

# Outcome evaluation: *Reading Planet: Rocket Phonics* trial 2021–2023

Independent evaluation by Dr Helen L. Breadmore, Associate Professor Psychology in Education, School of Education, University of Birmingham

14<sup>th</sup> November 2023

## Abstract / Executive summary

---

This paper reports the first outcome evaluation to assess the impact of exposure to two complete years of the new *Reading Planet: Rocket Phonics* programme. Impact was assessed by examining the reading attainment, spelling and phonics knowledge of Year 1 pupils in 11 schools which had self-selected to use the new *Reading Planet: Rocket Phonics* programme, and comparing this to 13 business-as-usual control schools. There were no significant differences between conditions in any outcome measures. This means that the pupils in schools using the *Reading Planet: Rocket Phonics* programme progressed with their early literacy skills no worse and no better than controls. Descriptive statistics showed some further evidence of promise. Means suggested that pupils in schools using *Reading Planet: Rocket Phonics* had literacy attainment slightly below controls in Autumn 2022 which had shifted to slightly above controls in Spring 2023 (quantified by PiRA standardised scores). However, those differences were small and were not statistically significant. In spelling the differences appeared more promising – means suggested that in Summer 2023 pupils in schools using *Reading Planet: Rocket Phonics* had slightly better spelling scores than controls (equivalent to spelling one additional word correctly). However, the difference was not statistically significant and may have been due to chance. In-programme phonics assessments indicated that pupils in schools that implemented *Reading Planet: Rocket Phonics* clearly increased phonics knowledge over the course of Year 1 (a finding also observed in Year R previously – see Breadmore, 2023), although they did not differ from controls in their performance on the National Phonics Screening Check. Together, this evidence of promise is consistent with the expected mechanisms of change – the *Reading Planet: Rocket Phonics* programme is expected to increase phonics knowledge, leading to growth in word reading and spelling, which in turn support literacy acquisition. There are a number of issues with trial design that may have limited the ability to fully observe the impact of the *Reading Planet: Rocket Phonics* programme as compared to control. In particular, the non-randomised assignment of schools to conditions, risk of bias in sampling, lack of baseline assessment of phonics knowledge, and changes in control sample used in Year R compared to Year 1. These issues could be overcome in future trials, which should also include a priori indicators of compliance and fidelity to understand the impact that the quality of implementation has on outcomes.

# Contents

---

<b>Outcome evaluation: <i>Reading Planet: Rocket Phonics</i> trial 2021–2023</b> .....	1
Abstract / Executive summary.....	2
Contents .....	3
List of tables .....	4
List of figures .....	5
Introduction .....	6
The present study .....	6
Method.....	9
Participants .....	9
School level information about sample recruitment and retention.....	9
Participants assigned to <i>Reading Planet: Rocket Phonics</i> and control conditions .....	10
Design and procedure.....	11
Condition.....	12
Trial responsibilities .....	12
Primary and secondary outcome measures.....	12
Data diagnostics.....	15
Results .....	16
Data analysis plan .....	16
<b>Primary Outcomes (Literacy Attainment)</b> .....	16
Impact of condition on primary outcome measures .....	22
Secondary Outcomes: Spelling and Phonics Knowledge .....	22
Impact of condition of secondary outcomes .....	25
In-programme phonics assessment.....	28
Discussion.....	29
Sampling and sample size .....	29
Missing data .....	30
Timing of outcome measures .....	31
Validity of outcome measures .....	31
Programme differentiation, compliance and fidelity.....	32
Conclusion .....	33
References.....	34
Appendix 1 .....	36

## List of tables

---

Table 1: Summary of background information about schools assigned to <i>Reading Planet: Rocket Phonics</i> and control conditions. ....	10
Table 2: Summary background information about pupils included in primary outcome analyses in schools assigned to <i>Reading Planet: Rocket Phonics</i> and control conditions. ....	11
Table 3: Summary background information about pupils included in secondary outcome analyses (GAPS spelling subtest and National Phonics Screening Check) in schools assigned to <i>Reading Planet: Rocket Phonics</i> and control conditions.....	11
Table 4: Summary of pupil level scores on primary outcome (Progress in Reading Attainment – <i>New PiRA</i> ) at each time point in schools assigned to <i>Reading Planet: Rocket Phonics</i> and control conditions. ....	17
Table 5: Summary of primary outcomes, averaged within school clusters at each time point for pupils in schools assigned to <i>Reading Planet: Rocket Phonics</i> (RP) and Control (C) conditions.....	20
Table 6: Summary of pupil level scores on secondary outcomes (GAPs spelling subtest, National Phonics Screening Check, in-programme phonics assessments) at each time point in schools assigned to <i>Reading Planet: Rocket Phonics</i> and control conditions. ....	23
Table 7: Summary of secondary outcomes (raw scores on GAPs spelling subtest, National Phonics Screening Check raw score), averaged within school level clusters for pupils in schools assigned to <i>Reading Planet: Rocket Phonics</i> (RP) and control (C) conditions.....	24
Table 8: Summary of pupil level scores on primary outcome (Progress in Reading Attainment – <i>New PiRA</i> ) at each time point in schools assigned to <i>Reading Planet: Rocket Phonics</i> and control conditions limited to only those participants also included in secondary outcomes analysis. ....	36

## List of figures

---

Figure 1: Primary and secondary outcome measures across time.....	13
Figure 2: Violin plot illustrating the distribution of pupil level New PiRA standardised scores by time (1 = Autumn 2022, 2 = Spring 2023, 3 = Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics). Mean scores plotted with a point.....	18
Figure 3: Violin plot illustration of distribution of school mean new PiRA standardised scores illustrated by time (Autumn 2022, Spring 2023, Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics). Mean scores illustrated with a point.....	21
Figure 5: Violin plot illustrating distribution of pupil level <i>GAPS</i> spelling subtest raw scores illustrated by time (1 = Autumn 2022, 2 = Spring 2023, 3 = Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics). Mean scores illustrated with a point.....	26
Figure 6: Violin plot illustrating distribution of school mean <i>GAPS</i> spelling subtest raw scores illustrated by time (1=Autumn 2022, 2=Spring 2023, 3=Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics). Mean scores illustrated with a point.....	26
Figure 7: Pupil level raw scores on the National Phonics Screening Check illustrated by time (Autumn 2022, Spring 2023, Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics).....	27
Figure 8: School mean raw scores on the National Phonics Screening Check illustrated by time (Autumn 2022, Spring 2023, Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics).....	27
Figure 9: Pupil level scores on in-programme phonics assessment at baseline and endline. ....	28
Figure 10: School mean scores on in-programme phonics assessment at baseline and endline. ....	28

# Introduction

---

Literacy is a core academic skill which impacts on every aspect of education and future success (Secretary of State for Education, 2022). In England, the Department for Education provide a list of validated Systematic synthetic phonics (SSP) programmes and encourages schools one of these programmes to teach phoneme-grapheme correspondences explicitly. The aim is to help children to read and spell words by decoding or ‘sounding out’ the letters. Early literacy instruction is most effective when it combines phonological awareness and phonics training, along with regular reading practice (Bus & van IJzendoorn, 1999; Ehri et al., 2001; Hulme et al., 2012). Here, we seek to evaluate the effectiveness of a new programme that aims to do this – the *Reading Planet: Rocket Phonics* programme from Hodder Education. For a thorough review of the context of the trial, see the Interim Report (Breadmore, 2023).

The *Reading Planet: Rocket Phonics* programme offers a complete solution to teaching phonics from the beginning of Reception (Year R) to the end of Key Stage 1 (followed by ongoing resources through the Reading Planet book scheme throughout Key Stage 2). The programme has been developed by a phonics expert (Abigail Steel), aligns well to curriculum guidance and has been validated by the Department for Education (DfE). Previous trials conducted during the development of the programme offered promising evidence of the suitability of the programme for early literacy instruction. However, the evaluations focused on gathering feedback to improve the quality and content of reading and teaching materials to increase satisfaction and acceptability (i.e., to improve adherence and fidelity), rather than directly measuring pupil outcomes. Case studies and testimonials from individual schools suggest that teachers observed improvements in attainment (see <https://www.risingstars-uk.com/case-studies/english-reading-impact-studies/reading-planet>). However, these reports are not objective, and without a control group comparison it is unclear what proportion of observed changes in attainment can be attributed to the *Reading Planet: Rocket Phonics* programme. Hence, this is the first trial to directly measure pupil outcomes to examine the impact of the complete *Reading Planet: Rocket Phonics* programme delivered for two academic years from the beginning of Year R to the end of Year 1.

## The present study

The aim of the present study is to evaluate the impact of the *Reading Planet: Rocket Phonics* programme. The programme was developed for implementation as the main method of literacy instruction from the beginning of Year R and throughout Key Stage 1. Interim reports published in 2023 examined progress over the first academic year of delivering the programme (Breadmore, 2023; Breadmore & Wood, 2023). The results from the Year R cohort showed promising signs of progress in both phonics knowledge and reading attainment. Although scores did not differ significantly from control schools, at the end of that year the pupils in *Reading Planet: Rocket Phonics* schools were on average around 5 standard scores ahead of those in control schools (the equivalent of 10 months difference in reading age). Scores for pupils in Year 1 were extremely varied and difficult to draw conclusions from. These pupils had not been exposed to *Reading Planet: Rocket Phonics* from the beginning of their literacy education, their prior learning had been disrupted by the impact of the COVID-19 pandemic, and it appeared that implementation of the programme had been variable with this cohort. In the present evaluation we therefore only follow the pupils who were in Year R during the first year of the trial, measuring the impact of the programme through their second year in formal education (Year 1), after having received *Reading Planet: Rocket Phonics* from the beginning of formal literacy instruction.

As is typical for a trial of this type, we compare the progress of children exposed to the ‘intervention’ (*Reading Planet: Rocket Phonics*) to those in a ‘business-as-usual control’ condition (i.e., receiving other methods of phonics instruction). The *Reading Planet: Rocket Phonics* programme is a complex intervention - it contains multiple elements which are likely to impact on outcomes and it may be difficult to identify which of those elements are the active ingredients. For example, if reading attainment increases with exposure to *Reading Planet: Rocket Phonics*, this could be due to the programme raising children’s phonics skill, or because it provides opportunities to apply those skills to decodable reading books, or because continuing professional development (CPD) provision increases teacher knowledge and self-efficacy in teaching literacy, or some interaction between these factors. In this outcome evaluation, we focus on the pupil outcomes; – literacy attainment (primary outcome), spelling and phonics knowledge (secondary outcomes).

The primary outcome, **literacy attainment**, aligns to the overall goal of the programme and is operationalised as attainment on the Progress in Reading Assessment (*New PiRA* – Ruttle et al., 2021). The ability to manipulate phonemes and graphemes, and to learn and use phoneme–grapheme correspondences is essential for early reading and spelling. However, this phonics knowledge is only a steppingstone – a starting point on the difficult journey to becoming literate. SSP programmes do not aim to teach every phoneme–grapheme correspondence rule, and indeed not all English words are most efficiently decoded using such correspondences. SSP programmes aim to get children started so that they can go on to learn more efficient processes (Breadmore et al., 2019; Rose, 2006; Torgerson et al., 2018). That is why combining phonics with opportunities to apply this knowledge in reading and spelling is recommended. It gives children the chance to apply their knowledge and learn further skills through ‘self-teaching’. Progress in Reading Assessment (*New PiRA* – Ruttle et al., 2021) has good reliability and face validity as a measure of literacy attainment in this context. It can be group administered in the second half of each academic term and has been standardised and norm-referenced accordingly.

