it may be that this focused attention is a key component, especially on initial transition to secondary school.

In one of the schools there was a specified day for AR quizzes. This meant that pupils had to complete the book reading before that day. This regularity was not observed in other schools. Pupils generally completed the quiz whenever they finished reading the book. The teachers just monitored the progress of reading by checking the number of days the pupil had kept the book.

STAR tests were repeatedly taken as recommended in AR but not all schools had taken the STAR test at any standard point of time during the period of trial. The groups had taken these tests whenever the teacher felt the need for it. However, schools had maintained the record of tests and used the information for important advice to pupils.

The process of NGRT post-test was simple. The online set-up was pre-arranged and during the last week of the trial all schools conducted the post-test. In one of the schools dates of birth were incorrectly entered for about eight pupils and therefore the NGRT version they received was very low and less challenging. The school leaders detected this error during the test and it all happened in front of evaluator team members who were observing the process. These eight pupils were retested the next day after the error in dates of birth was corrected in the main GL system.

Formative findings

According to one teacher, conducting AR requires a lot of administrative work such as AR colour coding and banding of books, monitoring pupils and advising them on book selection, and following STAR reports for individual pupils and the whole class. It requires space in school planning for the timetable just as given to other subjects.

Suggestions for improvement of AR implementation are based on what the evaluators observed and what school leaders, teaching staff and pupils experienced and reported.

Tutoring pupils with very low reading ability

The AR protocol does not say if teacher's instruction can be involved for pupils at very low reading levels. Independent reading for these pupils is initially frustrating. Teachers said that it is too much to expect them to read complete books and pass the quizzes if pupils do not have basic reading skills such as word recognition, reading fluency and a bit of awareness to decode the text. If these readers with very low reading levels are to use AR then there must be some scope or advice on initial tutoring lessons for them.

AR quiz challenge

School leaders and teaching staff suggested that the AR quizzes need to be made more challenging in order to overcome doubts about whether pupils have read the books independently and not sought the answers from other resources. Collaboration is also possible among a group of pupils who share knowledge of specific books and the quiz. It was suggested that if a book had more than one quiz available then it could also reduce the chances of cheating.

Writing tasks

AR targets mainly reading comprehension but there is a scope to add writing tasks in the intervention. AR quizzes are always multiple choice questions, but teachers felt that if there were small writing activities as well the intervention could have been more effective.

Timings for AR

Schools using AR need to incorporate at least 40 minutes in the timetable so that pupils on the intervention do not have to miss other lessons. AR is difficult to adopt in regular teaching of subjects other than English. However, pupils always had the choice of selecting books that were non-fiction or relevant to other subjects such as Geography, History and other science subjects.

Can school leaders conduct an RCT?

An additional issue for this aggregated trial was whether schools could conduct evaluations of their own interventions. The overall answer is a guarded 'yes' on the basis of these four schools. Having the schools run the intervention was advantageous in several respects. Permission to innovate was easier, no schools and few pupils dropped out after being deemed eligible, and there was no developer pushing the advantages of their product. These schools showed themselves capable of randomising the eligible pupils to treatment groups, apparently without subversion. This may be partly due to the fact that the trial used a waiting-list design. School leaders stated that the initial day of training by the evaluators was very helpful in sorting out priorities and the logic of the order of events here. They said that conducting the research project was an educational experience for them. These meetings helped the evaluators to know school leaders' views of the research project and to find what challenges teachers faced in conducting small-scale research projects in their schools. The trial may

also have been easier for schools because there was no explicit pre-test, and so identification of eligibility could lead directly to randomisation.

In terms of managing the intervention, school leaders were free to make decisions regarding venues for the intervention, purchasing materials, choice of equipment, timings and class adjustment without any developer's direct involvement. This lack of developer intervention was a considerable advantage, freeing schools and the evaluators from the pressure to demonstrate success experienced in other EEF-funded trials so far. This independence was partly possible because AR is fully developed and also supported through accessible online content.

However, there are some limitations and barriers. It was observed that most staff involved became advocates for AR increasingly during the trial, and schools had already made arrangements to continue with and expand the use of AR for future years. They did not have the mental equipoise needed to conduct a fair test.

