ABRACADABRA (ABRA) and Reading and Understanding in Key Stage 1 (RUKS)

Evaluation Report

September 2022

Kerry Bell, Hannah Ainsworth, Louise Elliott, Caroline Fairhurst, Louise Gascoine, Pam Hanley, Andrew Haynes, Catherine Hewitt, Alex Mitchell, Carole J. Torgerson, David J. Torgerson, Katie Whiteside.

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

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

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

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

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

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

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from the University of York and Durham University:

Kerry Bell, Hannah Ainsworth, Louise Elliott, Caroline Fairhurst, Louise Gascoine, Pam Hanley, Andrew Haynes, Catherine Hewitt, Alex Mitchell, Carole J. Torgerson, David J. Torgerson, Katie Whiteside.

The principal investigator was Dr Kerry Bell with maternity cover provided by Hannah Ainsworth (October 2017 to October 2018).

Contact details
Dr Kerry Bell
York Trials Unit
Lower Ground Floor ARRC
Department of Health Sciences
University of York
YO10 5DD

Tel: 01904 321325
Email: **kerry.bell@york.ac.uk**

Acknowledgements

We would like to give thanks to all the schools, teachers, teaching assistants, and children who participated in the study. We would also like to thank all the regional trainers for their support of the project. A special thanks to the full implementation team, Professor Clare Wood, Dr Janet Vousden, Dr Helen Johnson, Dr Luisa Tarczynski-Bowles, Professor Rob Savage, Professor Phil Abrami, Professor Anne Wade, Dr Anna Cunningham, and Ms Joanne Lloyd. Finally, many thanks to the EEF for supporting this research, particularly Anneka Dawson, Dr Florentina Taylor, Thomas Martell, Matthew van Poortvliet, Katharina Keck, Dr Rachael Morris, Sarah Tillotson, and Aoife Duff.

Executive summary

The project

ABRACADABRA (ABRA) is web-based interactive software designed to develop pupils' literacy skills through phonics, fluency, and comprehension activities built around a series of age-appropriate texts. The ABRA software was developed by a team from Concordia University and McGill University. Based on ABRA, researchers at Nottingham Trent and Coventry Universities have developed Reading and Understanding in Key Stage 1 (RUKS), a 20-week reading support programme that can be delivered either through an online, ICT-based model that uses the ABRA software or through an offline, paper-based model. The ICT and paper-based models cover equivalent content, aiming to improve children's decoding and reading comprehension skills, thereby improving their reading attainment. Both versions of the programme begin with one and a half days of training for teachers and teaching assistants who then deliver four 15-minute sessions per week to all Year 1 children (age five to six) in small groups of four to five pupils.

One hundred and fifty-seven schools and 4,007 pupils took part in this effectiveness trial. It was a three-armed, cluster randomised controlled trial with randomisation at the school level. Most schools were randomised to one of three arms: ICT, paper-based, or control (business as usual). Schools without the ICT facilities needed for the ICT programme were randomised to one of two arms: paper-based or control. The primary outcome was pupil reading attainment measured using the Progress in Reading Assessment (PiRA). The secondary outcomes included measures of decoding, letter-sound knowledge, and reading attitudes. A process evaluation was also conducted, including observations, interviews, focus groups, and surveys. The evaluation started in January 2018 and finished in July 2019.

Table 1: Key conclusions

Key conclusions	
1.	Children in schools receiving the paper-based programme made the equivalent of two additional months' progress in reading, on average, compared to children in control schools. This result has a high security rating.
2.	Children in schools receiving the ICT-based programme made no additional progress in reading, on average, compared to children in control schools. This result has a high security rating.
3.	Children in both the paper-based and ICT-based programme schools made additional progress in decoding and letter-sound knowledge compared to children in control schools. The paper-based programme also had a small positive impact on pupil attitudes towards reading.
4.	The implementation and process evaluation found that the training model worked well. Schools felt well supported to implement the programmes and reported that they were able to implement it as intended, although a lower proportion of deliverers in the ICT arm reported delivering the intended number of sessions. Around half of deliverers felt the programmes were of equal or better standard than their usual provision and around half said they would use it again (58% in the paper-based arm and 45% in the ICT arm).
5.	79% of the deliverers of the ICT programme reported experiencing technology issues. These largely centred on unstable internet connections. This may help to explain the lack of impact on reading attainment observed for the ICT programme. Some teachers also suggested that the Canadian pronunciation and cultural context in the ICT programme may have been challenging for pupils.

EEF security rating

These findings have a high security rating. This was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of schools. The trial was a well-designed three-armed randomised controlled trial and was well-powered. Fifteen percent of pupils who started the trial were not included in the final analysis due to pupil absence on testing days or because their school withdrew from the evaluation. Pupils in the three trial arms were similar in terms of prior attainment.

Additional findings

Pupils in schools receiving the paper-based ABRA programme made, on average, two additional months' progress in reading (measured using the Progress in Reading Assessment) compared to those in the control group equivalent while pupils in schools receiving the ICT programme made, on average, no additional progress compared to the control group. These are our best estimates of impact, which have a high security rating. As with any study, there is some uncertainty around the result: the possible impacts of the paper-based programme range from no additional progress in reading to positive effects of up to four months' additional progress while possible impacts of the ICT programme range from negative effects of two months less progress to positive effects of up to two months' additional progress in reading.

While only the paper-based programme had positive impacts on overall reading attainment compared to business as usual in control schools, both versions of the programme had positive impacts on measures of pupils' decoding and letter-sound knowledge. Pupils in schools receiving the paper-based programme made an average of two months' additional progress in decoding and two months' additional progress in letter-sound knowledge compared to pupils in control schools. Similarly, pupils in schools receiving the ICT programme made an average of one month's additional progress in decoding and two months' additional progress in letter-sound knowledge compared to pupils in control schools. A small positive impact on pupils' attitudes to reading was also identified for pupils in schools receiving the paper-based programme compared to pupils in control schools and schools receiving the ICT programme. There were no notable differences in the impacts of either programme on reading attainment for pupils eligible for free school meals compared to other pupils.

Data gathered for the implementation and process evaluation suggests that the training model worked well and prepared staff to deliver the programme. Adaptations to the programme, such as altering activities to meet the needs of specific children, were more common in the paper-based arm than in the ICT arm. This may have contributed to the greater impact observed for the paper-based programme. Findings from pupil focus groups suggest that pupils were engaged with both versions of the programme, though more so for the ICT version where children enjoyed the game-like interface.

In the EEF's previous efficacy trial of ABRA (McNally et al., 2016), positive impacts on overall reading attainment were observed for both versions of the programme while in the present trial positive impacts were only found for the paper-based programme. Technology issues experienced by schools delivering the ICT programme in this evaluation may have contributed to the lack of impact observed for the ICT arm. In the previous trial, schools in the ICT arm were provided with a laptop with an offline version of the ABRA software installed, meaning schools were not reliant on internet connectivity. Use of the programme online was deemed a more feasible and scalable approach for this effectiveness trial.

Cost

The average cost of the ICT programme for one school was £1,585 or £20 per pupil per year when averaged over three years. The average cost of the paper-based programme was £1,915, or £25 per pupil per year when averaged over three years. Schools also covered staff time costs. Staff delivering the intervention attended one and a half days of training, delivered ABRA for one hour a week for 20 weeks, and spent around 25 minutes a week on session preparation.

Impact[1]

Table 2: Summary of impact on primary outcome

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P Value	EEF cost rating
Reading: ICT vs Control[1]	0.00 (-0.14, 0.14)	0	🔒🔒🔒🔒🔓	2123 (998, 1125)	0.99	£ £ £ £ £
Reading: ICT vs Control (everFSM)[1]	0.03 (-0.16, 0.22)	0	N/A	519 (263, 256)	0.75	N/A
Reading: Paper-based vs Control[2]	0.15 (0.01, 0.30)	2	🔒🔒🔒🔒🔓	2401 (1172, 1229)	0.04	£ £ £ £ £
Reading: Paper-based vs control (everFSM)[2]	0.12 (-0.05, 0.29)	2	N/A	604 (312, 292)	0.16	N/A

[1] Model 1 includes schools that had sufficient ICT facilities and so were randomised between the three trial arms.
[2] Model 2 includes all schools in model 1 (excluding those randomised to the ICT group), plus the eight schools which were only randomised between paper-based and control (that is, schools without ICT facilities).

[1] The impact evaluation of ABRA was done using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

Introduction

Background

Poor literacy skills affect individuals at every stage of their life. Children with low literacy levels are less likely to achieve at school; adults with low literacy levels have reduced job opportunities (Mulcahy et al., 2019). In England, national Key Stage 2 (KS2) results have previously indicated that only 73% of pupils met the expected standard in reading (DfE, 2019**).** Given that KS2 attainment is a predictor of GCSE attainment (Benton and Sutch, 2014), this is of considerable concern. As such, it is important that research continues to identify effective approaches to increase early literacy skills. A tertiary review recommended that interventions including phonics approaches to increase reading acquisition should be evaluated in large scale randomised controlled trials (RCTs; Torgerson et al., 2018) whilest an EEF-funded review of the use of teaching assistants (TAs) found beneficial impacts on pupil attainment when TAs were used to deliver structured small group interventions (Sharples et al., 2015).

ABRACADABRA (ABRA) is a web-based, interactive software that provides modular 'game'-based activities covering the four broad skill areas for literacy development. The ABRA resources are freely available and have been widely used in Canada, Australia, China, Hong Kong, Kenya, and, more recently, Rwanda. ABRA provides alphabetics (phonics and phonemic awareness), fluency, comprehension, and writing activities around a series of age-appropriate texts with the aim of increasing skills in reading. To engage children in reading and writing and to increase their motivation, the gaming elements are integrated. The majority of individual games are linked in terms of content to a series of electronic texts (mainly 'stories', some nonfiction) suitable for beginner readers. Extension activities for some of the tasks within ABRA can be found in the 'Teacher Area' of the site.

A number of studies have evaluated ABRA in a wide range of contexts and conditions, though many of these are generally longitudinal and quasi-experimental rather than randomised controlled trials (RCTs), which are considered the gold-standard for effectiveness research (Hariton and Locasciio, 2018). However, several small-scale RCTs led by the developers of the software have been conducted in Canada, where the ABRA toolkit was first developed (Comaskey et al., 2009; Savage et al., 2009), as well as a larger effectiveness trial also conducted in Canada (Savage et al., 2013). A synthesis of the research, conducted by the developers, revealed positive effects in student literacy achievement (Abrami et al., 2020) in a variety of different geographic and classroom contexts.