Secondary outcomes help us to examine more directly whether the programme is successful in training the underlying skills that it targets. In the present study we explore spelling and phonics knowledge as secondary outcomes. **Spelling** was measured using the spelling subtest of Progress in Grammar, Punctuation and Spelling Assessment (Charles et al., 2018), which can be group administered in the second half of each academic term and has been standardised accordingly. The subtest has not been norm-referenced. We assessed **phonics knowledge** using data from the National Phonics Screening Check. The National Phonics Screening Check is a mandatory assessment which is administered to most pupils in England towards the end of Year 1 each year<sup>1</sup>. Content is determined by the Standards and Testing Agency and is embargoed until after administration. As such, we did not know the precise content when designing the trial or during programme delivery. Hence, this measure is more distant from the programme and more objective since it is not under control of programme developers. However, this secondary outcome measure is only available at endline.

To summarise, the aim of the present outcome evaluation is to assess the impact of the *Reading Planet: Rocket Phonics* programme on Year 1 literacy attainment, spelling and phonics knowledge. This is achieved in a controlled trial, comparing the performance of pupils attending schools delivering the *Reading Planet: Rocket Phonics* programme to those in a business-as-usual control. Our hypotheses are;

1. Pupils who complete two years of the *Reading Planet: Rocket Phonics* programme will demonstrate increased literacy attainment compared to those in the control condition.

---

<sup>1</sup> There were exceptions due to the extraordinary circumstances of the COVID-19 pandemic in 2020 and 2021, but this trial was not affected by this.

2. Pupils who complete two years of the *Reading Planet: Rocket Phonics* programme will have better spelling attainment than those in the control condition.
3. Pupils who complete two years of the *Reading Planet: Rocket Phonics* programme will have more secure phonics knowledge than those in the control condition.



## Method

---

### Participants

Participants were recruited via their schools, who decided whether to participate, self-selecting their assignment to each arm of the trial. We begin by describing school recruitment and school level sampling information, and then describe participant demographics.

### School level information about sample recruitment and retention

This outcome evaluation is informed by information from 24 schools – 11 in the *Reading Planet: Rocket Phonics* condition and 13 controls.

**Reading Planet: Rocket Phonics Condition:** Fourteen schools were initially recruited by Hodder Education to trial the *Reading Planet: Rocket Phonics* programme in the 2021/22 academic year, self-selecting to this condition. One school withdrew prior to any data collection. The interim reports (Breadmore, 2023; Breadmore & Wood, 2023) describe outcomes for Year R and Year 1 pupils during that first year of the trial. Here we report outcomes for 11 of these schools who continued in the second year of the trial, implementing *Reading Planet: Rocket Phonics* with Year 1 pupils in the 2022/23 academic year. These schools included 3 academy converters, 3 academy sponsor led and 5 community schools.

**Control Condition:** Fifteen schools were recruited by Hodder Education to act as controls in the second year of the trial. These schools were recruited through a call for expression of interest which was sent to primary schools which had previously purchased from Hodder Education. However, two schools did not provide outcome data, hence the outcome evaluation is informed by information from 13 control schools. This included 2 academy converters, 1 academy sponsor led, 3 community, 1 foundation, 2 voluntary aided and 4 voluntary controlled schools. Only one of the control schools included in this outcome report took part in the first year of the trial.

Summary information about school characteristics is illustrated in Table 1 and indicates that schools assigned to *Reading Planet: Rocket Phonics* and control conditions differed in several important ways. This information was sourced from publicly available information from the DfE school performance comparison website (<https://www.compare-school-performance.service.gov.uk/>). The table includes national averages for state-funded primary schools for comparison. Overall, *Reading Planet: Rocket Phonics* schools tended to be a little larger than control schools (which were more variable in size). In terms of the demographics of their pupils, *Reading Planet: Rocket Phonics* schools tended to have a high proportion of children who have been eligible for Free School Meals compared to the national average, and this is also substantially larger than control schools. The proportion of children with English as an Additional Language was around national average in *Reading Planet: Rocket Phonics* schools, but control schools had far fewer (well below a nationally representative proportion of children for whom English was not their first language). The proportion of children registered for SEND support was in line with national averages for both control and *Reading Planet: Rocket Phonics* schools. Putting this background information about schools together, there is no evidence to suggest systematic biases in the selection of schools that would present an advantage for *Reading Planet: Rocket Phonics*. Indeed, any biases present would be expected to advantage pupils in control schools.

**Table 1: Summary of background information about schools assigned to *Reading Planet: Rocket Phonics* and control conditions.**

	<i>National Average</i> <sup>2</sup>	<i>Reading Planet: Rocket Phonics</i>	<i>Control</i>
<i>N schools</i>		11	13
<i>Last OFSTED rating</i>		7 Good 4 Requires improvement	3 Outstanding 9 Good 1 Requires improvement
<i>N pupils on school roll</i> <sup>1</sup>	4,727,089	$\bar{x}$ 366.9 (SD 116.8) 204–623	$\bar{x}$ 256.8 (SD 204.5) 45–831
<i>% with SEN support</i> <sup>1</sup>	12.6	$\bar{x}$ 12.6 (SD 4.1) 3.4–18.8	$\bar{x}$ 13.9 (SD 4.9) 4.8–23.7
<i>% English not first language</i> <sup>1</sup>	21.2	$\bar{x}$ 23.2 (SD 21.2) 2.4–81.4	$\bar{x}$ 9.6 (SD 9.6) 0–27.4
<i>% Eligible for Free School Meals at any time during the past 6 years</i> <sup>1</sup>	23	$\bar{x}$ 36.4 (SD 19.5) 11.0–60.6	$\bar{x}$ 18.0 (SD 14.9) 0–53.4

**Table note:** <sup>1</sup>Mean ( $\bar{x}$ ), standard deviation (SD), range. <sup>2</sup>National averages in state-funded primary schools sourced from school census 2019.

## Participants assigned to *Reading Planet: Rocket Phonics* and control conditions

A complete dataset is needed to conduct outcome analyses, and therefore a priori participant level exclusionary criteria were developed to ensure a complete dataset was available for analysis. This method excludes all mid-phase leavers and joiners, pupils with severe SEND who were unable to access some assessments, and unexplained missing data. Note also that parents maintained the right to withdraw their child's data from the evaluation: any such withdrawals are not included in any pupil counts. On inspection of the data it was apparent that removing participants with any missing data would result in a very large proportion of data excluded as missing. To mitigate this, exclusions were applied separately for primary and secondary outcomes analyses. Analyses of primary outcomes excluded all pupils with missing data points for literacy attainment (*New PiRA*). Analyses of secondary outcomes excluded all pupils with any missing data points for literacy (*New PiRA*), spelling (*GAPS spelling subtest*) or phonics knowledge (*National Phonics Screening Check*).

A total of 1236 participants were recruited into the trial, but a large proportion were excluded due to missing data (429 exclusions in analyses of primary outcome measures, 635 exclusions in analyses of secondary outcomes). The number of pupils per school were quite variable and even more so in control schools. After exclusions had been applied,

- 807 pupils remained and were included in the primary outcome analyses (398 from control schools and 409 from *Reading Planet: Rocket Phonics* schools). There are a similar number of participants in each condition, despite differences in the number of schools. Demographic information about these participants is described in Table 2.
- 601 pupils remained and were included in the secondary outcome analyses (283 from control schools and 318 from *Reading Planet: Rocket Phonics* schools). Despite there being fewer schools, there was more complete data from *Reading Planet: Rocket Phonics* participants than controls. Demographic information about these participants is described in Table 3.

**Table 2: Summary background information about pupils included in primary outcome analyses in schools assigned to *Reading Planet: Rocket Phonics* and control conditions.**

	<i>Reading Planet: Rocket Phonics</i>	<i>Control</i>
<i>N schools</i>	11	13
<i>Total N participating pupils</i>	409	398
<i>Gender (Male: Female: Other including missing)</i>	215:194:0	187:21 <sup>1</sup> :0
<i>Age 1st September 2021 (in total months)<sup>1</sup></i>	$\bar{x}$ 65.59 (SD 3.64), 57–74	$\bar{x}$ 65.34 (SD 3.58), 60–72
<i>Cluster size (N pupils per school)<sup>1</sup></i>	$\bar{x}$ 37.18 (SD 9.10), 14–49	$\bar{x}$ 30.62 (SD 25.62), 5–104
<i>N EAL (EAL:Non-EAL:Missing)</i>	123:283:3	57:341:0
<i>N Pupil premium (PP:Non-PP:Missing)</i>	117:292:0	80:318:0

**Table note:** <sup>1</sup>Mean ( $\bar{x}$ ), standard deviation (SD), range overall.

**Table 3: Summary background information about pupils included in secondary outcome analyses (GAPS spelling subtest and National Phonics Screening Check) in schools assigned to *Reading Planet: Rocket Phonics* and control conditions.**

	<i>Reading Planet: Rocket Phonics</i>	<i>Control</i>
<i>N schools</i>	11	13
<i>Total N participating pupils</i>	318	283
<i>Gender (Male: Female: Other including missing)</i>	158:160:0	126:15 <sup>7</sup> :0
<i>Age 1st September 2021 (in total months)<sup>1</sup></i>	$\bar{x}$ 65.77 (SD 3.65), 60–72	$\bar{x}$ 65.49 (SD 3.65), 60–72
<i>Cluster size (N pupils per school)<sup>1</sup></i>	$\bar{x}$ 28.91 (SD 7.02), 14–39	$\bar{x}$ 21.77 (SD 11.93), 5–104
<i>N EAL (EAL:Non-EAL:Missing)</i>	91:224:3	35:248:0
<i>N Pupil premium (PP:Non-PP:Missing)</i>	82:236:0	36:247:0

**Table note:** <sup>1</sup>Mean ( $\bar{x}$ ), standard deviation (SD), range overall.

## Design and procedure

*Reading Planet: Rocket Phonics* Schools were recruited in Summer 2021 and formally entered the trial in September 2021 (at the beginning of the Autumn term when the children were in Year R) continuing through to July 2023 (end of Summer term in Year 1). Allocation to condition was self-selected (i.e., not randomised) and clustered at school level. Hence, the trial had a two-arm clustered design – assignment to the *Reading Planet: Rocket Phonics* programme or control conditions was determined at school level, with all pupils within a school receiving the same assignment. This report focuses on outcomes in Year 1 (Autumn 2022, Spring 2023, Summer 2023). Dependent variables and the within-subjects repeated measurement design of these variables is described later, in Primary and secondary outcome measures (see p12).

## Condition

The independent variable condition is measured between-subjects and has 2 levels – *Reading Planet: Rocket Phonics* or control, which are described below. Schools in both conditions completed termly *New PiRA* and *GAPS* spelling subtest assessments for all pupils in Year 1. In addition, schools in the *Reading Planet: Rocket Phonics* condition also completed an in-programme assessment of phonics knowledge at baseline and endline.

### *Reading Planet: Rocket Phonics programme*

*Reading Planet: Rocket Phonics* is a DfE validated SSP programme which aligns to Letters and Sounds. Schools assigned to this condition received the complete programme and package of support for delivery – a suite of materials to support both whole-class teaching and consistent daily phonics practice, as well as training for teachers. This included:

- Teacher’s guides
- Sounds mats and friezes for classroom displays
- 150 A5 flashcards for introducing and revising target letter-sounds
- 48 fully decodable and aligned target practice physical and digital reading books that focus on specific phoneme–grapheme correspondences that have been taught
- 114 additional fully-decodable reading books for further reading practice
- 6 multimedia teaching storybooks to support whole-class teaching
- 6 termly pupil practice booklets for daily practice and application to learning
- Access to CPD videos and webinars with author and phonics expert Abigail Steel
- In-programme phonics assessments to monitor progression at the start and end of each year.

