Accelerated Reader

Addendum 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 1 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.

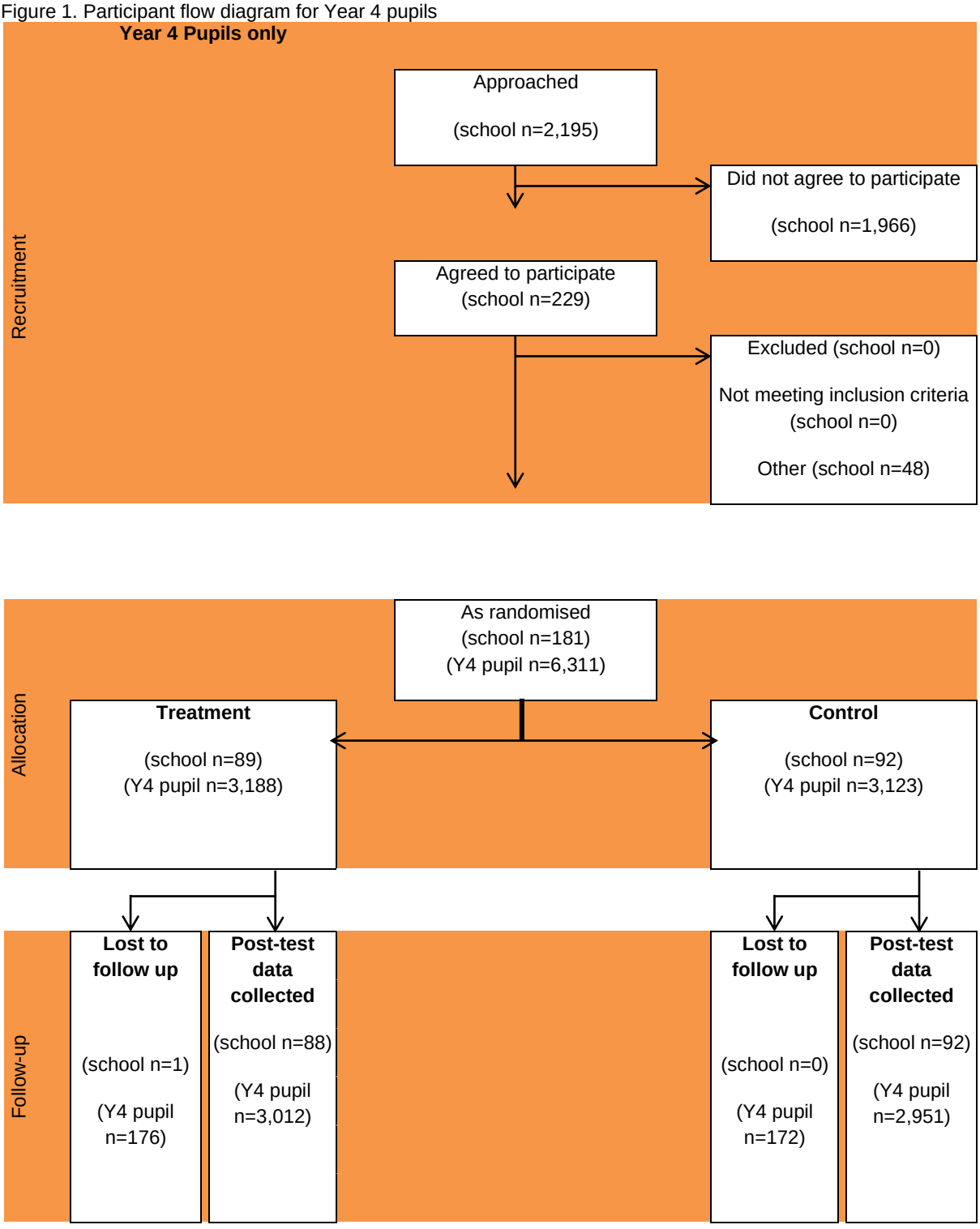
Impact evaluation: Year 4 cohort

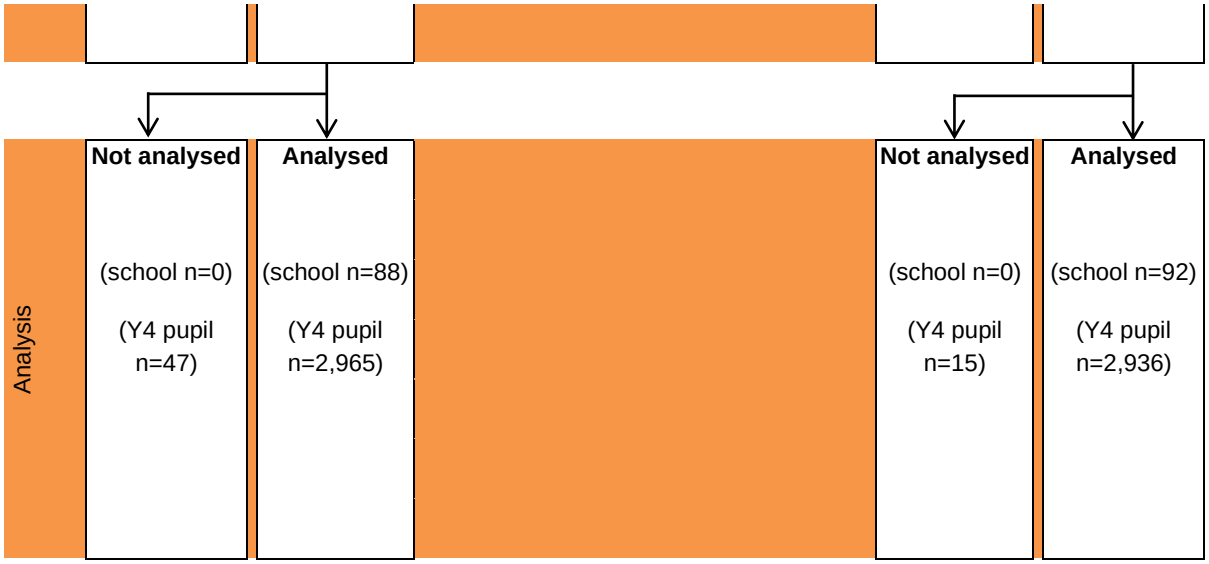
Participants

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 2. 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 3. 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 4. 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).