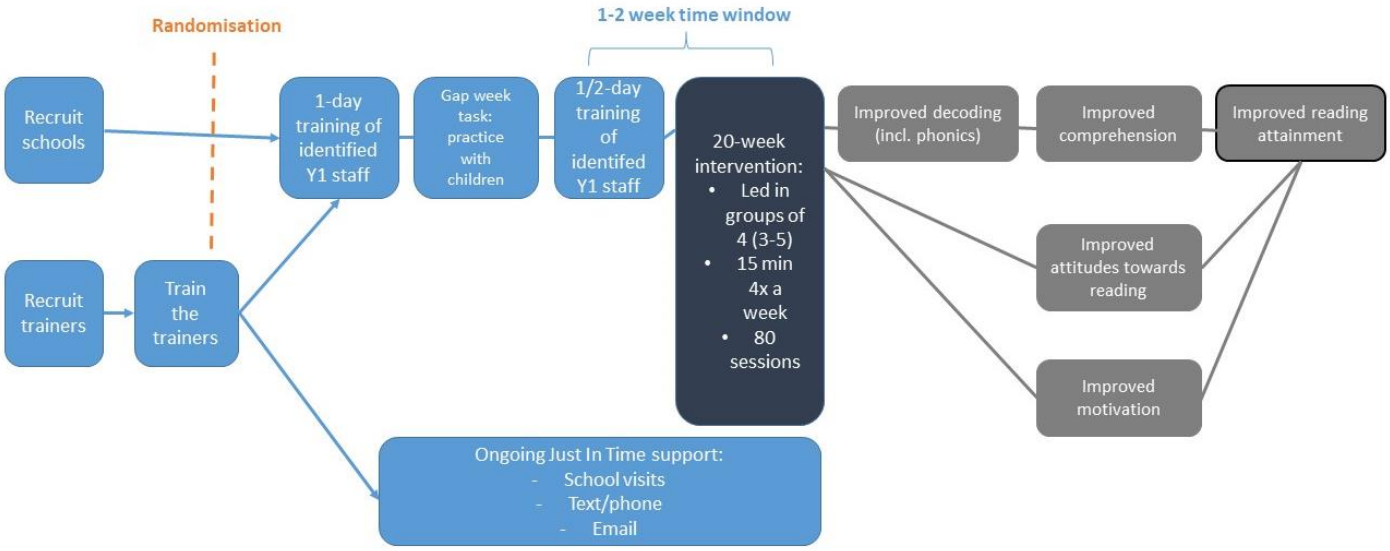
Based on the ABRA platform, researchers at Nottingham Trent and Coventry Universities (hereafter referred to as 'the implementation team') developed the Reading and Understanding in Key Stage 1 (RUKS) programme, a 20-week reading support programme that is non-targeted and takes place in Year 1 of primary school. School staff deliver the programme to small groups of Year 1 pupils using either the ABRA online software or a comparable paper-based method. In 2016 an EEF-funded efficacy trial found that pupils who received the RUKS programme delivered online via ABRA as a computer-based programme (ICT) or the equivalent paper-based programme made between two- and three-months' additional progress in literacy compared to pupils who received standard provision (McNally et al., 2016). A more marked effect was observed for pupils eligible for free school meals ('FSM pupils') and those with below average pre-test reading scores (McNally et al., 2016). The findings were consistent with evidence from the EEF Teaching and Learning Toolkit, which indicates that technology is most effective when used to facilitate new approaches to teaching and learning rather than as an end in itself (), and with the EEF Key Stage 1 Literacy Guidance, which recommends a balanced and engaging approach to developing reading that involves teaching both decoding and comprehension skills (as covered by the RUKS programme) (Higgins et al., 2020).

Given the positive findings arising from the efficacy trial, the EEF funded the present effectiveness trial to evaluate the impact of the reading support programme when delivered at scale and to investigate further any differences between the ICT programme using the ABRA software and the equivalent paper-based reading support programme. Differences will be explored both in terms of the impact, as measured through the primary and secondary outcomes discussed below, and through the implementation and process evaluation. We report the findings of a three-armed, pragmatic ('real world'), cluster RCT.

RUKS programme

The RUKS programme is based around the ABRA software and was developed by the implementation team. It comprises a 20-week programme of 15-minute lessons of a particular structure (rather than teachers choosing from available activities). The implementation team describe the main aim of the RUKS programme as increasing reading ability through a balanced literacy curriculum that supports phonics development while also working on fluency, vocabulary, and comprehension. The logic model in Figure 1 outlines the mechanism by which the programme was expected to work. The model starts with the recruitment of schools and the cascade, 'train-the-trainers' model of training with ongoing just-in-time support provided to schools throughout implementation. Implementation involves four 15-minute sessions per week for at least 20 weeks, a total of at least 80 sessions. The core hypothesis of the logic model is that the programme leads to improvements in decoding skills, including phonics, and comprehension and subsequent improvement in overall reading attainment. The programme was also anticipated to improve attitudes towards reading and motivate children to read, which both feed into the main outcome of improved reading attainment.

Figure 1: RUKS programme logic model



For this evaluation, the RUKS reading support programme was delivered in one of two ways: online using ABRA (ICT delivery model) or via equivalent paper-based materials, prepared in collaboration with Professor Morag Short. As described in the efficacy trial report (McNally et al., 2016), Robert Slavin, director of the Center for Research and Reform in Education at Johns Hopkins University, conducted an independent review of the two programmes on behalf of the EEF that concluded that the two iterations of the programme are well matched in terms of content and delivery, the content of the activities, and the processes involved in delivering them. The programme, irrespective of mode of delivery, comprises decoding, fluency, and comprehension activities based around a series of age-appropriate texts.

Members of school staff who received specialist continuing professional development (CPD) training were responsible for delivering the programmes in school. A pragmatic approach was taken to training with a train-the-trainers, 'cascade' model being implemented. This differed from the training model evaluated in the efficacy trial where members of the implementation team from Coventry University delivered the training directly to schools. Members of the implementation team delivered training sessions to trainers from a variety of backgrounds—principally from training organisations, consultancies, and universities. These trainers, referred to hereafter as 'regional trainers', then delivered training to school staff in their respective regions. Regional trainers received five days of training: one three-day course followed by two one-day courses at a central venue. Additionally, regional trainers were required to undertake two days of preparation in their own time. Regional trainers were trained to deliver training, to provide support via scheduled school visits, and provide ad hoc support for both the ICT delivery model (ABRA) and the paper-based delivery model reading programmes. They received additional support from members of the implementation team where required. Eleven regional trainers were recruited, each local to one of the five recruitment regions—Manchester, Teesside, Newcastle, West Midlands, and East Midlands. One regional trainer dropped out because of illness and was replaced before the training programme began and another left during the project leaving a final total of ten.

Training occurred after randomisation. Schools randomised to one of the programme arms were provided with 1.5 days of CPD delivered in a venue near their school. They also received two scheduled visits (in Weeks 2 to 4 and Week 10) from their regional trainers to support the programme delivery, and just-in-time support delivered on demand, either remotely or in person.

Schools were encouraged to use the RUKS programme during the course of one academic school year (2018/2019) for a minimum of 20 weeks but were able to continue beyond 20 weeks if they wished. Based on the evidence from the previous efficacy trial (McNally et al., 2016), schools were instructed to group pupils in Year 1 into small groups of three to four pupils. Schools were asked to deliver the RUKS programme in four 15-minute sessions per week, supported by a member of school staff. This member of staff was chosen by the school.

Both versions of the RUKS programme focus on teaching Year 1 pupils, in small groups, a balanced reading curriculum to improve attainment. Essential elements of both programmes are:

- a systematic and structured approach to reading;
- evidence-based skills learned through a variety of activities linked to real books;
- small group work and discussion;
- development of a rich language environment; and
- valuing and using pupils' reflections.

The key difference between the programmes is the use of technology in the ICT delivery model and paper-based materials in the paper-based delivery model.

Both programmes were intended to be supplemental to usual literacy teaching, rather than acting as a replacement, and schools were advised to deliver the programme outside core literacy teaching time but ideally during literacy-based teaching, for example, during topic work time. Schools were expected to deliver a minimum of four 15-minute sessions per week for at least 20 weeks. How schools specifically used each of the programmes was explored within the IPE. Schools were also requested to encourage pupils to use their respective programme materials, that is, either the digital system or the paper-based resources, outside the specific group sessions, for example, at home or during lunch/breaks, however, this was not measured in any way as part of this evaluation.

In contrast to the efficacy trial (McNally et al., 2016), some key differences were instigated for the present trial for both the ICT and paper-based delivery models. These are presented in full, including the rationales behind each change, in Appendix C but include:

- Training was suitable for, and was delivered to, a range of nominated school staff, unlike in the efficacy trial where training was restricted to teaching assistants.
- Programme schools had to pay £200 to cover the costs of training rather than receiving it free of charge.
- Randomisation was between- rather than within-school
- An extra story book was added to the programme.
- Formalised progression rules using record sheets were added.
- The activity logs and registers were updated to capture usage outside the core usage (four 15-minute sessions a week for twenty weeks).
- Four types of programme extension were offered: (1) extended sessions of 20 to 30 minutes, (2) extended programme duration to 24 or 30 weeks, (3) the introduction of related activities (for example, 'grow a beansprout when working on the book about this'), and (4) give extra tuition to those struggling.

ICT delivery model (ABRA)

Both versions of the programme are intended to be delivered over a 20-week period with four 15-minute sessions per week. The content of the ICT version of the programme was delivered online using the ABRACADABRA (ABRA) web-based technology platform, which represents a modular 'game'-based literacy intervention taking a balanced approach to decoding and comprehension. The games are related and are linked, based on content, to a series of texts (typically stories) of an age-appropriate level. There are extension activities for some of the tasks for those children who progress more quickly through the programme. Full details of the ICT delivery model (ABRA) programme are provided in the efficacy trial **final report** (**McNally et al., 2016**). In addition to the changes made to both the ICT and paper-based delivery models outlined earlier, the following changes were also made to the ICT delivery model for the present research. These are presented in full in Appendix C but include:

- It was stipulated that the 20-week ICT programme be delivered on a laptop or desktop computer connected to the internet, but not via tablets.
- Nominated school staff were provided with ICT facilities during training to access and practice with the software.
- Nominated school staff were provided with 'best practice guidelines' sheets.
- ABRA has been updated from a Flash-based version to an HTML 5 version. The screenshots in the McNally et al. (2016) EEF report may differ slightly from the current version of ABRA
- Schools were previously provided with a laptop with the ABRA software preloaded, whereas in the present trial schools had to use their own devices to access the software online.