### *Business-as-usual control*

The schools assigned to the control condition were not provided with any of the *Reading Planet: Rocket Phonics* materials and were instead asked to continue with business-as-usual – using their planned teaching methods and resources. These schools reported using a range of different phonics resources, including Little Wandle, Bug Club, Letters and Sounds, Phonics Play, Twinkle, Read Write Inc, FFT and teacher-made resources.

## Trial responsibilities

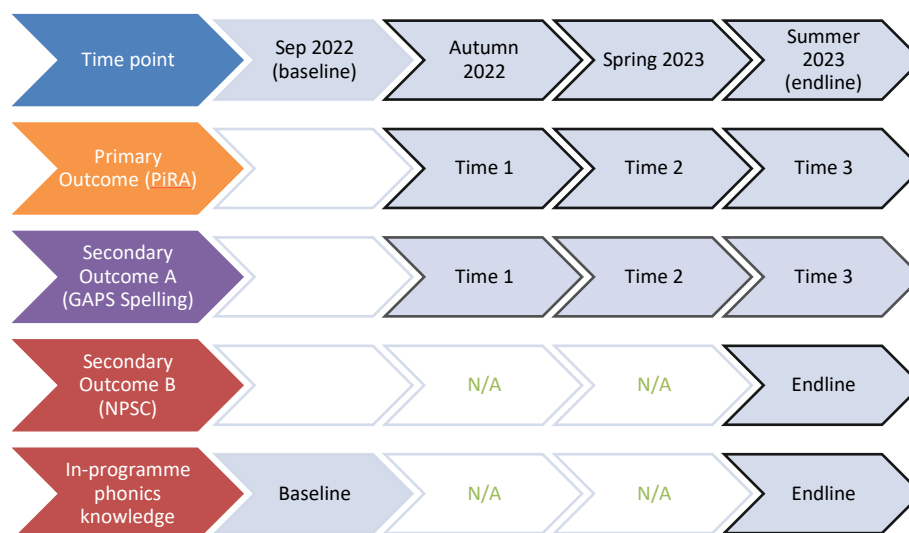
This outcome evaluation report was prepared by an independent evaluator (Dr Helen Breadmore, Associate Professor Psychology in Education, University of Birmingham), who was commissioned by Hodder Education, the publishers of the *Reading Planet: Rocket Phonics* programme. The outcome evaluator consulted on the design of the trial, performed data analysis, and interpreted and reported those results in this report. A process evaluation was conducted by another evaluator (Professor Clare Wood and Violeta Baikousi, Nottingham Trent University), independent both from the outcome evaluator and Hodder Education. Neither evaluator had responsibility for recruitment and retention of participants, schools, nor collection of outcome data. This, along with provision and delivery of the *Reading Planet: Rocket Phonics* programme, was the responsibility of Hodder Education (led by Kate Mayhew, Schools Engagement Manager). Both evaluations were conducted by chartered psychologists in accordance with British Psychological Society ethical guidelines and given favourable opinions by university ethics committees.

## Primary and secondary outcome measures

The primary outcome, *Progress in Reading Assessment (New PiRA)*, was measured repeatedly, within-subjects at 3 time points (Autumn, Spring, Summer). Note that *New PiRA* Autumn was administered after

half-term, in accordance with the manual, and hence not at baseline. Two secondary outcomes were measured – *GAPS* spelling subtest was measured at 3 time points (Autumn, Spring, Summer) and the National Phonics Screening Check was measured at one time point (endline – Summer). In addition, participants in the *Reading Planet: Rocket Phonics* condition completed in-programme assessments of phonics knowledge at 2 time points (baseline – September, endline – Summer). Those scores will be summarised descriptively to illustrate progress through the year. Figure 1 illustrates the measures administered over the course of the year.

**Figure 1: Primary and secondary outcome measures across time.**



*Note: Orange indicates primary outcome measure of literacy attainment, purple indicates secondary outcome measure of spelling, red indicates secondary outcome measure of phonics knowledge.*

### Primary outcome

The primary outcome of interest is overall literacy attainment. **Literacy attainment** was measured using the paper version of the *New PiRA* (Ruttle et al., 2021), a termly standardised and norm-referenced assessment of overall reading ability designed for whole-class administration by the classroom teacher. The assessment takes approximately 40 minutes.

As a general measure of literacy attainment, *New PiRA* has good content validity. The test is aligned to 2014 national curriculum and Key Stage 1 SATs content, and it was recently standardised with a large nationally representative sample in Autumn 2019, Spring 2020 and Summer 2021. Comprehension questions at this level focus on literal comprehension rather than drawing on prior knowledge. The test also has good internal reliability. The manual reports Cronbach's alpha ( $\alpha$ ), confidence bands and SEM for the standardisation sample for each termly assessment (Ruttle et al., 2021). The reported Cronbach's alpha ( $\alpha$ ) for the 3 assessments Year 1 Spring to Year 1 Summer are all in the range considered good to excellent (.84, .91, .82 respectively) (Breadmore & Carroll, 2021; Evers, Hagemester, et al., 2013; Evers, Muñiz, et al., 2013). The norm-derived population appear large, of similar demography to the present cohort, and free from bias. It is worth noting that Autumn and Spring norming data were collected in 2019 before the COVID-19 pandemic, while Summer norming data were collected in 2021 following 2 years of substantial disruption to education. Nonetheless, care has been taken to adjust the standardisation of Summer tests to account for this. *New PiRA* administration in the present study was consistent with the standardised instructions outlined in the manual. All tests were administered on paper, by the classroom teacher, shortly after half term. Teachers scored each item for accuracy using the

marking guide and entered item level raw scores into MARK (My Assessment and Reporting Kit) online. All other scores were then automatically generated through the system.

At each time point, raw *New PiRA* test scores can be converted to summed strand scores, cohort standardised scores, age standardised scores, Hodder scores and a descriptive Performance Indicator. Each test varies in the number of items and alignment to different strands, making it difficult to make comparisons at this level between tests. When using measures of attainment which are closely aligned to curriculum and which have been standardised for use at a specific time point (e.g., Spring Year 1), cohort standardised or raw scores are the most appropriate for interpretation. Where possible, other scores are summarised in this report for descriptive purposes but are not analysed. Cohort standardised scores are particularly helpful for interpreting how important any observed differences between groups are because they are directly meaningful and can be compared across the different versions of the *New PiRA* test. The average pupil at the same stage of education in the normative sample receives a standardised score of 100, and the standard deviation is 15. Hence, a standardised score of 85 is 1 standard deviation below the mean, while a standardised score of 115 is 1 standard deviation above the mean, and so 68% of pupils should receive scores within this range of ability (for further discussion see Breadmore & Carroll, 2021, p. 20; Ruttle et al., 2021, p. 32). On the other hand, given the large sample of pupils who completed the measures in the present study, analyses of raw scores are acceptable but cannot be directly interpreted to understand absolute changes in performance over time. For example, if an individual presents with a lower raw score in Summer than Autumn, that does not mean that their reading ability has decreased, it is more likely to mean that the Summer test is more difficult than the Autumn test.

### *Secondary outcomes*

Secondary outcomes were spelling and phonics knowledge. See Figure 1 for a summary of when each secondary measure was administered.

The first secondary outcome of interest is spelling. **Spelling** was measured using the *GAPS (Progress in Grammar, Punctuation and Spelling Assessment)* spelling subtest (Charles et al., 2018), a termly standardised assessment of overall reading ability designed for whole-class administration by the classroom teacher. The complete assessment takes approximately 40-50 minutes, has been carefully aligned to the curriculum and provides norm-referenced scores with reliability reported in the Technical Manual. However, since only the spelling subtest was administered to reduce testing burden (reducing the administration time to around 20 minutes) we will only report raw scores on this measure, which could vary from 0 to 20.

The next secondary outcome of interest is **phonics knowledge** measured using the National Phonics Screening Check, which was administered in line with government guidance at endline only. The content of the National Phonics Screening Check is determined by the Standards and Testing Agency and is a mandatory assessment which is administered to most pupils in England towards the end of Year 1 each year. In 2023, section 1 contained 12 pseudo-words (*fod, bep, zat, ult, heng, shob, farn, chesh, trun, glork, remp, deeps*) and 8 real words (*chip, hill, jars, shack, twin, crack, tuft, ports*), and section 2 contained 8 pseudo-words (*jer, drave, teap, phibe, slimp, caint, spelt, strave*) and 12 real words (*boy, scoop, foam, white, swept, dreams, scrap, strike, elbow, tailor, sequins, whirling*). The test was administered and scored by the classroom teacher during the test week – week beginning 12th June (Scoring and Testing Agency, 2023a, 2023b). Pupils score 1 mark for each correctly read word and the summed score (out of 40) was used as the dependent variable in the analyses. The threshold mark (released after administration of the test) was 32.

Finally, pupils in the *Reading Planet: Rocket Phonics* programme also completed in-programme phonics assessments, an identical test administered at baseline and endline. The in-programme phonics assessment is curriculum linked with high face validity – it is designed to carefully match the content of the *Reading Planet: Rocket Phonics* programme. Components of the test measure each of the skills that are taught in the *Reading Planet: Rocket Phonics* programme, including letter sound knowledge, oral blending and segmenting, word reading, letter formation, word segmenting, sentence reading and sentence writing. The test is individually administered to pupils by their classroom teacher. Scores on each component are summarised, and a total score generated for the purposes of analyses. This score could range from 0 to 274. This measure has been adjusted following validation in the first year of the trial. The same assessment was administered at baseline and endline to all participants in the *Reading Planet: Rocket Phonics* condition and therefore raw score is appropriate for analyses.

## Data diagnostics

Data were collected by Hodder Education, anonymised, and then shared with the independent evaluator. The format of the data ensured that the independent evaluator could not identify pupils, teachers or schools during the process of data analysis. The evaluator checked data for completeness and out-of-range outliers. Out-of-range scores and outliers were treated as missing data. For example, *New PiRA* raw scores of 0 could reflect refusal to attempt the test for behavioural reasons or failure to provide any correct responses due to ability; in any event, such results cannot be standardised and therefore were treated as missing. No scores on *New PiRA*, *GAPS* spelling or the National Phonics Screening Check were identified to be above the maximum plausible score. However, there were out of range scores on subscales and total score on the in-programme phonics assessment. Participant level exclusionary criteria were applied after data diagnostics was complete (described on p10).



## Results

---

### Data analysis plan

Data analyses were comprised of phases 1-3 for each outcome measure in turn. Data were analysed using SPSS v28.0.1.1 and R v4.3.0 running in RStudio v2023.03.0.

**Phase 1:** Descriptive exploration of outcomes. These analyses provide a summary of the distribution of scores and general pattern of results for the variables of interest without considering the clustering of pupils within schools. The aim is to confirm that the measures are suitable for analysis using inferential statistics. For example, that data do not violate the assumptions of parametric tests.

**Phase 2:** Descriptive exploration of school level cluster scores and tests of baseline equivalence.

**Phase 3:** Inferential statistics empirically test hypotheses relating to primary and secondary outcomes. These analyses test whether changes in performance over the course of the year differ by condition (*Reading Planet: Rocket Phonics*/control). Data were analysed using hierarchical linear mixed effects regression models with maximum likelihood using the lme4 package (v1.1-34) in RStudio, with random intercepts by participant and school (accounting for clustering of participants within schools). The fully specified model is:

$$\text{Outcome score} = \text{Condition} * \text{Time} + (1 | \text{Participant/School})$$

Statistical significance was ascertained using likelihood ratio tests to compare full and null models (interaction terms tested by comparing fully specified model to one with additive effects, main effects compared null main effect to additive model).