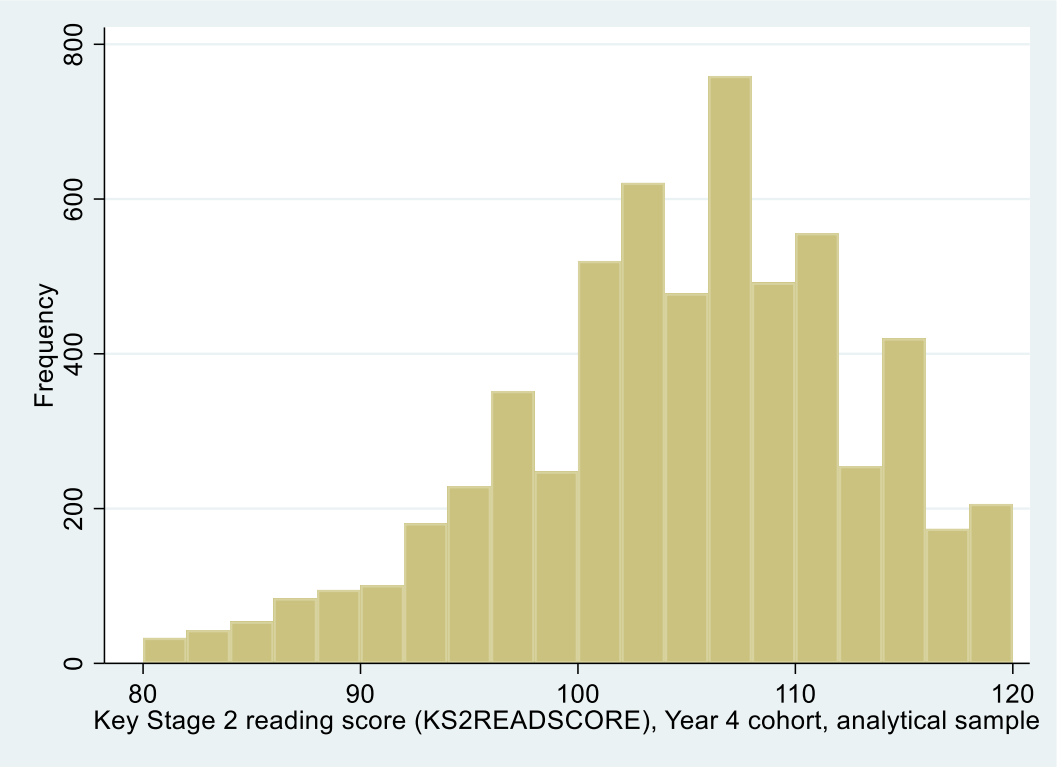
Outcomes and analysis

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.

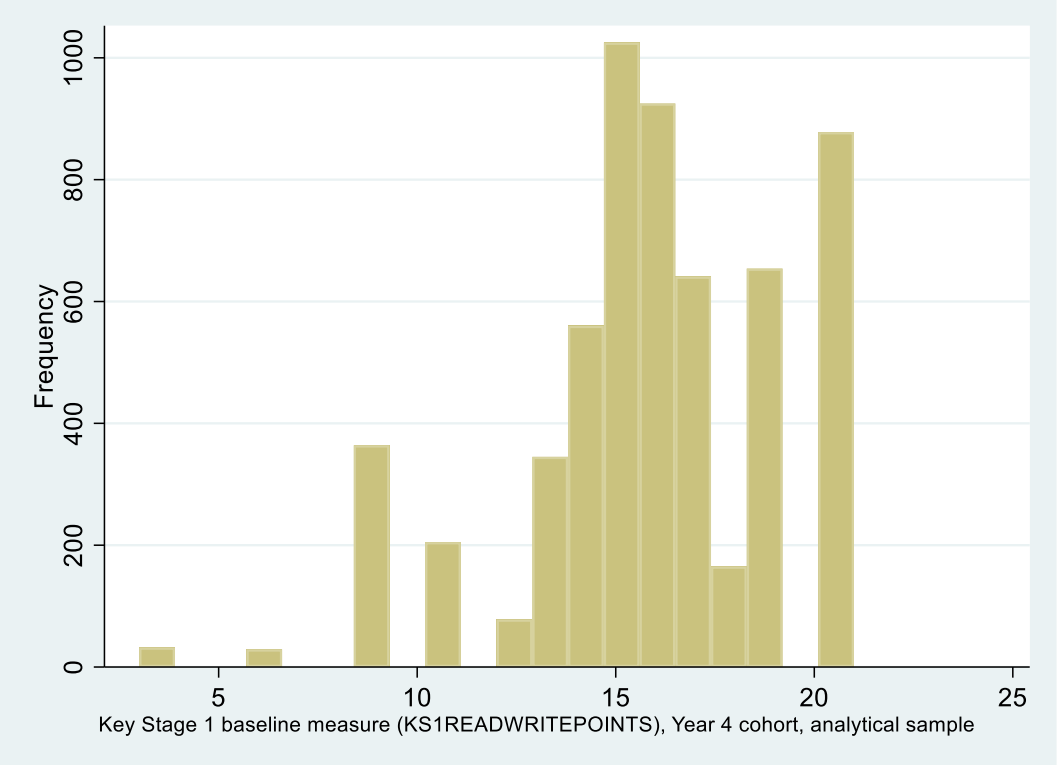
Figure 2. Distribution of Key Stage 2 reading score outcome measure, analytical sample. (smallest bin based on 36 observations)



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 5. 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 6. 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. **Error! Reference source not found.** and **Error! Reference source not found.** present the results of the analysis for the FSM group. Table 9. Subgroup analysis (ITT), Non-FSM groupTable 9 and **Error! Reference source not found.** 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 7. Subgroup analysis (ITT), FSM group

	Raw means				Effect sizes		
Outcome	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 8. 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 9. Subgroup analysis (ITT), Non-FSM group

	Raw means				Effect sizes		
Outcome	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 10. 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 11. 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