A full overview according to the TIDieR checklist (Hoffman et al., 2014) is presented in Appendix E.

Schools allocated to the ICT delivery model arm were required to pay a subsidised rate of £200 for the training of school staff.

Paper-based delivery model

The paper-based version of the programme uses all the same stories, vocabulary items, questions, words, and letter sounds in all of the activities. It is essentially identical to the ICT version, just delivered using a different medium. The only real difference is the use of paper-based activities and additional resources such as magnetic letters and cards. Full details of the paper-based delivery model are provided in the efficacy trial **final report** (McNally et al., 2016). In addition to the changes made to both the ICT and paper-based delivery models outlined above, the following change was also made to the paper-based delivery model. Changes are presented in full in Appendix C:

- The paper-based version is a paper version (not pencil and paper).

A full overview according to the TIDieR checklist is presented in Appendix E.

As per the ICT delivery model, schools allocated to the paper-based delivery model were required to pay a subsided rate of £200 for the training of school staff.

Control

Schools in the control group were asked to continue with usual teaching with Year 1 pupils in the 2018/2019 academic year, taking a 'business as usual' approach. It was anticipated that, in most schools, some pupils would receive small group teaching as part of usual practice (for example, for pupils performing below the expected level or needing additional support/stretch). In order to explore the effects of small group teaching *per se*, if their usual teaching included some form of small group teaching, schools in the control group were asked to deliver such teaching—similar in length and delivery to that which would be delivered in the programme schools—to approximately four pupils, if their usual teaching included some form of small group teaching. The content of this additional small group teaching was at the schools' discretion, but schools were advised that this should be literacy-based but could not be ABRA or the paper-based equivalent. Prior to randomisation, all schools were asked to pre-identify and inform the evaluation team of the details for the (approximately) four pupils they intended to deliver small group teaching to if their school was allocated

to the control group and if their usual teaching included some form of small group teaching. We collected details of any small group teaching in terms of its content and delivery through surveys and interviews via the IPE. Schools in the control arm received a thank you payment of £500 for their participation in the trial.

Evaluation objectives

As stated in the **trial protocol** (Ainsworth et al., 2018) and **statistical analysis plan** (Fairhurst and Mitchell, 2019), the primary objective of this analysis was:

- To investigate the effectiveness of an ICT-based small-group reading support programme (ABRA) delivered at scale and an equivalent paper-based small group reading support programme delivered at scale to pupils in Year 1 (ages five to six) on literacy development.

The specific research questions were:

Primary research questions

1. How effective is the ICT delivery model of the reading support programme (ABRA), compared to the 'business as usual' group, in increasing the literacy skills of pupils in Year 1?
2. How effective is the paper-based delivery model of the reading support programme, compared to the 'business as usual' group, in increasing the literacy skills of pupils in Year 1?

Secondary research questions

3. How effective is the ICT delivery model of the reading support programme (ABRA), compared to the 'business as usual' group, in increasing the literacy skills of pupils in Year 1 who are eligible for FSM?
4. How effective is the paper-based delivery model of the reading support programme, compared to the 'business as usual' group, in increasing the literacy skills of pupils in Year 1 who are eligible for FSM?
5. How effective is the ICT delivery model of the reading support programme (ABRA), compared to the paper-based model, in increasing the literacy skills of pupils in Year 1?
6. How effective is the ICT delivery model of the reading support programme (ABRA), compared to the paper-based model, in increasing the literacy skills of pupils in Year 1 who are eligible for FSM?

The research questions for the implementation and process evaluation (IPE) were:

1. What does baseline practice in participating schools look like (control and programme) in terms of teaching or interventions targeted at improving literacy in Year 1?
2. To what extent do the schools and teachers implementing the programmes adhere to the intended model and dosage?
    a) Linked to the above, how effectively has the training provided to the trainers been cascaded to the school staff?
    b) What variability in implementation exists across different participating settings? Are there any barriers or adaptations?
    c) How well have components of the programmes been delivered and how well have pupils and school staff engaged with it?
    d) What is the reach of the programmes across the programme schools—for example, what proportion of training has been attended? How many sessions have been completed?
3. During the implementation period, what other practices do evaluation schools use that focus on improving literacy at Key Stage 1?
    a) What small group interventions have been used in control schools and/or did any compensatory activities occur?
    b) Have programme schools used any other literacy-focused interventions or small group teaching in addition to the programmes?
    c) Have the programmes been delivered in addition to usual literacy teaching, or to what extent have the programmes been substituted for usual literacy teaching?
4. Are there perceived or actual benefits for specific groups of pupils (for example, SEN, EAL, GRT)?

5. How did schools group pupils for small group work (programme and control schools)?
    a) Does the way pupils are 'grouped' have any impact of the effectiveness of the programmes (ICT and paper-based delivery models)? NOTE: This was explored through observational data only.
6. What are the views of specified stakeholders (teachers, pupils, developers) about the implementation and effectiveness of the programmes during the trial period?

Ethics and trial registration

Ethical approval was obtained from Durham University's School of Education Ethics Committee in March 2018. The Health Sciences Research Governance Committee at the University of York was informed of the outcome of this ethical review as per University procedures. Ethical approval was also obtained from Coventry University on behalf of the implementation team (October 2018).

Data sharing agreements were in place with participating schools.

Trial monitoring was undertaken by a trial management group, which was responsible for the day-to-day running and management of the trial. Led by the principal investigator (KB, and HA during KB maternity leave), the trial management group comprised all members of the evaluation team. The team met regularly throughout the trial and further included representatives from the implementation team and the EEF, where appropriate.

Trial registration number: ISRCTN37208856.

Data protection

All data was stored and processed in accordance with the Data Protection Act (2018).

Schools were informed of the data requirements at recruitment meetings and through the Memorandum of Understanding (MOU; Appendix F). After agreeing to take part in the evaluation, all schools sent letters (Appendix G) to the parents or carers of the pupils in the participating class(es) containing information sheets that outlined the data schools were providing about the pupils in the trial and how it would be used. Parents and carers were given an option to withdraw their child from data sharing before the baseline data collection stage. Because the programme was delivered at the class level in many schools, parents were not given an option to withdraw their child from receiving the programme.

At the start of the trial, schools provided pupil details including full names, unique pupil numbers (UPN), and date of birth to the evaluation team for all pupils in the participating class(es) to allow linkage with National Pupil Database (NPD) data. This data linkage included sensitive data such as FSM (free school meal) status. Access to pupil details was limited to specific members of the evaluation team. During the trial, the University of York was deemed to be a data controller (as defined by the data protection legislation). The data used for the statistical analysis will be archived in the EEF data archive, as managed by FFT Education, within the Office for National Statistics Secure Research Service. The EEF should be recognised as the data controller for all archived datasets following trial completion.

All personal data was processed under Article 6 (1) (e) ('processing necessary for the performance of a task carried out in the public interest') and Special Category data under Article 9 (2) (j) ('processing necessary for ... scientific ... research purposes') of the General Data Protection Regulation (2016, applicable in the U.K. from May 2018). All data transfers to and from schools were via encrypted spreadsheet sent through the University of York's Dropoff service (a secure online file transfer service). All pupil outcome and assessment data, from baseline through to final follow-up, was collected on paper and identified by each pupil's unique trial ID. Paper assessment data is held securely in a controlled access area in locked cabinets within York Trials Unit (YTU) at the University of York. The trial management systems and trial data are held on secure University of York servers with access limited to specified members of YTU staff. Electronic data and paper documents including identifiable personal data will be securely archived and disposed of by YTU when the youngest participating young person is aged 25. This is in line with the Limitations Act 1980 and the Records Management Code of Practice for Health and Social Care 2016. Anonymised electronic data and paper documents will be kept indefinitely. Data sharing agreements were put in place between the participating institutions (evaluation team and implementation team). Data sharing agreements (Appendix H) were also in place with participating schools.

All results have been anonymised so that no school or individual pupil will be identifiable in the report or dissemination of any results.

Project team

The independent evaluation was conducted by researchers from University of York and Durham University.

University of York

Dr Kerry Bell, Principal Investigator (PI).

Hannah Ainsworth, Interim PI providing maternity cover (October 2017 to November 2018).

Louise Elliott, Data Manager, Testing Co-ordinator and Co-Investigator.

Caroline Fairhurst, Statistician and Co-Investigator.

Dr Pam Hanley, Lead Process Evaluator.

Andrew Haynes, Trial Support Officer.

Professor Catherine Hewitt, Senior Statistician and Co-investigator.

Alex Mitchell, Statistician.

Professor Carole Torgerson, Methodologist, Educationalist and Co-Investigator.

Professor David Torgerson, Methodologist and Co-Investigator.

Dr Katie Whiteside, Trial Co-ordinator

Durham University

Dr Louise Gascoine, process evaluator and co-investigator.

Members of the evaluation team involved in the process evaluation were Kerry Bell, Pam Hanley, Louise Gascoine, Katie Whiteside, and Andrew Haynes.

The programme was developed and facilitated by a team of researchers led by Dr Janet Vousden (formerly Coventry University, now Nottingham Trent University). The full implementation team comprised:

Dr Janet Vousden, Coventry University/Nottingham Trent University.

Professor Clare Wood, Nottingham Trent University.

Dr Helen Johnson, Coventry University.

Professor Rob Savage, UCL.

Professor Phil Abrami, Concordia University, Canada.

Professor Anne Wade, Concordia University, Canada.

Dr Luisa Tarczynski-Bowles, Coventry University.

Dr Anna Cunningham, Nottingham Trent University.

Ms Joanne Lloyd, Coventry University.

The ABRA evaluation was supported at the EEF by:

Anneka Dawson (until Dec 2017), EEF Evaluation Manager.
Dr Florentina Taylor (until December 2020), EEF Evaluation Manager.

Katharina Keck (until November 2021), EEF Senior Evaluation Manager.

Dr Rachael Morris, EEF Evaluation Manager.

Thomas Martell (Until Jan 2019), EEF Programme Manager.

Matthew van Poortvliet (Until July 2020), EEF Programme Manager.

Eleanor Stringer: EEF Head of Programmes (until June 2021).

Sarah Tillotson (Until March 2022), EEF Senior Programme Manager.

Aoife Duff, EEF Programme Manager.

Methods

Trial design

The evaluation was a three-arm cluster RCT with randomisation at the school level. Participants were Year 1 children.