### Primary Outcomes (Literacy Attainment)

Table 4 illustrates literacy attainment in Autumn 2022, Spring 2023 and Summer 2023 (Table 8 in Appendix 1 provides comparable primary outcomes for those pupils with complete datasets including secondary outcomes). Cohort standardised scores confirm that pupils were typical of the population – mean standard scores approximate 100 and include the full range of abilities. The overall means are very similar between conditions. On average, *Reading Planet: Rocket Phonics* participants begin with literacy attainment slightly below those in Control schools, and by Spring and Summer they are slightly above control schools. Figure 2 illustrates pupil outcomes at each time point graphically. Note that in both conditions the standardised scores in Spring 2023 have less variance than in Autumn 2022 and 2023. Nonetheless, both conditions appear equally effected by this.



**Table 4: Summary of pupil level scores on primary outcome (Progress in Reading Attainment – New PiRA) at each time point in schools assigned to *Reading Planet: Rocket Phonics* and control conditions.**

	<i>Autumn 2022</i>		<i>Spring 2023</i>		<i>Summer 2023</i>	
	<i>Reading Planet: Rocket Phonics</i>	<i>Control</i>	<i>Reading Planet: Rocket Phonics</i>	<i>Control</i>	<i>Reading Planet: Rocket Phonics</i>	<i>Control</i>
<i>N schools (participants)</i>	11 (409)	13 (398)	11 (409)	13 (398)	11 (409)	13 (398)
<i>Raw New PiRA test scores<sup>1</sup></i>	$\bar{x}$ 14.68 (SD 5.99) 1-25	$\bar{x}$ 14.92 (SD 5.58) 1-25	$\bar{x}$ 13.08 (SD 6.33) 1-25	$\bar{x}$ 12.41 (SD 6.29) 1-25	$\bar{x}$ 16.25 (SD 5.85) 1-25	$\bar{x}$ 15.92 (SD 6.35) 1-25
<i>Cohort standardised scores<sup>1</sup></i>	$\bar{x}$ 99.48 (SD 15.95) 63-127	$\bar{x}$ 100.12 (SD 14.87) 63-127	$\bar{x}$ 100.6 (SD 14.61) 73-128	$\bar{x}$ 99.08 (SD 14.48) 73-128	$\bar{x}$ 99.08 (SD 14.55) 64-125	$\bar{x}$ 98.31 (SD 15.51) 64-125

**Table note:** <sup>1</sup>Mean ( $\bar{x}$ ), standard deviation (SD), range overall (does not account for school clusters).

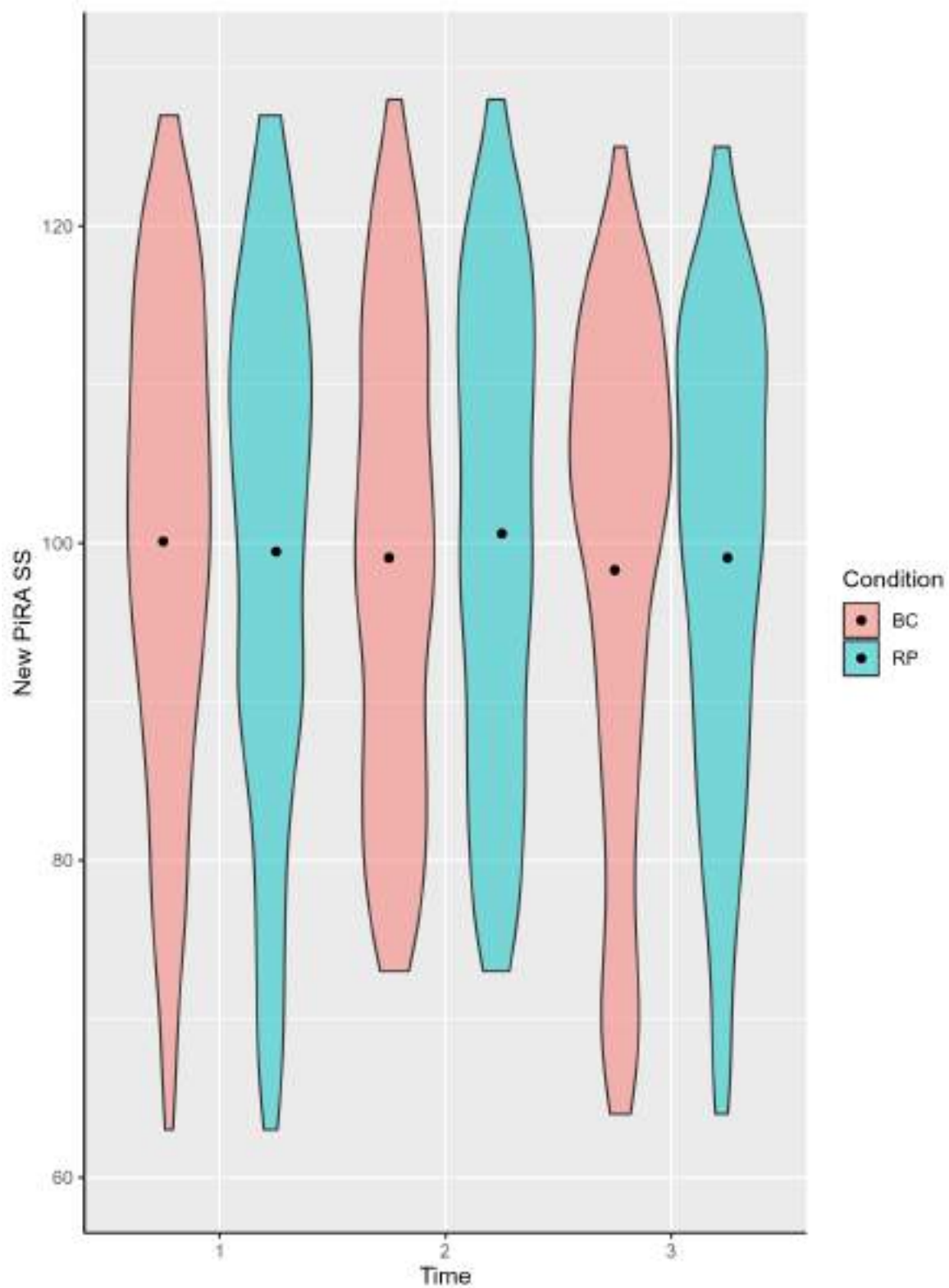


Figure 2: Violin plot illustrating the distribution of pupil level New PiRA standardised scores by time (1 = Autumn 2022, 2 = Spring 2023, 3 = Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics). Mean scores plotted with a point.

Data were aggregated into school level clusters to explore variability in outcomes between schools. A summary of outcomes in *Reading Planet: Rocket Phonics* and control conditions using school level clusters is provided in Table 5 and Figure 3. Note the large degree of variability both within and between clusters. To test for baseline invariance in literacy attainment, we conducted a one-way analysis of variance (ANOVA) with the between-subjects factor condition (school clusters – *Reading Planet: Rocket Phonics*, control) and the dependent variable mean cohort standardised score on the Autumn 2022 *New PiRA*. This indicated that there were no significant differences in baseline knowledge between conditions;  $f(1,22) = 0.27$ ,  $p = .61$ ,  $\eta^2 = .01$ , Cohen's  $D = .21$ . Inspecting the group means confirms that although the participants in *Reading Planet: Rocket Phonics* showed a slight tendency to have lower scores than controls in Autumn 2022, the differences are negligible. Therefore, we do not need to be concerned about baseline literacy attainment when interpreting outcome analyses.

Figure 3 illustrates mean standardised scores clustered by school. This confirms a high degree of variability between schools. Nonetheless, comparing Figure 3 to Figure 2 reveals the amount of variability within schools that is lost by collapsing the data by school. As a result of this, analyses are conducted using linear mixed models with random intercepts by participants and schools to account for participants being clustered within schools while maximising power to detect a significant difference by retaining between-participant variability.

**Table 5: Summary of primary outcomes, averaged within school clusters at each time point for pupils in schools assigned to *Reading Planet: Rocket Phonics* (RP) and Control (C) conditions.**

	RP	C	RP	C	RP	C
<i>N schools (N participants)</i>	11 (409)	13 (398)	11 (409)	13 (398)	11 (409)	13 (398)
<i>Primary outcome: Reading attainment</i>	Autumn 2022		Spring 2023		Summer 2023	
<i>Raw score<sup>1</sup></i>	$\bar{x}$ 14.85 (SD 2.55) 10.37-18.4	$\bar{x}$ 15.45 (SD 2.97) 11.26-20.6	$\bar{x}$ 13.13 (SD 2.83) 7.5-17.5	$\bar{x}$ 13.72 (SD 3.03) 9.89-18.8	$\bar{x}$ 16.42 (SD 2.01) 13.64-20.37	$\bar{x}$ 16.57 (SD 2.52) 12.33-20.48
<i>Reading age equivalence (years; months)</i>	5;04-5;06	5;06-5;09	5;03-5;04	5;03-5;04	5;09-5;11	5;09-5;11
<i>Cohen's D effect size estimate</i>	.22		.20		.07	
<i>Cohort standardised score<sup>1</sup></i>	$\bar{x}$ 99.92 (SD 6.8) 87.92-109.34	$\bar{x}$ 101.5 (SD 7.91) 90.33-115.2	$\bar{x}$ 100.72 (SD 6.55) 87.71-110.82	$\bar{x}$ 102.07 (SD 6.96) 93.22-113.6	$\bar{x}$ 99.49 (SD 5.06) 92.69-109.4	$\bar{x}$ 99.9 (SD 6.2) 90.11-109.96
<i>Cohen's D effect size estimate</i>	.21		.20		.07	