School leaders did not always appreciate the importance of aspects of the evaluation. For example, when pressed they were happy to support the evaluators who were trying to locate and test missing pupils. But they did not do this themselves, and had no real concept of the dangers of attrition (despite discussion in the training days).

Although the school leaders were given guidance, materials and a template to develop school reports, in addition to the two training days on evaluation, none of the schools submitted an individual school report on the results. The design of an RCT makes analysis simple – the headline result is just the standardised average difference between the two groups at post-test. But this is seemingly beyond the capacity of school leaders.

The process of evaluation was always guided and monitored by the expert evaluators. The communication between school leaders and the evaluation team was excellent and exchange of information was quite frequent. The protocol of a rigorous RCT would not have been followed without the evaluators' advice and constant communication with school leaders. Therefore, on the basis of this small study, it seems that schools can conduct evaluations under certain conditions but that they would always need expert assistance at a few crucial points.

Control group activity

Pupils in the control group were not given access to AR features nor did teachers follow any protocol of AR activities with them. In general, pupils in the control group continued their usual activities in the school which sometimes used other reading interventions. Pupils in the control group were not restricted in access to books in the library or to other equipment such as tablets and iPads.

Conclusions and implications

The headline finding, coupled with the low attrition and initial balance between groups, is that Accelerated Reader is a successful approach in this context. The trial is medium in scale, but as shown in the background discussion it is not only the largest in the UK, but also the largest actual RCT ever conducted with AR.

Limitations

In one of the schools randomisation was by classes rather than eligible individuals. There is no reason to suspect that this has influenced the results, and the impact was similar across all schools.

The randomisation was done within the schools and pupils in both arms of the trial continued joint activities in the school other than AR sessions. Students in the treatment group were taken out of the regular classes for AR or they were doing AR after the school time. There is a minimal danger of treatment diffusion to the pupils in the control group.

The design adopted for this trial cannot consider any long-term impact of this intervention. Once the tests were completed, the control group received the intervention. There was no opportunity to follow AR pupils and see if they continued their attitude, interest and performance in independent book reading.

The schools participating in the trial had volunteered to conduct the AR intervention themselves. The schools are therefore not necessarily a representative sample of a larger population in the area.

AR developers were not involved in this evaluation which means that there was no conflict of interest. However, there is a possibility that school leaders' enthusiasm to take part in a funded research project resulted in a kind of 'Hawthorne' effect. The schools volunteered to conduct AR for the purpose of evaluation but all of them had already decided to do so because they felt that AR was a very good intervention.

The results of NGRT are based on pupils' performance on the screen-test. NGRT is a computerised screen-test which adapts the challenge according to pupil's initial responses. AR quizzes and STAR test are also screen-tests. It is possible that some of the difference between the groups can be explained by extra familiarity with this kind of testing. If all were given a paper-pencil test the results could have been slightly different.

Interpretation

The main objective of this RCT was to test the gains in reading comprehension after using AR against those of pupils in the control group. The criteria of a rigorous trial were maintained throughout, no school dropped out of the evaluation, AR was regularly administered, pupils in the control group did not have access to AR, and the pupil attrition rate was around 2.3%. The trial was modest in scale but is the largest robust evaluation of AR in the UK by some considerable margin. Therefore, the overall finding suggests that AR was an effective reading intervention with the effect size of around +0.24. This is approximately equivalent to an additional three months' progress in the reading age of pupils after 22 weeks. It was at least equally effective for FSM-eligible pupils and so could be part of any attempt to reduce the poverty gradient in literacy on transfer to secondary school. The estimated cost of using AR for one year is £9 per pupil assuming that staffing, access to internet and most importantly a wide range of books are available.

AR may be harder to implement with pupils with very low initial reading levels. These pupils may need some preparation of a more formal nature. Other than that, AR can be used with all abilities and with individuals or whole classes.

The schools have demonstrated that they can learn about RCTs and conduct many of the elements themselves without direct supervision. However, there are some aspects that would still require oversight by independent evaluators. This would be even more important in other designs (such as school-level allocation or non-waiting list) or where there is likely to be a conflict of interest. Schools naturally tend to be more concerned with the progress of each specific cohort than with the generation of safe knowledge for others.