Table 3: Trial design

Trial design, including number of arms		Three-arm cluster RCT
Unit of randomisation		School-level, via minimisation
Minimisation variables		7. Deliverer staff type: three levels—qualified; non-qualified; both 8. Number of pupils in the Year 1 cohort: two levels—: [cut point defined using median number of pupils per Year 1 cohort] ≤38; >38) 9. Percentage of pupils ever eligible for FSM in the Year 1 cohort (two levels—: [cut point defined using median percentage of pupils ever eligible for FSM in the Year 1 cohort] ≤21%; >21%) 10. Geographical area: five levels—West Midlands; East Midlands; Newcastle; Teesside; Manchester
Primary outcome	Variable	Reading attainment
	Measure (instrument, scale, source)	Progress in Reading Assessment (PiRA) Summer 1 Version • Scale: 70–130 • Source: in-person test
Secondary outcome(s)	Variable(s)	• Ability to read exception, regular and nonwords • Ability to sound out single letters and letter combinations • Reading attitudes

	Measure(s) (instrument, scale, source)	• Ability to read exception, regular and nonwords
		  ○ instrument: Diagnostic Test of Word Reading processes (DTWRP)
		  ○ scale: age-standardised score on a scale of 70–130
		  ○ Source: in-person test
		• Ability to sound out single letters and letter combinations
		  ○ instrument: total LeST
		  ○ scale: normalised z-score on a scale of -3.09–2.37
		  ○ source: in-person test
		• Reading attitude
		  ○ instrument: Reading Attitudes Questionnaire
		  ○ scale: total raw score on a scale of 23–92
		Source: in-person test

Baseline for primary outcome	Variable	Reading attainment

	Measure (instrument, scale, source)	• Progress in Reading Assessment (PiRA) Autumn 1 Version
		  ○ scale: 70–130
		  ○ source: in-person test

Baseline for secondary outcome(s)	Variable	Reading attainment
	Measure (instrument, scale, source)	• Progress in Reading Assessment (PiRA) Autumn 1 Version ○ scale: 70–130

		○ source: in-person test

A three-armed, pragmatic ('real world'), cluster RCT was conducted to explore the effectiveness of the RUKS programme (ICT or paper-based) for improving children's reading skills in Year 1. A pragmatic design was selected to reflect as accurately as possible the implementation of the programme in a real-life context. Most schools (those with ICT facilities) were randomly allocated to one of the three trial arms, ICT delivery, paper-based deliver, or business as usual, in the academic year 2018/2019. Due to the ICT delivery arm relying on access to a degree of technology, eight schools without ICT facilities were only randomised between the paper-based and business as usual arms to avoid excluding schools based on their IT equipment.

The programme is a non-targeted, full class programme. Initially, it was anticipated that smaller schools (single class intake) would deliver the programme to the whole Year 1 cohort whereas larger schools, with greater than two Year 1 classes, would deliver to a maximum of two classes. Schools were asked at recruitment to indicate how many classes and pupils they intended to deliver the programme to if allocated to one of the programme arms. As explained in the Participant Selection section below, this information determined the selection of classes or pupils to take part in the evaluation within each school. However, following the training of school staff, several schools felt that they did not have capacity to deliver the programme to the number of classes or pupils they had initially intended and opted to deliver the programme to a smaller number of pupils. Where the evaluation team were informed of this, we were able to randomly select a smaller subset of their original cohort to receive the programme, according to the number they felt they could manage (that is, randomly 'deselect' a number of pupils initially selected to receive the programme). However, at various stages throughout the implementation period, we were made aware of several other schools that had reduced the number of pupils receiving the programme due to capacity issues. This will be described in full detail below. All pupils regardless of whether they were deselected from the programme, were followed up for post-testing.

While there was no cost to schools for the programme materials, all schools allocated to deliver the ICT or paper-based programme were asked to contribute £200 to cover the cost of training school staff. All other costs, including the remainder of the training cost, were subsidised by the EEF.

We did not adopt a waitlist design; rather, schools allocated to continue business as usual (as described above) were offered £500 as an incentive to continue providing data for the evaluation. Prior to randomisation, all schools were asked to identify a sample of four Year 1 pupils to whom they would deliver small group teaching if they were allocated to the control arm and if their usual teaching included some form of small group teaching. Schools were advised that the content of this small group teaching was at their discretion but must be literacy-based, cannot be ABRA or the paper-based equivalent, and should be similar in length and delivery to that which would be delivered in schools allocated to the use the RUKS programme (that is, for 20 weeks, four 15-minute sessions per week).

The trial was designed, conducted, and is reported to CONSORT standards (Altman et al., 2012) in order to minimise all potential threats to internal validity such as selection bias and a range of post-randomisation bias.

Participant selection

The recruitment process was led by Professor Clare Wood, Nottingham Trent University, from the implementation team. The evaluation team supported the implementation team in recruiting primary schools by way of preparing recruitment materials and jointly attending recruitment events to talk about the trial with prospective schools. The implementation team also received support from the National Literacy Trust and the EEF. In total, five recruitment hubs were established in the West Midlands, East Midlands, Newcastle, Teesside, and Manchester with the aim of recruiting approximately 40 schools in each hub. The recruitment hubs were chosen in order to target schools serving deprived communities with the aim of recruiting schools with above the national average proportion of FSM pupils (14.1% among primary schools in 2017; DfE, 2017). Given the challenging recruitment target in this trial, we did not set a minimum threshold for FSM pupils in a school as an inclusion criterion.

Recruitment activities included recruitment events, face to face visits to schools, telephone and email communication, and advertisement. Schools were invited to attend recruitment events during the academic year 2017/2018. These

events comprised an explanation of the reading support programme from the implementation team as well as a presentation providing an overview of the trial design and expectations for participating schools from the evaluation team. Schools were given the opportunity to raise any concerns or queries about either the programme or the evaluation. Schools that expressed an interest in participating in the trial were asked to sign a memorandum of understanding (MOU; see Appendix F). This outlined the expectations of the implementation team (Coventry and Nottingham Trent Universities), the evaluators (University of York, Durham University), and the participating schools. Participating schools were asked to complete a baseline survey (see Appendix I) and conduct a baseline assessment (Progress in Reading Assessment, PiRA, Autumn 1, hereafter referred to as the 'pre-test') prior to randomisation. Schools were also asked to provide the evaluation team with pupil-level data necessary to identify children to complete the pre- and post-testing and to enable matching with the NPD.

For schools to be eligible to take part in the evaluation and to receive the programme, they were required to have:

- a Year 1 cohort;
- the necessary ICT equipment to take part in the ICT arm (access to a suitable computer in a suitable location): these schools were identified as eligible to be randomised to any of the three arms; schools without the necessary ICT equipment to take part in the ICT arm were identified as eligible to be randomised to either of the other two arms—paper-based or control;
- Schools willing to agree to the requirements of participation outlined in the MOU (Appendix F)
- Schools who could feasibly deliver programmes to a minimum of ten Year 1 pupils.

To ensure the capacity and willingness to participate in the trial, schools were asked to sign a comprehensive MOU which fully outlined the requirements for participating schools (Appendix F).

Alongside the MOU, schools were asked to provide information about:

- how many classes and/or pupils they intend to deliver the programme to;
- the type of staff members they will send to the training;
- the school's ICT resources;
- the total number of pupils in the Year 1 cohort;
- the percentage of pupils ever eligible for FSM across the whole school cohort; and
- the percentage of pupils ever eligible for FSM in the Year 1 cohort.

In addition to the MOU signed by the school's headteacher, opt-out consent was also sought from parents/carers regarding their child participating in the assessments for the evaluation and their child's data being shared for the purposes of the evaluation. At the start of the 2018/2019 school year, schools were asked to distribute an information sheet provided by the evaluation team to inform parents/carers of the study (Appendix G). Parents/carers could then withdraw their child from the evaluation before any data collection by returning a withdrawal from research form. After distributing the information sheet and allowing parents/carers sufficient time to respond, schools then securely sent the evaluation team a list of names, unique pupil numbers (UPNs) and baseline data for all pupils in Year 1 (2018/2019) whose parents/carers had not returned a withdrawal from research form.

The number of classes or pupils that each school advised they intended to deliver the programme to, if allocated to a programme arm, determined the number classes/pupils selected to take part in the evaluation and complete the baseline assessment within each school. Where possible, the evaluation team randomly selected classes/pupils (from those whose parents/carers had not returned a withdrawal from research form) using the following principles:

- For schools that intended to deliver the programme to one or more classes, one class was randomly selected to be assessed. Other classes were randomly selected according to the school's capacity to deliver the programme and assessment was optional for these classes (schools were asked to assess these classes if they had time and capacity to do so).

- For schools that intended to deliver the programme to a specified number of pupils, one class was randomly selected and then the specified number of pupils were randomly selected from within that class. If applicable, other classes were randomly ordered and pupils were randomly selected from these classes until the specified number of pupils that the school had capacity to deliver to was reached.

Nine schools requested that they select the classes/pupils to take part themselves, often for practical/logistical reasons, and this was permitted as a last resort to retain the schools in the trial. Since this occurred prior to randomisation, it was assumed that this would be balanced across the three arms so should not introduce selection bias. Of these nine schools, two were allocated to ICT delivery, two to paper-based delivery, and five to the control group

Written consent was required from parents/carers for any children selected to participate in focus groups with members of the evaluation team for the trial IPE. Parents/carers of children selected to participate in the focus groups received a letter from the evaluation team (via the school) providing information about the purpose of the focus group and a reminder of the overall aims of the study (Appendix R). Parents/carers were asked to return a consent form to the school if they were happy for their child to participate in the focus groups. The evaluation team also sought informed verbal consent from pupils before they participated in focus groups.

Outcome measures

Primary outcome

The primary outcome was measured (pre- and post-implementation) using the Progress in Reading Assessment (PiRA), which is available from Rising Stars Assessment.[2] There are three versions of the PiRA per year group, one for each school term: autumn, spring, and summer. The baseline (pre-test) assessment was administered in October 2018 so the PiRA Autumn 1 paper was used; the post-test was administered from May to July 2019 using the PiRA Summer 1 paper. The PiRA evaluates the general reading ability of pupils and is commonly used in schools to monitor reading progression over the school year. In particular, the PiRA assesses the following areas of reading ability: phonics, literal comprehension, and reading for meaning. This instrument strongly aligns with the overarching aim described in the logic model—to improve general reading ability.