**Table note:** <sup>1</sup>Mean ( $\bar{x}$ ), standard deviation (SD), range within clusters.

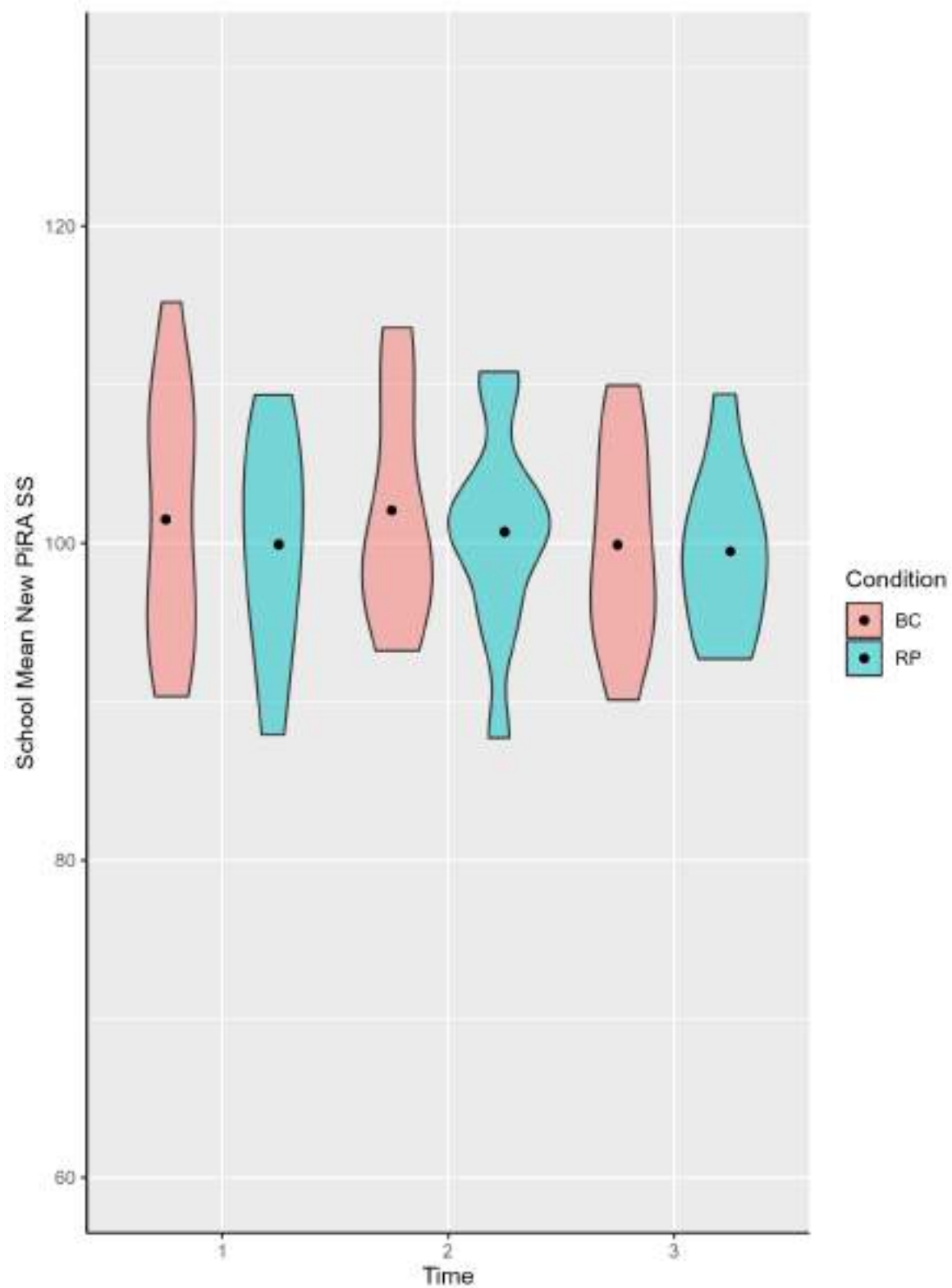


Figure 3: Violin plot illustration of distribution of school mean new PiRA standardised scores illustrated by time (Autumn 2022, Spring 2023, Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics). Mean scores illustrated with a point.

## Impact of condition on primary outcome measures

To assess whether there was a significant effect of condition on outcomes, standardised scores<sup>2</sup> on the *New PiRA* were modelled across the three time points using linear mixed effects models fit by maximum likelihood. A large amount of variance in the model was attributed to participant (137.03) and school (23.52), confirming the importance of specifying random intercepts by participant and school. Model comparison likelihood ratio tests indicated a significant main effect of Time, qualified by a significant interaction between time and condition; Time  $\chi^2(2)=10.3$ ,  $p=.0058$ , Condition  $\chi^2(1)=0.08$ ,  $p=.77$ , Condition\*Time  $\chi^2(2)=7.41$ ,  $p=.025$ . Follow-up tests examined the main effect of condition at each time point. Those tests confirmed that scores did not differ significantly between conditions at any time point, main effect of condition was not significant in Autumn  $\chi^2(1)=0.19$ ,  $p=.66$ , Spring  $\chi^2(1)=0.05$ ,  $p=.82$ , or Summer  $\chi^2(1)=0.01$ ,  $p=.91$ .

To explore the condition\*time interaction, follow-up tests examined the effect of time within each condition at paired time points (Autumn-Summer, Autumn-Spring, Spring-Summer). These analyses explore whether the changes in scores over time are similar or different in the two conditions. The means presented in Table 4 and Table 5 suggests that in the *Reading Planet: Rocket Phonics* condition standard scores increased from autumn to spring and then decreased from spring to summer. The overall change was not significant Autumn-Summer  $\chi^2(1)=0.49$ ,  $p=.48$ , the Autumn-Spring increase in scores was marginal  $\chi^2(1)=3.57$ ,  $p=.059$ , and the Spring-Summer decline was significant  $\chi^2(1)=8.12$ ,  $p=.0044$ . In Control schools on the other hand, the means presented in Table 4 and Table 5 suggest an overall decline in scores, with an initial increase in scores from autumn to spring followed by a decline from spring to summer. The overall decline in scores was significant Autumn-Summer  $\chi^2(1)=8.41$ ,  $p=.0037$ , although the neither the Autumn-Spring  $\chi^2(1)=3.22$ ,  $p=.073$ , Spring-Summer  $\chi^2(1)=2.16$ ,  $p=.14$  changes were. Overall, this suggests that while pupils in control school tended to show a small decline in performance on the *New PiRA* over the course of the year as compared to cohort, pupils in *Reading Planet: Rocket Phonics* showed less of a decline.

## Secondary Outcomes: Spelling and Phonics Knowledge

Table 6 illustrates secondary outcome measures. Raw score on the *GAPS* spelling subtest measure in Autumn, Spring and Summer, while raw score on the National Phonics Screening Check was only measured in Summer 2023. Note that sample size is reduced in these analyses due to missing data (participants were only retained in the analyses if they had a complete dataset), and further reduced in the final set of analyses where only pupils in schools completing the *Reading Planet: Rocket Phonics* programme completed in-programme phonics assessments.

Similarly to analyses of primary outcomes, data were aggregated into school level clusters to explore variability in outcomes between schools (see Table 7 for summary). To test for baseline invariance in spelling, we conducted a one-way analysis of variance (ANOVA) with the between-subjects factor

---

<sup>2</sup> These analyses were also conducted on raw scores. As expected, there was very little difference in the results. A large amount of variance was again attributed to participant (22.36) and school (3.76), confirming the importance of specifying random intercepts by participant and school. Model comparison likelihood ratio tests indicated a significant main effect of Time, qualified by a significant interaction between time and condition; Time  $\chi^2(2)=372.93$ ,  $p<.0001$ , Condition  $\chi^2(1)=0.08$ ,  $p=.77$ , Condition\*Time  $\chi^2(2)=7.97$ ,  $p=.019$ . Follow-up tests examined the main effect of condition at each time point. Those tests confirmed that condition did not have a significant effect at any time point, Autumn  $\chi^2(1)=0.19$ ,  $p=.66$ , Spring  $\chi^2(1)=0.05$ ,  $p=.82$ , Summer  $\chi^2(1)=0.01$ ,  $p=.93$ .

condition (school clusters – *Reading Planet: Rocket Phonics*, control) and the dependent variable mean cohort standardised score on the Autumn 2022 *New PiRA*. This indicated that there were no significant differences in baseline knowledge between conditions;  $f(1,22) = 2.15$ ,  $p = .16$ ,  $\eta^2 = .09$ , Cohen's  $D = -0.6$ . Therefore, we do not need to be concerned about baseline variance between conditions. Baseline equivalence is not tested for the National Phonics Screening Check, which was only assessed at Endline (Summer 2023).

**Table 6: Summary of pupil level scores on secondary outcomes (GAPs spelling subtest, National Phonics Screening Check, in-programme phonics assessments) at each time point in schools assigned to *Reading Planet: Rocket Phonics* and control conditions.**

Autumn 2022			Spring 2023		Summer 2023	
	Reading Planet: Rocket Phonics	Control	Reading Planet: Rocket Phonics	Control	Reading Planet: Rocket Phonics	Control
N schools (participants)	11 318	13 283	11 318	13 283	11 318	13 283
Secondary outcome 1: Spelling	$\bar{x}$ 10.25 (SD 5.32) 1-20	$\bar{x}$ 9.56 (SD 5.00) 1-20	$\bar{x}$ 9.99 (SD 5.54) 1-20	$\bar{x}$ 9.38 (SD 5.34) 1-20	$\bar{x}$ 11.16 (SD 5.46) 1-20	$\bar{x}$ 10.31 (SD 5.51) 1-20
Skewness	0.17		0.27		0.08	
Kurtosis	-0.95		-0.98		-1.20	
Secondary outcome 2: National Phonics Screening Check					$\bar{x}$ 34.96 (SD 5.34) 4-40	$\bar{x}$ 35.40 (SD 4.78) 4-40
Skewness					-2.58	
Kurtosis					9.50	
In-programme phonics assessment	$\bar{x}$ 118.69 (SD 37.54) 30-244				$\bar{x}$ 221.34 (SD 38.32) 78-274	
N participants	297				297	
Skewness	0.47				-1.06	
Kurtosis	0.74				0.62	

**Table note:** <sup>1</sup>Mean ( $\bar{x}$ ), standard deviation (SD), range overall (does not account for school clusters).

**Table 7: Summary of secondary outcomes (raw scores on GAPs spelling subtest, National Phonics Screening Check raw score), averaged within school level clusters for pupils in schools assigned to *Reading Planet: Rocket Phonics* (RP) and control (C) conditions.**

	RP	C	RP	C	RP	C
<i>N clusters (schools)</i>	11	13	11	13	11	13
<i>Secondary outcome 1: spelling</i>	Autumn 2022		Spring 2023		Summer 2023	
<i>N participating pupils in evaluation</i>	318	283	318	283	318	283
<i>Raw score<sup>1</sup></i>	$\bar{x}$ 10.24 (SD 1.88) 6.90-12.39	$\bar{x}$ 9.00 (SD 2.22) 5.1-13.76	$\bar{x}$ 10.01 (SD 1.93) 6.69-12.23	$\bar{x}$ 8.78 (SD 2.21) 3.79-13.47	$\bar{x}$ 11.15 (SD 1.3) 8.97-12.87	$\bar{x}$ 9.74 (SD 2.73) 4.34-13.6
<i>Cohen’s D effect size estimate</i>	-0.60		-0.59		-0.64	
<i>Secondary outcome 2: National Phonics Screening Check</i>					Summer 2023	
<i>N participating pupils in evaluation</i>					318	283
<i>Raw score<sup>1</sup></i>					$\bar{x}$ 35.06 (SD 2.27) 30.55-37.97	$\bar{x}$ 35.96 (SD 2.05) 32.07-38.00
<i>Cohen’s D effect size estimate</i>					0.41	

**Table note:** <sup>1</sup>Mean ( $\bar{x}$ ), standard deviation (SD), range within clusters.



## Impact of condition of secondary outcomes

To assess whether there was a significant effect of exposure to *Reading Planet: Rocket Phonics* on secondary outcomes, separate analyses were conducted on each measure. This was necessary because of variation in when outcomes were measured. The amount of missing data varied in each case and, as a result, samples vary across outcomes. In all cases, outcomes were modelled using linear mixed effects models fit by maximum likelihood, with random intercepts by school (and, where possible, participant). Spelling was assessed at three timepoints (Autumn, Spring, Summer) and so data were analysed similarly to the primary outcome measure (with fixed effects of Time and Condition). The National Phonics Screening Check was assessed only in Summer 2023 and so Time was not considered as a fixed effect. The in-programme phonics assessment was administered in Autumn 2022 and Summer 2023 but only in schools using the Rocket Phonics programme, and so Time (but not Condition) was considered as a fixed effect.