Future research and publications

AR is a well-developed intervention with strong technological support available for teachers, and this RCT has shown that it can work for pupils who are struggling to achieve age-appropriate reading levels. The main concern emerging from the process evaluation is the need for teacher instruction and a certain level of guided reading practice for pupils at very low reading levels. An effectiveness trial on a larger scale would be a useful next step.

Would the teacher's role have an impact? The teacher's role is supposedly just to provide advice on book selection and monitoring pupil's performance through quizzes. The teacher is just managing a group of readers who are doing AR. Would it make a difference if the teacher's role were also replaced by software, where pupils receive advice and motivation to read through computers?

What is the 'active' ingredient of AR? There are six key steps involved in the AR process which involved STAR assessment, teacher's advice and monitoring to pupils, book selection, independent book reading by pupil and pupil taking the quiz. What is the most effective element of this whole process? Is it regular independent reading time? Does it depend on pupils using online quizzes? Or is it the atmosphere of whole-group book reading competition that motivates pupils to read? Would any regime of reading books and completing quizzes give similar results or is it just using the Renaissance Learning Inc. recommended books and quizzes that are responsible for this impact?

Do pupils become lifelong readers? One of the claims of AR is that pupils inculcate the habit of regular reading after using AR. There is a question of how long the impact of AR persists. A longer-term study is possible.

The evaluators plan to use these results in a future paper, comparing the strengths and weaknesses of AR with other possible literacy catch-up programmes.

References

- Berk, R. and Freedman, D. (2001) Statistical assumptions as empirical commitments, <http://www.stat.berkeley.edu/~census/berk2.pdf>, accessed 030714
- Brooks, G. (2007) *What works for pupils with literacy difficulties? The effectiveness of intervention schemes*, London: DCSF Publications
- Bullock, J. (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; 0171 Advisor: Gerald Tindal. *DAI*, 66 (07A), 56-2529
- Carver, R. (1978) The case against statistical significance testing, *Harvard Educational Review*, 48, 378-399
- Falk, R. and Greenbaum, C. (1995) Significance tests die hard: the amazing persistence of a probabilistic misconception, *Theory and Psychology*, 5, 75-98
- Goodman, G. (1999) *The Reading Renaissance/Accelerated Reader Program, Pinal County school-to-work evaluation report*, Tucson, AZ: Creative Research, Inc. (ERIC Document Reproduction Service No. ED427299)
- Gorard, S. (2013) *Research Design: Robust approaches for the social sciences*, London: SAGE
- Gorard, S. (2013) The propagation of errors in experimental data analysis: a comparison of pre- and post-test designs, *International Journal of Research and Method in Education*, <http://dx.doi.org/10.1080/1743727X.2012.741117>
- Institute of Education Sciences (IES) (2008) *What Works Clearing House Intervention Report: Accelerated Reader*, US Department of Education
- Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K. and Busick, M. (2012) *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*, Washington DC: Institute of Education Sciences
- Nichols, J. (2013) *Accelerated Reader and its effect on fifth-grade students' reading comprehension* (Doctoral dissertation, Liberty University)
- Nunnery J., Ross, S. 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), 1–18
- Paul, T., VanderZee, D., Rue, T. and Swanson, S. (1996) *Impact of the Accelerated Reader technology-based literacy program on overall academic achievement and school attendance*. Paper presented at the National Reading Research Centre Conference on Literacy and Technology for the 21st Century (Atlanta, GA, October 4, 1996)
- Paul, T., VanderZee, D., Rue, T., and Swanson, S. (1996) *Impact of the Accelerated Reader Technology-Based Literacy Program on Overall Academic Achievement and School Attendance*, Madison, WI: The Institute for Academic Excellence
- Pavonetti, L. M., Brimmer, K. M., & Cipielewski, J. F. (2000) *Accelerated Reader [R]: What Are the Lasting Effects on the Reading Habits of Middle School Students Exposed to Accelerated Reader [R]*

in Elementary Grades? Paper presented at the Annual Meeting of the National Reading Conference (50th, Scottsdale, AZ, November 29-December 2, 2000)

Peak, J., & Dewalt, M. W. (1994) Reading Achievement: Effects of Computerized Reading Management and Enrichment, *ERS Spectrum*, 12(1), 31-34