At baseline, the PiRA was administered by class teachers but marked and scored by research assistants at the University of York. At post-test, the PiRA was administered by research assistants who were blind to the school allocation. The PiRA papers were then marked by a separate set of research assistants who were also blind to the school allocation. All research assistants were specifically employed for the respective tasks and were not members of the evaluation team. At both pre- and post-test, a sample of 10% of papers were double marked to ensure consistency of marking. Where discrepancies arose, questions were re-marked, and a final score confirmed. To the best of our knowledge, pre-testing was conducted under exam conditions as per our request to schools.

The primary outcome was the age-standardised PiRA score. This was obtained by calculating the raw PiRA score and using the conversion tables contained in the manual provided by Rising Stars Assessment to calculate the age-standardised score. It is standardised to an average score of 100, showing whether a pupil is above or below average relative to PiRA's national standardisation sample. A higher age-standardised score indicates greater attainment.

Secondary outcomes

The following secondary outcomes were measured at post-test only in a subset of up to ten pupils per school randomly selected from the pupils assessed for the primary outcome at pre-test. All secondary outcomes were collected, administered, and marked by research assistants blind to school allocation. The secondary outcomes were as follows.

Diagnostic Test of Word Reading Processes (DTWRP)

[2] More information on the PiRA can be found at **https://www.risingstars-uk.com/pira**

Diagnostic Test of Word Reading Processes (DTWRP)[3] assesses the reading of regular words, exception words, and non-words to enable the precise areas of difficulty experienced by individual pupils to be identified. To achieve higher scores on this assessment children must apply decoding and comprehension skills, which are core components of the logic model. The DTWRP takes approximately ten minutes for a pupil to complete and is delivered on a one to one basis. This test is more sensitive than a group reading test as it focuses on the decoding of single words (which is a focus of KS1), thus it can identify children who are at floor on PiRA because their decoding is poor.

The age-standardised DTWRP score was analysed as the secondary outcome. This was obtained by calculating the raw DTWRP score and using the conversion tables provided in the user manual to calculate the age-standardised score. A higher age-standardised score indicates greater attainment.

Letter Sound Test (LeST)

The Letter Sound Test (LeST)[4] assesses a person's ability to sound out single letters and letter combinations. To achieve higher scores on this assessment children must apply decoding skills, which is a core component of the logic model. The LeST takes approximately five minutes for a pupil to complete and is delivered on a one to one basis. This test also focuses on letter sounds (which is a focus of KS1) so is much more sensitive than a group reading test. The number of correct items, out of 51, are summed to produce a total raw score. This total raw score was then converted to an age (year group) standardised 'z-score' for 'Year 1' (ages five to six) that was used for analysis. A higher age-standardised score indicates greater attainment.

Reading Attitudes Questionnaire (RAQ)

The logic model hypothesises that exposure to the programme will result in improved attitudes towards reading and an increased motivation to read. The Reading Attitudes Questionnaire (RAQ) assesses a child's attitude towards and motivation in reading. The RAQ takes approximately five minutes for a pupil to complete and is delivered on a one to one basis. The aim of this secondary outcome was to measure a more process-based outcome and potentially a marker of more distal effects—since it is known that there is a positive relationship between motivation and reading. It was also felt that schools would be interested in this measure. (Note that to reduce the burden on schools, and to keep testing costs down, if all independent testing could not be completed in one day, we accepted some attrition on this measure.) The RAQ is scored by calculating the total raw score. A higher score indicates more positive attitudes and greater motivation towards reading.

Key Stage 1 attainment

As per the trial protocol, we had hoped to explore the impact of the reading support programme on KS1 assessments completed at the end of 2019/2020, however, this was not possible due to the Covid-19 pandemic, which resulted in a nationwide lockdown commencing 23 March 2020 and the cancellation of KS1 assessments.

Data collection procedures

All secondary outcomes were administered on a one to one basis to a sub-sample of pupils per school with the view of assessing ten pupils per school. The DTWRP and LeST were scored by the research assistants at the time of testing. Scores were then entered into a database by research assistants at the University of York. Data entry was double entered for a sample of 10% to ensure a low data entry error rate.

All assessments were completed on paper and research assistants adhered to the guidelines provided by each of the respective instrument suppliers. To ensure consistency of administration at post-test across the pool of 21 research assistants administering the assessments, one of the two members of the evaluation team responsible for training the research assistants (KB and LE) accompanied each research assistant on their first visit to a school to observe the test administration. Any issues arising were discussed after observation and additional training provided where indicated. Research assistants all attended a one-day training course delivered by members of the evaluation team (KB and LE)

[3] More information on the DTWRP can be found at **https://www.gl-assessment.co.uk/products/diagnostic-test-of-word-reading-processes/**
[4] More information on the LeST can be found at **https://www.motif.org.au/home/test/7**

prior to commencing testing. Research assistants were also provided with a training manual and administration script developed by members of the evaluation team (KB and LE). Following observations of all 21 research assistants, a training update was distributed to address any issues observed (all of which were minor).

Sample size

Table 4: Sample size details

		Protocol	
		Overall	FSM
MDES (Cohen's d)		0.20	0.23
Pre-test/post-test correlations	Level 1 (pupil)	0.45	0.45
	Level 2 (class)	0	0
	Level 3 (school)	0	0
Intracluster correlations (ICCs)	Level 2 (class)	0	0
	Level 3 (school)	0.15	0.15
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		two-sided	two-sided
Average cluster size		27	7
Number of schools[5]	ICT	67	67
	Paper-based	67	67
	Control	67	67
	Total:	201	201
Number of pupils[6]	ICT	1,809	452

	Paper-based	1,809	452
	Control	1,809	452
	Total	5,427	1,356

Protocol

The previous efficacy RCT (McNally et al., 2016) found an effect size of 0.138 for the ICT programme and 0.231 for the paper-based programme, with larger effect sizes among FSM pupils eligible for FSM (0.368 and 0.396, respectively). A total of 84% of pupils involved at randomisation were included in the primary analysis, with an average of 40 pupils per school. The intracluster correlation coefficient at analysis was 0.15 and the correlation between the pre-test and the post-test was 0.43. (Note, this correlation is not the raw correlation between the PiRA pre-test and post-test—it accounts for covariates and is taken from the R-squared of a regression.) The previous evaluation mandated that participating schools have a minimum of two teaching assistants and therefore the included schools are likely to be larger than the average U.K. primary school. Nationally, there were 27.1 pupils in the average primary school class in 2016.

This was a three-arm trial, with two primary research questions relating to the comparisons of the two RUKS programme arms against the shared control arm. In such a scenario, there is no consensus on whether adjustment for multiple testing is required (Wason et al., 2014). In a discussion during the initial set-up meeting of the research project attended by the evaluation team, implementation team, and the EEF, the decision was made not to apply a statistical correction for the fact that we have two primary research hypotheses; therefore, both comparisons were assessed at the 5% significance level.

We planned to recruit a sample of 201 schools (67 in each arm) to give 80% power to detect an effect size of approximately 0.20 of a standard deviation (SD) between either of the programme groups with the control group, assuming an average class size of 27, 15% attrition at the pupil-level at follow-up, an ICC of 0.15, alpha of 0.05, and a correlation of 0.45 between the pre-test and post-test. For the secondary outcomes, with ten pupils per school under otherwise identical assumptions (but assuming no attrition, that is, actually following up ten per school), the MDES was approximately 0.22.

Across all primary schools in England, in January 2016, the average percentage of children claiming FSM was 14.5% (DfE, 2016). In this trial, we aimed to recruit schools in deprived areas likely to have higher than average levels of pupils eligible for FSM. We assumed an average percentage of 25% in each school, as this was the average observed in schools randomised into a recent EEF trial (evaluation of 'ReflectED', still ongoing, unpublished). With an average of 27 pupils per school at randomisation, we expected an average of seven pupils to have FSM status (201 x 27 x 0.25 = 1,356 in total). With this number, assuming 15% pupil-level attrition at follow-up, an ICC of 0.15, alpha of 0.05, and a correlation between the pre-test and post-test of 0.45, we had 80% power to detect an effect size of 0.23 in the FSM subgroup in the primary analysis.

The parameters above used to calculate the MDES for all pupils and the FSM subgroup were based on assumptions made at the protocol stage. The observed parameters and calculated MDES at the randomisation and analysis stages are presented in Table 10. The MDES was calculated using Stata code written by the study statistician.

Randomisation

A York Trials Unit statistician, not involved in the recruitment of schools, used a dedicated computer program, MinimPy (Saghaei, 2011) to randomise schools 1:1:1 to either of the two programme groups or to the control group. Minimisation was undertaken to ensure the groups were balanced on the following factors:

- staff type—the member/s of staff identified by the school who would deliver the programme (three levels: qualified teacher or non-qualified teacher or both);

- number of pupils in the Year 1 cohort (two levels; median number of pupils per Year 1 cohort was used as cut point);
- percentage of pupils ever eligible for FSM in the Year 1 cohort (two levels; median percentage of pupils who have ever been eligible for FSM was used as a cut point); and
- geographical area—five levels: West Midlands, East Midlands, Newcastle, Teesside, and Manchester.

At the point at which schools were ready to be randomised, data from all engaged schools was used to calculate the median cut-points for the number of pupils in the Year 1 cohort and percentage of FSM pupils, for use in minimisation. These values were calculated as n = 38, and 21%, respectively. Schools were randomised on a rolling basis as and when they completed baseline tasks; however, in order to randomise schools in time for them to attend the relevant training, 33 schools had to be randomised and informed of their trial allocation before completing pre-tests. This is not ideal and has the potential to introduce bias. The impact of this is explored in a sensitivity analysis.

One of the programme arms relied on technology. In general, some schools, particularly small or rural schools, have recurrent problems with technology. We did not want to exclude schools from participation based on ICT facilities as this did not pose a barrier to usage of the equivalent paper based programme in usual practice. Schools that identified potential ICT limitations were randomised instead on a 1:1 basis between the paper-based delivery approach and business as usual only, but otherwise as described above.

Statistical analysis

Analysis was conducted in Stata v17 using the principles of intention to treat (ITT), where data was available, including the schools and pupils in the groups to which they were randomised irrespective of whether or not they actually received the RUKS programme. A detailed statistical analysis plan was published online.

Statistical significance was assessed using two-sided tests at the 5% level. Estimates of effect with 95% confidence intervals (CIs) and p-values are provided.