### Spelling

Analyses of spelling outcomes were conducted on data from 601 participants (see Table 3 in Method for summary of demographics). Pupil level outcomes in each condition (*Reading Planet: Rocket Phonics* or Control) at each time point (Autumn, Spring, Summer) are summarised in Table 6 and illustrated graphically in Figure 4. These scores include the full range of ability at each time point, the mean is central on the scale and overall skew and kurtosis at each timepoint were within a reasonable range – the data are normally distributed. Collapsing data by school limits variability – as you can see from the limited range of scores presented in Table 7 and Figure 5. This again highlights the importance of using linear mixed effects models to understand the impact of condition on outcomes.

To assess whether there was a significant effect of condition on outcomes, raw scores on the spelling subtest of *GAPS* were modelled across the three time points using linear mixed effects models fit by maximum likelihood. A substantial amount of variance in the model was attributed to participant (18.43) and school (2.27), confirming the importance of specifying random intercepts by participant and school. Model comparison likelihood ratio tests indicated a significant main effect of Time, a trend for an effect of condition, but the interaction between time and condition was not significant; Time  $\chi^2(2)=42.13$ ,  $p<.0001$ , Condition  $\chi^2(1)=3.04$ ,  $p=.081$ , Condition\*Time  $\chi^2(2)=3.04$ ,  $p=.22$ . Follow-up tests examined the main effect of condition at each time point. Those tests confirmed that scores did not differ significantly between conditions at any time point, main effect of condition was not significant in Autumn  $\chi^2(1)=2.39$ ,  $p=.12$ , Spring  $\chi^2(1)=1.97$ ,  $p=.16$ , or Summer  $\chi^2(1)=3.41$ ,  $p=.065$ . Together with the means illustrated in Table 6 and Table 7, this suggests a trend for pupils in schools using *Reading Planet: Rocket Phonics* to end the year with slightly better spelling ability than those in controls.

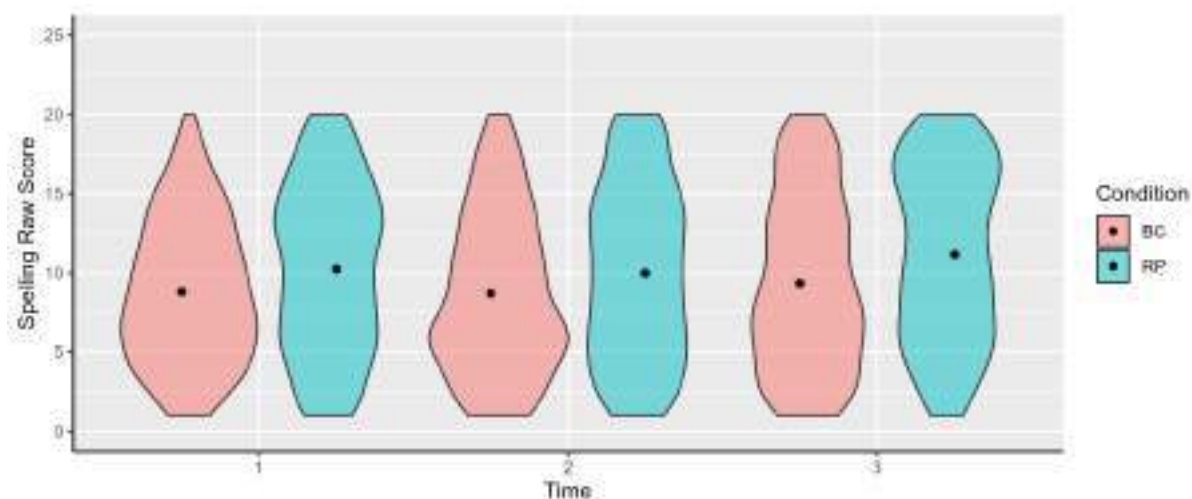


Figure 4: Violin plot illustrating distribution of pupil level *GAPS* spelling subtest raw scores illustrated by time (1 = Autumn 2022, 2 = Spring 2023, 3 = Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics). Mean scores illustrated with a point.

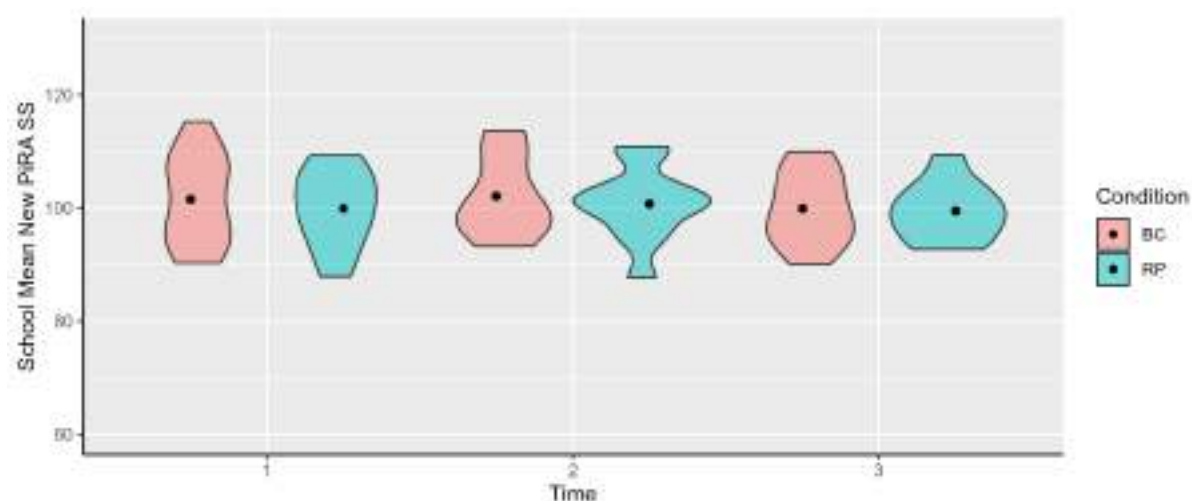


Figure 5: Violin plot illustrating distribution of school mean *GAPS* spelling subtest raw scores illustrated by time (1=Autumn 2022, 2=Spring 2023, 3=Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics). Mean scores illustrated with a point.

### *National Phonics Screening Check*

Analyses of outcomes on the national phonics check were conducted on data from the same 601 participants as the analyses of spelling outcomes. Pupil level outcomes in each condition (*Reading Planet: Rocket Phonics* or Control) are summarised in Table 6 and illustrated graphically in Figure 6. Data are not normally distributed – kurtosis exceeds acceptable levels and ceiling effects are clearly present – one standard deviation above the mean exceeds the maximum possible score on the assessment. Collapsing data by school further limits variability – as you can see by examining the very limited range of scores in Table 7 and Figure 7. Hence data are analysed using linear mixed effects models.

To assess whether there was a significant effect of condition on outcomes, raw scores on the National Phonics Screening Check were modelled across the three time points using linear mixed effects models fit

by maximum likelihood. It was not possible to include random intercepts by participant in these analyses (due to each participant contributing only 1 data point) but random intercepts by school were retained. A substantial amount of variance in the model was attributed to school (3.22), confirming the importance of specifying random intercepts by school. Model comparison likelihood ratio tests indicated that the main effect of condition was not significant  $\chi^2(1)=1.22$ ,  $p=.27$ . Together with the means illustrated in Table 6 and Table 7, this suggests that there were no reliable differences in pupil attainment on the National Phonics Screening Check in *Reading Planet: Rocket Phonics* and control schools.

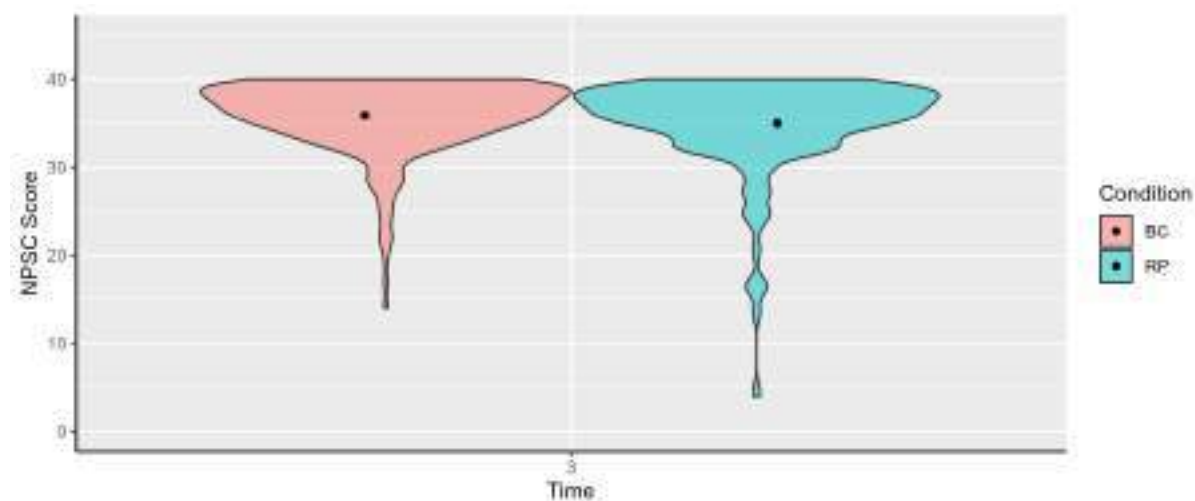


Figure 6: Pupil level raw scores on the National Phonics Screening Check illustrated by time (Autumn 2022, Spring 2023, Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics).

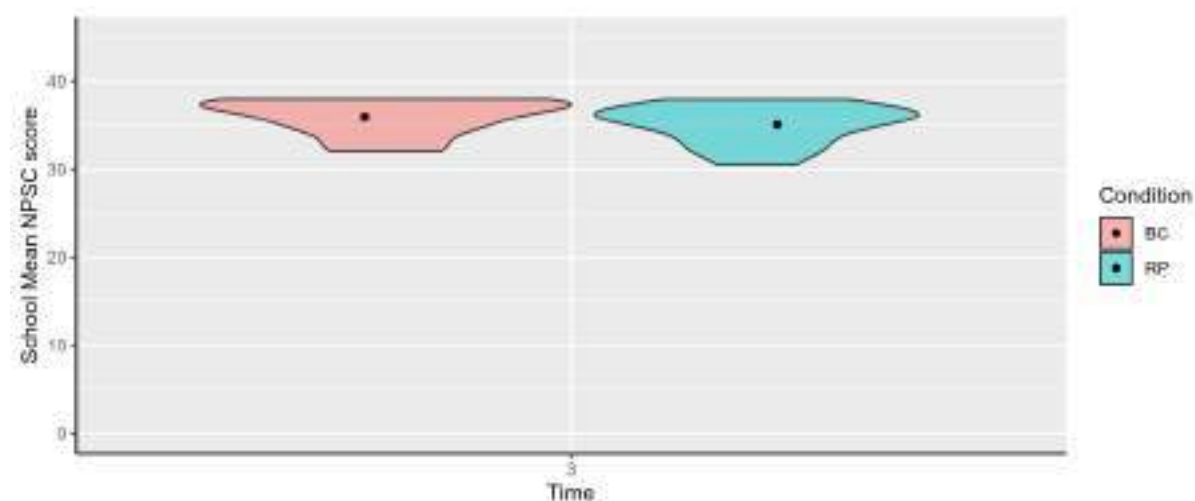


Figure 7: School mean raw scores on the National Phonics Screening Check illustrated by time (Autumn 2022, Spring 2023, Summer 2023) and condition (BC = business-as-usual control, RP = Rocket Phonics).