Ross, S., Nunnery, J. and Goldfeder, E. (2004) A randomized experiment on the effects of Accelerated Reader/Reading Renaissance in an urban school district: Preliminary evaluation report. Memphis, TN: The University of Memphis, Center for Research in Educational Policy

Scott, L.S. (1999) The Accelerated Reader program, reading achievement, and attitudes of students with learning disabilities. Unpublished doctoral dissertation, Georgia State University, Atlanta (ERIC Document Reproduction Service No. ED 434431)

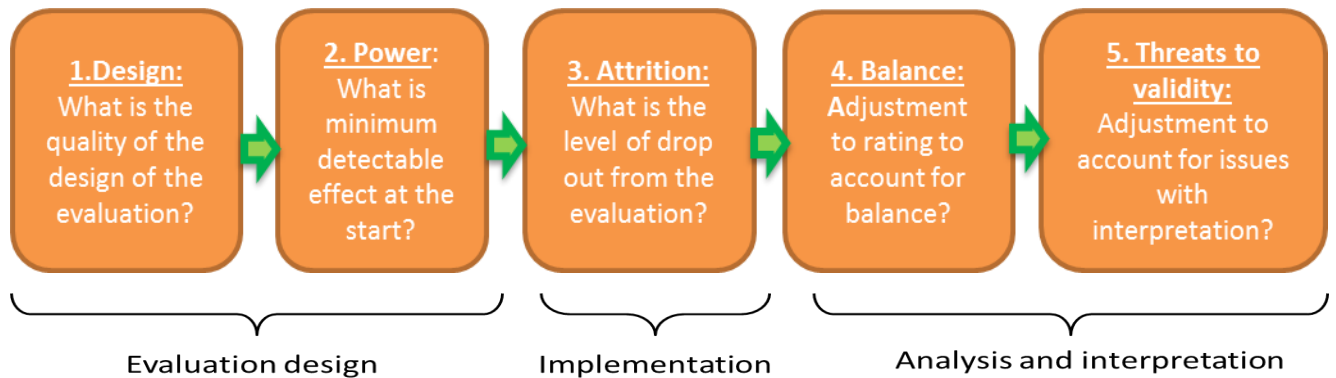
Topping, K. (2014) What kids are reading: The book reading habits of students in British Schools 2014: An Independent Study, Renaissance Learning Inc.: United Kingdom

Vollands, S., Topping, K. and Evans, R. (1996) Experimental evaluation of computer assisted self-assessment of reading comprehension: Effects on reading achievement and attitude, ERIC Document, ED 408 567

Vollands, S., Topping, K. and Evans, R. (1999) Computerized self-assessment of reading comprehension with the Accelerated Reader: action research, *Reading and Writing Quarterly*, 15, 3, 197-211

Watts, D. (1991) Why is introductory statistics difficult to learn?, *The American Statistician*, 45, 4, 290-291

Appendix 1: Padlock rating



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

The final security rating for this trial is 3 . This means that the conclusions have moderate security.

The trial was designed as an efficacy trial, partly administered by the schools. The study design could have achieved a maximum of 5 . However, the study was relatively small and lost 2 padlocks. The attrition was extremely low, and there was no indication of imbalance between the arms at baseline. There were a few threats to validity: the schools administered the randomisation without record; the teachers administered the testing, and a second analysis of the effect size did not corroborate the headline result. Therefore, the overall padlock rating is 3 .

Appendix 2: Cost rating

Cost ratings are based on the approximate cost per pupil of implementing the intervention over one year. Cost ratings are awarded using the following criteria.

Cost	Description
£	<i>Very low:</i> less than £80 per pupil per year.
£ £	<i>Low:</i> up to about £170 per pupil per year.
£ £ £	<i>Moderate:</i> up to about £700 per pupil per year.
£ £ £ £	<i>High:</i> up to £1,200 per pupil per year.
£ £ £ £ £	<i>Very high:</i> over £1,200 per pupil per year.

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

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

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

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



Education
Endowment
Foundation

The Education Endowment Foundation

9th Floor, Millbank Tower

21–24 Millbank

London

SW1P 4QP

www.educationendowmentfoundation.org.uk