School and pupil characteristics and measures of prior attainment are summarised descriptively by randomised group both as randomised and as analysed in the primary analysis models. No formal statistical comparisons were undertaken (Senn, 1994) except to report the differences in PiRA pre-test scores (raw and age-standardised) for those analysed as a Hedges' g effect size with 95% CI (Hedges, 2007). Continuous measures are reported as a mean and standard deviation (SD), while categorical data is reported as a count and percentage.

The trial was designed and conducted, and has been reported, to CONSORT standards (Altman et al., 2012; Wears et al., 2012; Moher et al., 2010). A CONSORT diagram (Figure 2) has been provided as part of the impact evaluation results to show the flow of schools and pupils through the trial.

Primary analysis

The raw and age-standardised PiRA post-test scores are summarised by trial arm, including the number and percentage above the national average (100 for the age-standardised score relative to PiRA's national standardisation sample). The age-standardised scores were calculated based on the child's age in years and completed months (one month brackets). Lookup tables for calculating the age-standardised score are provided in the user manual. No further use was made of raw scores and all proceeding analyses relate to the age-standardised score. The advantages of using the age-standardised score rather than the raw score include:

- it is standardised to an average score of 100, immediately showing whether a pupil is above or below average relative to PiRA's national standardisation sample; and
- it allows comparisons to take into account the pupils' ages: within a given Year 1 cohort, the difference in the age of two children can be up to one year; older pupils in the year may have a higher raw scores than younger pupils but could have a lower age-standardised score.

Therefore, the age-standardised scores were used for analysis.

The correlation between the pre- and post-test scores was estimated. Histograms of the pre- and post-test data distributions are displayed in Appendix D.

Two multilevel, mixed-effect, linear regression models at the pupil level were used to compare post-test PiRA age-standardised score between the groups. The first model included pupils from all three groups except those from the eight schools that were only randomised between the paper-based and control groups (because they did not have the technology to implement the ICT RUKS programme). This model was used to obtain the pairwise comparisons between the ICT programme and control groups and between the ICT programme and the paper-based programme. This model will, on occasion, be referred to as 'model 1' in this report. The second model ('model 2') excluded pupils in the schools randomised to the ICT group and was used to investigate the difference between the paper-based and control groups; that is, this model included schools with ICT facilities randomised to paper-based or control plus schools without ICT facilities who could only be randomised to paper-based or control.

Two models were necessary as it was not statistically appropriate to include schools that could never have been allocated to receive the ICT programme in a comparison involving this group. The strength of the statistical comparisons made in a randomised controlled trial is based on the fact that they are made between groups that are formed through randomisation, which produces groups that are, on average, balanced across all known and unknown confounders. The schools without adequate ICT facilities to deliver the ICT programme were only randomised between the paper-based and control groups; therefore, including them in a comparison against the ICT group breaks the randomisation. Even if the three groups appeared well matched on measured factors, there are likely to be differences between schools with and without adequate ICT facilities, which would bias the comparisons. Adjusted mean differences in scores between pairs of groups in the form of Hedges' g were extracted from the relevant model with a 95% CI and p-value. Both models were adjusted as follows.

Pupil-level fixed effects:

- baseline age-standardised PiRA score;
- gender (male/female);
- FSM (NPD variable EVERFSM_6_P); and
- Foundation Stage Profile—EYFSP; NPD variable FSP_GLD, defined as whether or not the pupil achieved a good level of development (GLD), that is, achieved level of 2 or 3 in each of the following areas of learning and development: communication and language, physical development, personal, social, and emotional development, literacy, and mathematics.

School-level fixed effects:

- allocation—two or three levels, according to model; ICT, paper-based, or control;
- staff type—three levels; qualified, non-qualified, both;
- number of pupils in the Year 1 cohort, as a continuous variable; and
- geographical area—five levels: West Midlands, East Midlands, Newcastle, Teesside, and Manchester.

Adjustment was made for clustering at the school level by including school as a random effect, a standard method for the analysis of cluster trials (Wears, 2002).

Model assumptions were checked as follows: the normality of the standardised residuals were checked using a histogram and qq-plot and the homoscedasticity of the residuals was assessed using a scatter plot of fitted values against the residuals. Only visual inspections of the plots were used. If the model assumptions were in doubt, a sensitivity analysis was planned in which transformations of the outcome or covariate data would be tried to improve the model fit.

Secondary analyses

Sensitivity analysis of primary outcome to results of pupils who completed the pre-test after their school was informed of their trial allocation

For pupils who completed the pre-test after their school was informed of their trial allocation, the PiRA pre-test results were summarised descriptively alongside the results of pupils who completed the pre-test before their school was informed of their trial allocation.

The primary analysis models were repeated excluding pupils who completed their pre-test after their school was informed of their trial allocation.

Sensitivity analysis of primary outcome analysis to replace pre-test PiRA covariate with EYFSP result in G09 for literacy (reading) score

The primary analysis models were repeated replacing the pre-test PiRA score as a measure of prior attainment with the EYFSP result in G09 for literacy (reading) score (NPD variable FSP_LIT_G09). This was done in order to take advantage of an increase in power should the post-test PiRA age-standardised score be more strongly correlated with the EYFSP G09 than the pre-test PiRA age-standardised score. We also wished to explore whether routinely collected data could serve as a pre-test for similar future trials.

Selection of pupils to receive programme

The primary analysis models were repeated excluding pupils who were pre-tested but then randomly deselected by YTU to receive the programme immediately following their school attending training.

Small group teaching

To investigate the effects of small group teaching, it was planned for the primary analysis models to be repeated in the subset of pupils who were identified by their schools at baseline as those to be taught in a small group if their school was allocated to teaching as usual. However, due to a low return of data on small group teaching, it was not possible to carry out this analysis.

Analysis in the presence of non-compliance

Schools were asked to keep registers to indicate when and which pupils took part in programme sessions in the ICT and paper-based groups. Data from the registers was entered by the implementation team into Excel spreadsheets and sent to the evaluation team via the University of York DropOff facility (a secure transfer service). Data on the number of sessions completed by each pupil was summarised descriptively.

A Complier Average Causal Effect (CACE) analysis for the primary outcome was planned to assess the effect of the RUKS programme in the compliers. Compliance was defined in the SAP at the pupil level in three ways as follows:

- minimal compliance—completed at least four sessions of the programme (Y/N);
- full compliance—completed 80% (n = 64) of the planned 80 sessions (Y/N); and
- number of sessions completed (continuous variable).

A two-stage least squares instrumental variable (IV) approach was planned using random group allocation as the IV (Dunn et al., 2005). However, due to a low return of registry data, the CACE analysis was not carried out.

Missing data analysis

The amount of missing baseline and outcome data was summarised and reasons for missing data explored and provided, where available. A multilevel mixed-effect logistic regression model was run to assess for statistically significant predictors of missing primary outcome data at the pupil level, including all available pupil- and school-level baseline data as fixed effects and school as a random effect. Significant predictors and possible mechanisms for the missing data are discussed.

The impact of missing data on the primary analysis models was additionally assessed using multiple imputation by chained equations, predicted by pre-test PiRA age-standardised score, school, allocation, and any variables found to be significant in the 'drop-out' model described above. A 'burn–in' of 150 was used and 30 imputed datasets were created. The primary analysis models were then rerun within the imputed datasets and Rubin's rules were used to combine the multiply-imputed estimates.

FSM subgroup analyses

Pupil UPNs, as obtained during the recruitment period, were used to access additional pupil characteristics from the NPD (for example, FSM status). The effect of the RUKS programme on FSM pupils was assessed via the inclusion of FSM status (using the EverFSM indicator EVERFSM_6_P in the NPD) and an interaction term between FSM status and allocation in the primary analysis models. This was followed by repeating the primary analyses in the subgroup of pupils who have ever been eligible for FSM.

Post-hoc sensitivity analysis assessing the impact of pupils whose PiRA post-test was facilitated and marked unblinded on the primary analysis models

Two schools conducted the PiRA just before the post-test was due for their own purposes, rather than those for the research. These tests were thus facilitated and marked by the teachers and not by blinded research assistants as for all other schools. These two schools, nevertheless, sent the data to the evaluation team for use in the research. Data on the PiRA from these schools was not included in the primary analysis. However, the primary analysis models were repeated including the pupils whose PiRA post-test was facilitated and marked by a teacher from the pupil's school unblinded to allocation. This analysis was not pre-specified in the SAP as schools marking their own PiRA tests was an unexpected occurrence.

Secondary outcome analysis

The secondary outcomes of DTWRP, LeST, and RAQ were analysed exactly as described for the primary outcome of PiRA using the same fixed and random effects. As these were not assessed at baseline, the PiRA age-standardised score at pre-test was included as the measure of prior attainment in the models.

Sensitivity analysis of secondary outcomes excluding pre-test PiRA score covariate

The secondary outcome analysis models were repeated excluding the covariate for the pre-test PiRA age-standardised score. This sensitivity analysis was carried out in order to assess the impact of adjusting for the pre-test PiRA age-standardised score on the secondary outcome analyses.

Sensitivity analysis of secondary outcomes replacing pre-test PiRA score covariate with the EYFSP result in G09 for literacy (reading) score

The secondary outcome analysis models were repeated replacing the pre-test PiRA score as a measure of prior attainment with the EYFSP result in G09 for literacy (reading) score (NPD variable FSP_LIT_G09). This was done in case FSP_LIT_G09 was more strongly correlated with the secondary outcomes than the pre-test PiRA score.

Estimation of effect sizes

Effect sizes, in the form of Hedges' g, were calculated by dividing the adjusted mean difference between the RUKS programme and control group by the pooled variance obtained from the unconditional model. A 95% CI for the effect size was calculated by dividing the 95% confidence limits for the adjusted mean difference by the same standard deviation. The pooled variances used to calculate the effect sizes are presented in Appendix D.

Estimation of ICC

The school-level intracluster correlation coefficient (ICC) for the post-test outcomes was extracted from each multilevel analysis model, with the 95% CI. The ICC associated with school for the pre-test scores was also presented with a 95% CI.

Implementation and process evaluation

Research methods

In line with EEF guidance, the aim of the implementation and process evaluation (IPE) was to generate and analyse 'data to examine how an intervention is put into practice, how it operates to achieve its intended outcomes and the factors that influence these processes' (Humphrey et al., 2016). The IPE in this evaluation sought to explore and develop an understanding of the programme delivery (including its fidelity, quality, and identifying elements of successful delivery) in relation to the programme specification as described (prior to implementation) by the implementation team.