## In-programme phonics assessment

The in-programme phonics assessment was administered at baseline and endline but only in schools using the *Reading Planet: Rocket Phonics* programme. After removing participants with any missing data on this measure, 297 pupils from 11 *Reading Planet: Rocket Phonics* schools remained. Those data are summarised by participant in Table 6 and Figure 8. There are a wide range of scores at both time points and while there is some skewness at endline, this is within acceptable limits and good degree of variation remains. Collapsing data by school indicates that overall, mean scores at all schools improved over the year, but the amount of change varied (see Figure 9). For example, it is quite striking that the school with the lowest average attainment at baseline had the highest average attainment at endline. To assess whether there was a significant change in phonics knowledge over the course of the year summed scores were modelled across the two time points using linear mixed effects models fit by maximum likelihood. This confirmed that there is a highly significant increase in phonics knowledge from Time 1 to Time 3  $\chi^2(1)=647.51, p<.0001$ , but the random effect of school also contributes a great deal of the variance in the model (122.3). Hence it is clear that the amount of growth differed between schools.

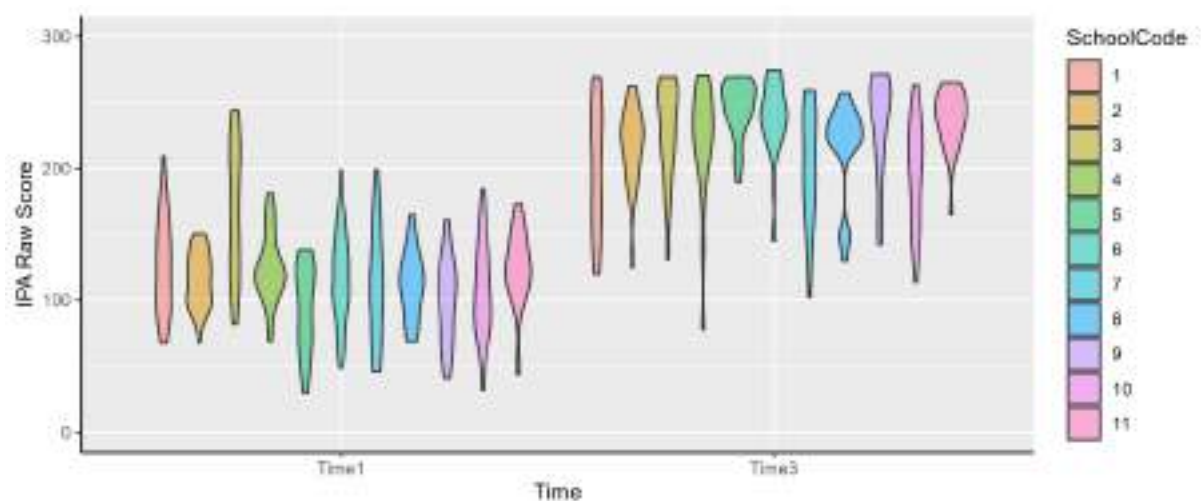


Figure 8: Pupil level scores on in-programme phonics assessment at baseline and endline.

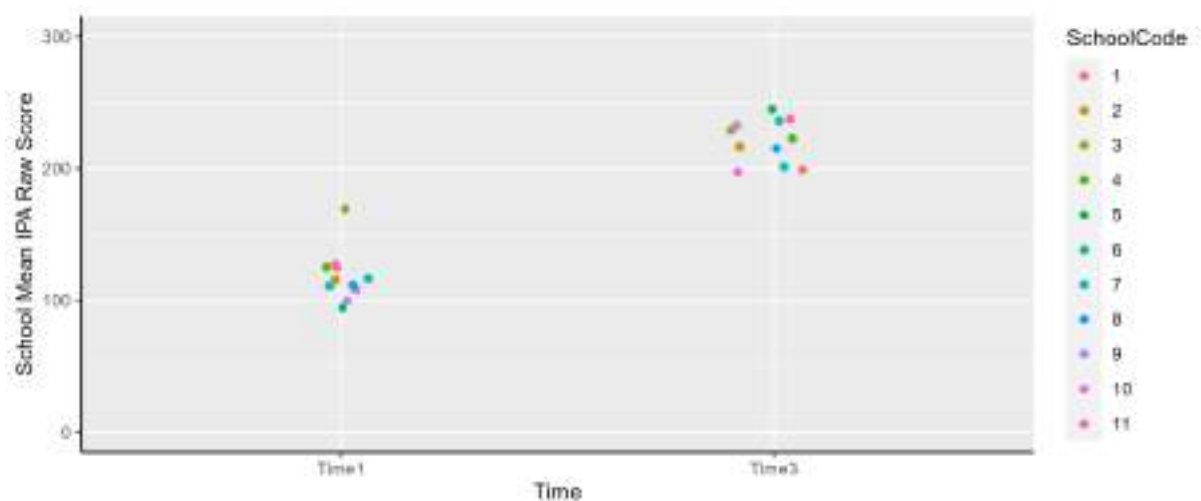


Figure 9: School mean scores on in-programme phonics assessment at baseline and endline.

## Discussion

---

The present evaluation reports the outcomes of a two-year trial of the *Reading Planet: Rocket Phonics* programme. We compare literacy attainment, spelling and phonics knowledge of Year 1 pupils clustered within 11 schools implementing *Reading Planet: Rocket Phonics* (from the beginning of Year R) to 13 business-as-usual controls. Analyses of primary outcomes revealed little evidence that the pupils in schools that were using the *Reading Planet: Rocket Phonics* showed very similar performance to those in controls. There were no significant differences in *New PiRA* cohort standardised scores, the primary outcome. This means the programme had neither positive nor negative effects on literacy attainment. Analyses of secondary outcomes revealed some evidence of promise in spelling attainment. Pupils in schools implementing *Reading Planet: Rocket Phonics* tended to have slightly higher average spelling scores in Summer 2023 than those in control schools, although the difference was not significant. Analyses of in-programme phonics assessments within the schools implementing *Reading Planet: Rocket Phonics* indicated that pupils showed a rapid and significant increase in phonics knowledge. Even those individuals starting Year 1 with very low phonics knowledge showed growth over the course of the year. This is important preliminary evidence for the success of the programme, as the assumed mechanism of change is that *Reading Planet: Rocket Phonics* first increases phonics knowledge, then improves word reading and spelling, and that this in turn support literacy attainment. Nonetheless, there were no significant differences on National Phonics Screening Check scores between pupils in schools implementing *Reading Planet: Rocket Phonics* and controls.

There are several factors related to the design of the trial which limit our ability to draw strong conclusions from these results. These limitations include risk of biases in sampling, concerns about data quality, the timing and validity of outcome measures, programme differentiation and compliance/fidelity. In each case, it is likely that the limitation decreased the likelihood of detecting an effect of implementation of *Reading Planet: Rocket Phonics*. As such, it is unclear whether failure to detect a significant difference between conditions is due to a true lack of effect of the programme, or due to issues with trial design. Each limitation is described below with recommendations of how further research could overcome these limitations.

### Sampling and sample size

Several factors related to the sample and methods of recruitment into the sample limit the conclusions that can be drawn from this study because of the risk that sampling biases have influenced results. A key factor is that schools self-selected the condition of the trial that they joined rather than being randomly assigned. It is not only possible, but likely that the schools that volunteered to implement *Reading Planet: Rocket Phonics* were doing so because they had concerns about the effectiveness of their existing phonics or literacy provision more broadly. In contrast, schools in the control condition may have been more confident in the quality of their existing provision. A key recommendation for future trials is to use stratified random assignment to condition to reduce the risk of biases relating to school assignment to condition.

The summary of background information about schools (see Table 1) highlights some key differences that may have impacted on progress in early literacy instruction, and which differ between conditions. For example, more of the control schools had received recent OFSTED ratings of “Outstanding”, while more of the *Reading Planet: Rocket Phonics* Schools had been rated as “Requires Improvement”. Control schools are typically smaller than those implement *Reading Planet: Rocket Phonics*. School size may be an important factor to consider in future trials – there was wide variation in school size in this study (ranging from 45 to 831 pupils on school role) but there were not enough schools to explore whether outcomes

varied across classes in single-form or multi-form entry schools. The Implementation and Process Evaluation may speak to this point.

## Missing data

A large proportion of data could not be used due to it being incomplete. Indeed, two control schools were removed from analyses due to incomplete outcome measures. Even in schools that were retained in the trial, an unusually large proportion of participants did not provide sufficient data to be included in analyses. Of 1236 participants who entered the trial, only 807 (65%) contributed to primary outcomes analyses and only 601 (49%) contributed to secondary outcome analyses. This presents us with several limitations – the achieved sample is smaller than planned and this limits power to detect a significant difference between conditions, and there is an increased risk of obtaining a biased sample as it is unclear that all participants are equally at risk of missingness.

There are good reasons to suspect that some participants are particularly likely to have missing data, and that certain groups of participants may be disproportionately affected by incomplete data. For example, persistently absent pupils are more likely to absent from school on assessment days, pupils with SEND may sometimes struggle to access assessments. Comparing the demographics of the achieved sample of pupils included in the primary outcomes (see Table 2) to the population summarised in the school census data (presented in Table 1) indeed suggests that the achieved sample were not nationally representative. These data suggest that in the primary outcomes analysis, the achieved sample of participants in schools implementing *Reading Planet: Rocket Phonics* included a greater proportion who had EAL and/or were in receipt of Pupil Premium compared to school census data (43% EAL, 40% PP, compared to 17% and 25% nationally). These pupils are more likely to show an attainment gap and slower growth in attainment. As such, this sampling bias may have limited the effect of the programme. Future research needs to ensure that participants in the achieved sample are more representative of their cohort.

What factors may have contributed to the large proportion of missing data? A key concern in this study is the impact of testing burden, both in terms of the number of assessments that were administered, and teachers having full responsibility for administration, scoring and data input. This could have resulted in higher levels of pupil, teacher or school refusal than had fewer tests had been administered. A smaller battery of tests and fewer timepoints for data collection may reduce the burden on resources and test fatigue, which could reduce missing data in future trials. The quality, completeness and reliability of the dataset would also be increased by using one-to-one or small group assessments, particularly given the young age of participants. Teacher administration is valuable with pupils of this age, as it usually improves ecological validity, reduces test anxiety and helps to ensure that a pupil's performance on the test is more likely to be representative of their usual performance in classroom activities. However, some issues with data quality were clearly identified (e.g., out-of-range scores recorded on the in-programme phonics assessments) and it is impossible to know whether or not there are other, hidden issues with data quality. One way to overcome this in future trials would be to have trained assessors present to invigilate teacher administered assessments, and for trained scorers to input data while implementing data quality checks (e.g., it is common practice to have a second marker check 10% of data input).

Compliance with test administration may also have been influenced by the design decisions to allow schools to self-select to condition and to have uneven incentives for participation in the trial. Schools implementing *Reading Planet: Rocket Phonics* were incentivised by receiving all resources needed to deliver the programme in addition to the assessments for free. They are likely to have maintained a much closer relationship with the delivery team throughout and may have felt a particular responsibility to complete assessments as a result. Control schools were already using the assessments, and were



incentivised by receiving all assessments for free. This may have encouraged them to administer some assessments, but perhaps not have been enough to ensure that they administered a representative sample of assessments. There is evidence in attrition rates to support this hypothesis, not only were two control schools removed from the trial due to missing outcome measures, but of those included 43% of participants in the remaining 13 control schools (283/651) had sufficient data to contribute to secondary outcomes compared to 54% of participants in 11 *Reading Planet: Rocket Phonics* schools (318/585). Future trials should consider how to incentivise all schools to provide more complete data, but particularly control schools. It is also worth noting that control schools were also more experienced in administering the assessments, which could potentially also have impacted on their results.

## Timing of outcome measures

The timing of when the outcome measures were administered may also have limited power to detect a significant difference. In particular, the first time point for the administration of *New PiRA* and *GAPS* spelling subtests was not at baseline in Year R, when literacy instruction first began to differ between conditions. Instead, the first time we could compare literacy attainment and spelling was after Autumn half-term in Year 1, and differences in phonics knowledge could not be assessed until endline. As a result, there is no way to know whether pupils' had begun the year or started school with differences in phonics knowledge in literacy attainment. This is important to establish because children who begin the year with lower literacy skills than their peers are likely to end the year with lower literacy skills than their peers, and usually the gap widens over time – these weaknesses often impact on the amount of progress that is made as well current attainment. While the Interim Evaluation found no significant differences in baseline phonics knowledge at the beginning of Year R (Breadmore, 2023), this was true only after excluding outliers – those pupils with unusually high phonics knowledge on school entry. There are several reasons why this cannot be assumed to also be true for the present study due, but it cannot be assessed due to differences in control group sampling in Year 1 compared to Year R.

Indeed, it seems likely that differences in control group sampling may have influenced the likelihood of detecting significant differences between conditions at the end of Year 1. The interim report showed promising signs of pupils in schools implementing *Reading Planet: Rocket Phonics* having higher literacy attainment than controls in Summer Year R (a nonsignificant mean 5 standard score advantage on *New PiRA*). In contrast, whereas the pupils in the present control group had very similar or slightly higher literacy attainment than those implementing *Reading Planet: Rocket Phonics* in Autumn Year 1 (see Table 4). This could mean that the initial progress was not sustained, or it could be due to differences in the control sample. Notwithstanding the challenges of assessing pupil knowledge and attainment on school entry, given that SSP programmes aim to support the earliest stages of literacy acquisition (decoding) it is crucial that future studies measure phonics knowledge at the beginning of Year R, and control group sampling should not change between Year R and Year 1.

## Validity of outcome measures

The assumed mechanism for change in outcome scores is that *Reading Planet: Rocket Phonics* initially increases phonics knowledge, then the application of that knowledge to read and spell individual words, and that this in turn leads to increases in subsequent literacy attainment. Future trials need to test this mechanism of change more carefully, aiming to understand relationship between early phonics knowledge and subsequent literacy attainment. This is important, because while the decoding skills acquired from high quality phonics programmes are an important first step in learning to be literate, they are certainly not sufficient to achieve skilled literacy. It seems important to assess what level of basic competency in phonics knowledge is sufficient, and at what stage literacy instruction should move beyond phonics to ensure that all children are able to read, spell and write well. For example, there could

be a more careful examination of the influence of the programme on understanding of orthographic and morphological conventions (such as affixes and root morphemes), vocabulary development, as well as text level skills such as knowledge about syntax, grammar, inferencing.

### Programme differentiation, compliance and fidelity

One of the challenges for assessing the impact of any phonics programme in England is limited programme differentiation – business-as-usual in control schools will also have been an SSP programme. Further research should focus on better understanding the differences in their implementation of SSP compared to *Reading Planet: Rocket Phonics* and use close measures to assess whether those differences lead to measurable changes in outcomes. The present study offers some evidence of promise in relation to this – *Reading Planet: Rocket Phonics* particularly focuses on consolidating learning through application of phonics knowledge to both word reading and spelling throughout the programme. Consistent with this, spelling attainment appeared to be developing somewhat more rapidly than control (although differences were not significant).

The implementation and process evaluation {Wood, 2023 #6978} suggested that some schools were not implementing *Reading Planet: Rocket Phonics* with high fidelity. The programme is a highly resourced complex intervention and it seems that not all users choose to use all of the available resources. As such, it is important to understand whether and how implementation fidelity influences the impact of the programme and, more specifically, which components of the programme are the non-negotiable “active ingredients” and which aspects teachers have the flexibility to adapt to their context to support implementation. Future trials should explore this by including a wider range of settings and exploring differences in implementation in different classes within schools as well as between schools.



## Conclusion

---

In conclusion, this outcome evaluation shows limited evidence of promise in Year 1, with no significant differences emerging between conditions in any outcome measures. Pupils in schools implementing *Reading Planet: Rocket Phonics* increased their phonics knowledge during Year 1, performing similarly to pupils in control schools in the National Phonics Screening Check at the end of the year. On average, pupils in *Reading Planet: Rocket Phonics* appeared to have slightly better spelling scores in Summer than those in control schools, although these differences were not significant. There were no significant differences in literacy attainment, although the means did indicate small positive changes. Pupils in *Reading Planet: Rocket Phonics* schools had standard scores slightly below controls in Autumn 2022, which had shifted to slightly above controls in Spring 2023. However, those differences were small and were not statistically significant. It is difficult to draw conclusions from null results, which indicate that implementation of *Reading Planet: Rocket Phonics* resulted in no measurable changes in outcomes compared to business-as-usual. However, a null result should not be interpreted negatively, particularly in the context of business-as-usual controls also delivering phonics programmes for early literacy. There are a number of issues with trial design that need to be resolved in future to better understand the impact of the programme. Taken together, the evidence from the Interim Evaluation and the present evaluation provides some promising evidence that *Reading Planet: Rocket Phonics* supports the earliest stages of literacy acquisition. A larger scale randomised-controlled trial with baseline assessment of phonics knowledge and endline assessment of phonics knowledge, word reading, spelling and literacy attainment at the end of Year R and Year 1 could lead to better understanding of the impact of the programme. Clear compliance and fidelity indicators would also enable consideration of the impact of the quality of programme delivery, with a view to understanding variation of implementation within and between schools.

## References

---

- Breadmore, H. L. (2023). *Interim Outcome Evaluation: Reading Planet Rocket Phonics from Rising Stars 2021-2023*. R. S. f. H. Education. <https://www.risingstars-uk.com/getmedia/e3ef5f86-3845-4f4d-8a52-2df144b4d855/Interim-Outcome-Report.pdf>
- Breadmore, H. L., & Carroll, J. M. (2021). *A sytematic review of standardised measures of attainment in literacy, mathematics, and science*. Evidence Review commissioned by the Education Endowment Foundation. [https://educationendowmentfoundation.org.uk/public/files/Review\\_of\\_Attainment\\_Measures\\_final.pdf](https://educationendowmentfoundation.org.uk/public/files/Review_of_Attainment_Measures_final.pdf)
- Breadmore, H. L., Vardy, E. J., Cunningham, A. J., Kwok, R. K. W., & Carroll, J. M. (2019). *Literacy Development: Evidence Review*. Education Endowment Foundation. [https://educationendowmentfoundation.org.uk/public/files/Literacy\\_Development\\_Evidence\\_Review.pdf](https://educationendowmentfoundation.org.uk/public/files/Literacy_Development_Evidence_Review.pdf)
- Breadmore, H. L., & Wood, C. (2023). *Interim evaluation report: The Reading Planet Rocket Phonics programme 2021-2023 impact trial*. R. S. f. H. Education. [https://www.risingstars-uk.com/media/Rising-Stars/Reading/Rising-Stars\\_Rocket-Phonics\\_impact-report\\_pf5.pdf](https://www.risingstars-uk.com/media/Rising-Stars/Reading/Rising-Stars_Rocket-Phonics_impact-report_pf5.pdf)
- Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of educational psychology*, 91(3), 403-414. <Go to ISI>://000082829900001
- Charles, A., Clarkson, B., McDonald, A., Metcalfe, C., Ruttle, K., Short, R., & Skeffington, S. (2018). *Progress in Grammar, Punctuation and Spelling Assessment. Manual Stage 1: Years 1-2*. RS Assessment from Hodder Education.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the national reading panel's meta-analysis. *Reading Research Quarterly*, 36(3), 250-287. <https://doi.org/10.1598/RRQ.36.3.2>
- Evers, A., Hagemeister, C., Høstmælingen, A., Lindley, P., Muñiz, J., & Sjöberg, A. (2013). *EFPA review model for the description and evaluation of psychological and educational tests: Test review form and notes for reviewers*.
- Evers, A., Muñiz, J., Hagemeister, C., Høstmælingen, A., Lindley, P., Sjöberg, A., & Bartram, D. (2013). Assessing the quality of tests: revision of the EFPA review model. *Psicothema*, 25(3), 283-291. <https://doi.org/10.7334/psicothema2013.97>
- Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science*, 23(6), 572-577. <https://doi.org/10.1177/0956797611435921>
- Rose, J. (2006). *Independent review of the teaching of early reading. Final Report, March 2006*.
- Ruttle, K., McCarty, C., Lallaway, M., Bennett, M., Pepper, L., Kilburn, V., . . . Hiorns, C. (2021). *New PiRA primary: Progress in Reading Assessment Year R, Key Stage 1 and Key Stage 2 Test Guidance (September 2021)*. RS Assessment from Hodder Education.
- Scoring and Testing Agency. (2023a). *Phonics screening check: 2023 scoring guidance*. C. Copyright. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1165181/2023\\_phonics\\_scoring\\_guidance.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1165181/2023_phonics_scoring_guidance.pdf)
- Scoring and Testing Agency. (2023b). *Phonics screening check: Administration guidance*. C. Copyright. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1142992/2023\\_Phonics\\_screening\\_check\\_administration\\_guidance.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1142992/2023_Phonics_screening_check_administration_guidance.pdf)
- Secretary of State for Education. (2022). *Opportunity for all: Strong schools with great teachers for your child*. Crown Copyright Retrieved from <https://www.gov.uk/government/publications/opportunity-for-all-strong-schools-with-great-teachers-for-your-child>

Torgerson, C., Brooks, G., Gascoine, L., & Higgins, S. (2018). Phonics: reading policy and the evidence of effectiveness from a systematic 'tertiary' review. *Research Papers in Education*, 1-31.  
<https://doi.org/10.1080/02671522.2017.1420816>

## Appendix 1

---

**Table 8: Summary of pupil level scores on primary outcome (Progress in Reading Attainment – *New PiRA*) at each time point in schools assigned to *Reading Planet: Rocket Phonics* and control conditions limited to only those participants also included in secondary outcomes analysis.**

	<i>Autumn 2022</i>		<i>Spring 2023</i>		<i>Summer 2023</i>	
	<i>Reading Planet: Rocket Phonics</i>	<i>Control</i>	<i>Reading Planet: Rocket Phonics</i>	<i>Control</i>	<i>Reading Planet: Rocket Phonics</i>	<i>Control</i>
<i>N schools (participants)</i>	11 (318)	13 (283)	11 (318)	13 (283)	11 (318)	13 (283)
<i>Raw New PiRA test scores<sup>1</sup></i>	$\bar{x}$ 14.68 (SD 5.99) 1-25	$\bar{x}$ 14.92 (SD 5.58) 1-25	$\bar{x}$ 13.08 (SD 6.33) 1-25	$\bar{x}$ 12.41 (SD 6.29) 1-25	$\bar{x}$ 16.25 (SD 5.85) 1-25	$\bar{x}$ 15.92 (SD 6.35) 1-25
<i>Cohort standardised scores<sup>1</sup></i>	$\bar{x}$ 99.48 (SD 15.95) 63-127	$\bar{x}$ 100.12 (SD 14.87) 63-127	$\bar{x}$ 100.6 (SD 14.61) 73-128	$\bar{x}$ 99.08 (SD 14.48) 73-128	$\bar{x}$ 99.08 (SD 14.55) 64-125	$\bar{x}$ 98.31 (SD 15.51) 64-125

**Table note:** <sup>1</sup>Mean ( $\bar{x}$ ), standard deviation (SD), range overall (does not account for school clusters).