The specific research questions that the IPE aimed to answer were as follows:

- RQ1 What does baseline practice in participating schools look like (control and programme) in terms of teaching or interventions targeted at improving literacy in Year 1?
- RQ2 To what extent do the schools and teachers implementing the programmes adhere to the intended model and dosage?
  a. Linked to the above, how effectively has the training provided to the trainers been cascaded to the school staff?
  b. What variability in implementation exists across different participating settings? Are there any barriers or adaptations?
  c. How well have components of the programmes been delivered and how well have pupils and school staff engaged with it?
  d. What is the reach of the programmes across the programme schools (for example, what proportion of training has been attended? How many sessions have been completed)?
- RQ3 During the intervention period what other practices do evaluation schools use that focus on improving literacy at KS1?
  a. What small group interventions have been used in control schools and/or did any compensatory activities occur?
  b. Have programme schools used any other literacy focused interventions, or small group teaching, in addition to the programmes?
  c. Have the programmes been delivered in addition to usual literacy teaching, or to what extent have the programmes been substituted for usual literacy teaching?
- RQ4 Are there perceived or actual benefits for specific groups of pupils (for example, SEN, EAL, GRT)?
- RQ5 How did schools group pupils for small group work (programme and control schools)?
  a. Does the way pupils are 'grouped' have any impact on the effectiveness of the programmes (ICT and paper-based delivery models)? NOTE: This will be explored through observational data only.
- RQ6 What are the views of specified stakeholders (teachers, pupils, developers) about the implementation and effectiveness of the programmes during the trial period?

Method

The IPE took place over the duration of programme delivery (school year 2018/2019) using a mixed-methods approach. A combination of surveys, interviews, observations, and focus groups were used as appropriate to engage with a variety of stakeholders, principally pupils (programme schools only), teachers, teaching assistants, regional trainers, and the developers (see Table 5 to Table 7 for IPE methods overview).

The school visits were completed by a team of five researchers. Each researcher conducted their first one or two visits in a pair so that ratings and observations could be discussed and moderated where necessary. Subsequently, researchers made the visits on their own.

Sampling

As outlined in the trial protocol, 25 schools were visited for IPE data collection (see Tables 5-7 for IPE methods overview). The sampling strategy was purposive. For the programme schools, a range of characteristics were taken into account

including geographic and urban/rural location, cohort size, and proportion of FSM pupils as well as assigned trainer and school role of facilitator (teacher and/or TA). The control schools selected for visits were drawn from 18 schools that confirmed that they were planning to deliver a small group literacy intervention to selected pupils. The sample of 25 schools in the IPE were sufficiently representative of the overall trial sample in terms of year group size and proportion of FSM.

Fidelity

Fidelity was explored in relation to the description of each of the programme arms outlined in the trial protocol (Ainsworth et al., 2018). In brief, the programme was stated as comprising the following.

ICT delivery model (ABRA)

- Programme lessons for all children in Year 1 using ABRA software
- Lessons in groups of 3-4 pupils
- A minimum of four 15-minute lessons per week for 20 weeks over the course of the academic year

(see Appendix E for full details)

Paper-based delivery model

- Programme lessons for all children in Year 1 using paper-based reading support programme
- Lessons in groups of 3-4 pupils
- A minimum of four 15-minute lessons per week for 20 weeks over the course of the academic year

(See Appendix E for full details.)

Fidelity to the programme (how closely schools adhered to the programme as outlined above) was measured using a range of approaches:

- in-school observations of programme sessions (facilitators and pupils);
- in-school interviews with a mix of teachers and teaching assistants;
- in-school focus groups with pupils;
- online questionnaires completed by a mix of school staff (teachers, heads, and teaching assistants); and
- online questionnaires completed by regional trainers.

Stakeholder views

In addition to exploring fidelity, the IPE also sought to determine the efficacy (from the stakeholders' perspectives, defined as school staff and pupils) and feasibility of integration of the two programme delivery models into teaching practice alongside exploring what usual practices would be for supporting reading for Year 1 pupils.

Data collection and analysis

The IPE comprised three main stages:

- pre-implementation (before programme commencement)—a survey to establish baseline practice in all participating schools (control and programme) and observations to establish fidelity of the training from the train-the-trainers sessions to training the nominated in-school facilitators;
- during implementation of the programme—school visits to explore what the two programme delivery models looked like in practice within the participating programme schools, including observations of programme lessons as well as facilitator interviews and pupil focus groups; and
- post-implementation—teacher/facilitator questionnaires, a developer interview, and regional trainer questionnaires to establish stakeholder opinions at the end of the implementation period. (teacher/facilitator questionnaires, developer interview and regional trainer questionnaires)

Copies of all data collection tools can be found in Appendices I to O.

Pre-implementation stage

The pre-implementation stage involved observing the processes of training the regional trainers and training school staff and also identifying the usual (baseline) teaching practices for teaching Year 1 reading in all schools (control and programme). Methods and sampling strategies were chosen to maximise data collection given the limited timeframe while minimising the burden for schools.

To explore the full cascade model of training, members of the evaluation team conducted observations at the various stages of training using a semi-structured observation framework supplemented by fieldnotes. Pre-implementation, all sessions of the three-day course for regional trainers in July 2018 were observed and the evaluation team reviewed the presentation slides used at the subsequent one-day session in September 2018 rather than attending in person. Members of the evaluation team also attended six of the training sessions for teachers: four full days (two ICT, two paper-based) and two half-days (one ICT, one paper-based) as well as obtaining copies of the attendance records from all the sessions.

A baseline survey was distributed to the project lead of all participating schools. This aimed to establish usual (or baseline) teaching practices in literacy (programmes used, how pupils grouped, role of TAs, and so forth) and the use of IT in school and personally. Schools were asked to ensure it was completed by:

- the headteacher, project lead, or literacy coordinator;
- all Year 1 class teachers; and
- all staff members who would receive training in the RUKS programme (ICT or paper-based) and who would deliver the programme (if school randomised to an appropriate trial arm).

Schools were advised that a response to the survey was mandatory in order to proceed to randomisation, a strategic approach to ensure the highest level of response. However, due to time constraints, seven schools were randomised in the absence of completed survey data. In total, 150 of 157 schools provided a response (96%), representing a total of 501 individuals. Completed surveys were submitted between 17 September and 5 November 2018.

Table 5: IPE methods overview—pre-implementation stage

Research methods	Method of data collection	Who or what?	N	Research questions addressed	Why?
Observation	Observation framework/fieldnotes	Training sessions for regional trainers	1 x 3-day	RQ2a	To establish the expected model in both of the programme arms.
Observation	Observation framework/fieldnotes	Training sessions for school staff	4 x 1-day 2 x 0.5-day	RQ2a	To explore fidelity in terms of cascaded training.
Documentary evidence	Training attendance records	School staff	All	RQ2d	To establish compliance and fidelity in terms of staff attendance at training.
Survey	Online survey	Trainers (who train the school staff)	10	RQ2a	To explore trainer perspective on the training provided to them and how this equipped them to train the school staff.

Research methods	Method of data collection	Who or what?	N	Research questions addressed	Why?
Survey	Online surveys	Year 1 class teachers	232	RQ1	Establish baseline for literacy practice.
		Other (TA etc.)	134		Who would schools plan to send to training?
		Headteacher or literacy coordinator	135		To gain an overview of the wider school perspective on the programmes and approaches to literacy in KS1.

Implementation stage

IPE activities undertaken during the programme implementation period sought to provide insight as to what the programme looked like in practice. Multiple data collection methods were employed across a geographical spread of schools throughout the programme period. In total, 25 schools were visited comprising of ten in each of the programme arms and five in the control arm:

- Thirty-one sessions (12 ICT, 13 paper-based, and six control) were observed to explore any variations in delivery across different settings and with different pupils.
- Pupil focus groups were held straight after programme sessions with groups of three to five pupils (24 in total at six schools) to elicit their views of each of the programme arms. Focus group discussions centred on likes and dislikes of the programme sessions, reactions to group working, and opinions of different elements of the programme.
- School staff involved in the delivery of programme sessions were interviewed at a selection of the visited schools (one interview in each of 15 schools including three control schools). This was two more interviews than planned (an extra paper-based and an extra ICT school where the deliverers were keen to be interviewed after their pupils had participated in focus groups). The interviews were used to gather opinions and suggestions about the implementation of the two programme delivery models and similar information about the 'business as usual' approaches being used in the control group. The interviews were conducted with Year 1 teachers and teaching assistants and occasionally staff with other roles such as reading support leaders.

Evaluation team members made five of the visits in pairs to enhance reliability and consistency, and another four of the observations were conducted when the regional trainer was carrying out their Week 10 visit. Where permission was given by the relevant individuals, interviews and focus groups were audiotaped and transcribed for ease of analysis. Audio data was stored securely and anonymously in accordance with the Data Protection Act (2018).

During the implementation period, a further training day for regional trainers was held (January 2019). This was timed to be prior to the Week 10 visits and focused on the use of the fidelity of implementation tool. The session was observed by two members of the evaluation team who were also trained in the use of the tool.

Table 6: IPE methods overview—implementation stage

Research methods	Method of data collection	Who or what?	N	Research questions addressed	Why?
Observation	Observation framework/fieldnotes	Training sessions for regional trainers	1 x 1-day	RQ2	To develop and moderate application of tool to assess fidelity of implementation.
Case study	Observation framework/fieldnotes (including Fidelity of Implementation tool developed with implementation team)	A range of participating schools ICT arm Non-ICT arm Control TOTAL:	 12 13 6 31	RQ2, RQ2a, RQ2b, RQ2c, RQ3, RQ3a, RQ5a	To explore what the programmes look like in the classroom (including adaptations and/or barriers), school context, pupil engagement, and embeddedness.
Case study	Interviews Pre-agreed coding frame for analysis (thematic)	School staff ICT arm Paper-based arm Control TOTAL:	 6 6 3 15	RQ2 RQ2a, RQ2b, RQ2c, RQ3, RQ3b, RQ3c, RQ4, RQ5, RQ6	To discuss observed sessions; differences in implementation; adaptations and barriers (actual or perceived); pupil engagement; perceived value of programmes (including for specific sub-sets of pupils).
Case study	Pupil focus groups (max 20 minutes with 4–5 pupils in each group)[7]	ICT arm Paper-based arm TOTAL:	3 3 6	RQ2, RQ2c, RQ6	To discuss observed sessions; explore pupil understanding and perception of programmes.

Post-implementation stage

The final, post-implementation stage of the IPE comprised surveys of both school staff and regional trainers as well as an interview with members of the implementation team. School surveys were distributed via email to all participating schools (programme and control). Surveys were tailored to the various types of respondents, for example, a teaching assistant, a non-delivering teacher, or a delivering teacher. These surveys, in conjunction with pre-randomisation baseline surveys, were used to explore any changes in practice over the course of the trial, interventions or programmes received by children in the control arm, and stakeholder views on programme implementation for schools in the programme arms. A total of 283 surveys from 142 of 157 schools were returned (90%), although ten were deemed unusable because the allocation they reported in the survey did not match their actual allocation, possibly due to confusion or human error during the completion of the survey. The survey was completed between 27 June and 23 July 2019.

The regional trainer survey was also distributed via email and aimed to capture views on the training they had received, their confidence in training school staff, their relationship with the schools they were responsible for, and their own opinions of the programme. Eight of the ten regional trainers contacted completed the survey in September and October 2019.

[7] Research (and experience of the researchers) indicates that focus groups with students in KS1 settings is appropriate if consideration is given to how the discussion is facilitated, group size, and length of the focus group (Heary and Hennessy, 2002; Morgan, Gibbs, Maxwell and Britten, 2002; O'Reilly, Ronzoni and Dogra, 2013).

The two members of the implementation team responsible for developing the RUKS programme participated in a paired, semi-structured interview in September 2019.

Table 7: IPE methods overview—post-implementation stage

Research methods	Method of data collection	Who or what?	N	Research questions addressed	Why?
Survey	Online survey	Year 1 class teacher	137	RQ2, RQ2a, RQ2b, RQ2c, RQ2d, RQ3, RQ3a, RQ3b, RQ3c, RQ6	To assess: how well programmes have been embedded in schools; teacher confidence in programmes; fidelity of implementation (teacher reported usage, adaptations made, and opinions); associated costs. To estimate school staff time required to facilitate the programmes. In control schools—to establish the presence of any compensatory practice or change in practice related to awareness of interventions.
		TAs (programme only)	81		
		Headteacher or literacy coordinator	14		
Survey	Online survey	Regional trainers	8	RQ2, RQ2a, RQ2b, RQ2c, RQ6	To explore how well their training prepared them to deliver to teachers; confidence in delivery; ease of school access; responsiveness of teachers; perceptions of programme.
Paired interview	Semi-structured, face to face	Implementation team	1	RQ2a, RQ2c, RQ6	To reflect on the programme period and explore any changes that may have occurred. in terms of the programmes. To consider the future feasibility and sustainability of the programmes.

Quality assurance

To ensure consistency, observations and visits were conducted by more than one member of the evaluation team in the following cases:

- all four of the observed ‘train-the-trainer’ regional trainer sessions;
- four of the six observed school staff training sessions; and
- five of the 24 school visits.

After the joint school visits, the members of the evaluation team compared their observation sheets, discussed and resolved any discrepancies, and came to an agreement about the ratings. All members of the evaluation team who did observations independently first attended one of these joint moderation visits.

Analysis

As presented in Table 6 and Table 7, we took a pragmatic approach to the data collection for the IPE and, accordingly, different analytical approaches were applied in terms of analysis. The analysis conducted was driven by the logic model (Figure 1) and all data was interpreted through the lens of the overarching research questions. The logic model included outcomes that are not captured through the impact evaluation but are explored through the IPE. For example, the training process and barriers and facilitators to successful implementation.

Interviews and focus groups were audio recorded, transcribed verbatim, and then data managed using NVivo Version 12. An inductive approach to data analysis was primarily used for interviews, focus groups, and free text answers to questions in the online pre- and post-implementation surveys. This approach allowed us to (1) ‘to condense extensive and varied raw text data into a brief, summary format; and (2) to establish clear links between the research objectives and the summary findings derived from the raw data’ (Thomas, 2006, p. 237). On a practical level this involved multiple readings of the raw data to identify frequent and dominant themes from which a coding framework was developed and the transcripts subsequently coded. Schools that participated in the interviews and focus groups were allocated a unique ID and interview participants were described only by their role and trial arm to assure anonymity and confidentiality.

The free text survey responses yielded through surveys were analysed through thematic analysis where Braun and Clarke's (2006) six-phase process was used:

- familiarising yourself with the dataset;
- coding;
- generating initial themes;
- developing and reviewing themes;
- refining, defining, and naming themes; and
- writing up.

A mixture of inductive and deductive coding was used: we studied the qualitative data, drew up a list of codes through inductive analysis (that is, bottom-up from the data), then grouped them into themes. The themes were then interpreted deductively according to each research question. Each research question uses data from different sources, some in combination. The data is presented in a such a way as to make the findings the most accessible for the reader.

Costs

Programme costs
Data on costs was collected directly from the developers of the RUKS programme at Nottingham Trent University as well as schools via the post-implementation survey. In the context of the trial, the only direct monetary cost associated with the programme was the cost to schools for the provision of training. In the Costs section of this report, we detail the costs associated with implementing each version of the programme in a 'real-world' context.

Time costs
Data regarding the time involved at the school level to facilitate the programme for both delivery models were collected directly from school staff via the post-implementation survey and session logs.

Timeline

Table 8: Timeline

Dates	Activity	Staff responsible/leading
May 2017	Project start date	-
15 May 2017	Set-up meeting 1	All
22 August 2017	Set-up meeting 2	All
September 2017	IDEA workshop	Evaluation and implementation teams
December 2017	Ethics approval for MOU	Evaluation team
October 2017–April 2018	Protocol development	All
April 2018	Ethics application	Evaluation team
April 2018	ISRCTN application	Evaluation team
April 2018	Development of baseline survey	Evaluation team
January–August 2018	Recruitment of schools	All
September 2018	Participating pupil identification and informing parents/carers	Evaluation team and schools
September 2018	Baseline surveys—school staff and trainers	Evaluation team and schools
September 2018	Pre-testing	Administered by schools
September 2018	Randomisation	Evaluation team

Dates	Activity	Staff responsible/leading
October 2018	Training of identified school staff	Implementation team
October 2018–May 2019	Programme delivery	Schools
October 2018–May 2019	School support	Implementation team
October 2018–May 2019	School observations	Implementation and evaluation teams
October 2018–May 2019	Staff interviews	Evaluation team
March 2019	Pre-test marking	Evaluation team
April 2019–June 2019	Development of post-implementation survey	Evaluation team
May–July 2019	Post-testing	Evaluation team
July 2018	Post-implementation survey	Evaluation team
June–August 2019	Post-testing marking	Evaluation team
January-March 2022	Data analysis	Evaluation team
March 2022	Draft report	Evaluation team
August 2022	Final report	Evaluation team

Impact evaluation results

The data presented in this section resulting in low counts was obtained via the research team and was in no way informed by data from the ONS.

Participant flow including losses and exclusions

School recruitment and attrition

The target for recruitment was 201 schools. Memorandums of understanding were returned by 179 schools, 13 of which withdrew before randomisation. No data was collected for any of the schools that withdrew pre-randomisation. The reason for withdrawal for the majority of these schools was insufficient staff capacity for the project, with one school additionally mentioning that it felt that the programme would not fit with its current teaching of reading.

In total, 166 schools were randomised (ICT: n = 51; paper-based: n = 57; control: n = 58); this is 82.6% of the target of 201 schools. Of these, 166, eight (5.1%) reported not having access to ICT facilities and as a result were randomised to either paper-based or control only (paper-based: n = 3; control: n = 5). Nine (5.4%) of the randomised schools failed to return valid pre-test data (ICT: n = 4/51, 7.8%; paper-based: n = 1/57, 1.8%; control: n = 4/58. 6.9%). Of these nine, seven did not complete the child recruitment stage or provide pupil details, which were necessary to progress to pre-testing. Of these seven schools, four withdrew due to one of the following reasons: staff absence, not understanding the time commitment for the project, unforeseen circumstances, and concerns that the opt out consent approach was not GDPR compliant (despite reassurances). The remaining three schools that did not provide pupil details became unresponsive to contact attempts. One school that did provide pupil details did not go on to complete pre-testing as it wished to self-select children to be involved and then became unresponsive after the evaluation team agreed to the self-selection. Finally, one school provided pupil details and completed the pre-test, but the assessment booklets were unfortunately lost in the post as the school posted these rather than waiting for the evaluation team to arrange a courier collection.

In total, 157 schools—ICT: n = 47; paper-based: n = 56; control: n = 54—were randomised and returned valid pre-test data. Thirty-three (21.0%) carried out the pre-test after being informed of their allocation—ICT: n = 8/47 (17.0%); paper-based: n = 11/56 (19.6%); control: n = 14/54 (25.9%). These 157 schools were asked to post-test their participating pupils. Of these, three (1.9%) withdrew from post-testing, all of which were in the paper-based arm (5.4%). One school declined post-testing as it thought pupils would be too overwhelmed with the assessment process; this school had a high number of children with SEN in the class and had withdrawn from programme delivery due to being unable to fit the ABRA programme in alongside the many other interventions it delivered to meet the needs of the class. The second school withdrew from post-testing as the programme lead for the school left and the headteacher did not wish to continue in the trial in any capacity—this school had withdrawn from programme delivery prior to completing the programme training, citing a change in circumstances, which meant it no longer felt able to participate. Finally, the third school withdrew from post-testing as it did not wish to set aside time to facilitate the post-test.

In summary, 154 of 166 (92.8%) schools—ICT: n = 47/51 (92.2%); paper-based: n = 53/57 (93.0%); control: n = 54/58 (93.1%)—were randomised and returned both pre-test and post-test data.

Pupil recruitment and attrition

The target sample size was 5,427 pupils. In total, 4,014 pupils (74.0% of the target) completed the pre-test and were randomised (ICT: n = 1,158; paper-based: n = 1,471; control: n = 1,385). However, seven pupils from one school in the paper-based arm were randomised in error and are excluded from all further analyses. The school had stated it only intended to deliver the intervention to one class, and this class was the one put forward for randomisation. However, a second class (with these seven pupils) that the school had not put forward was randomised in error. It was this class that comprised the seven pupils randomised in error. The other pupils from this school remained in the study. Of the 4,007 remaining pupils, 168 (4.2%) were from the eight schools with no ICT facilities (paper-based: n = 47; control: n = 121).

Of the 4,007 eligible pupils randomised, 545 (13.6%) did not complete the post-test for the primary outcome—ICT: n = 120/1158 (10.4%); paper-based: n = 285/1464 (19.5%); control: n = 140/1385 (10.1%). Reasons for non-completion are given in **Error! Reference source not found.**9.

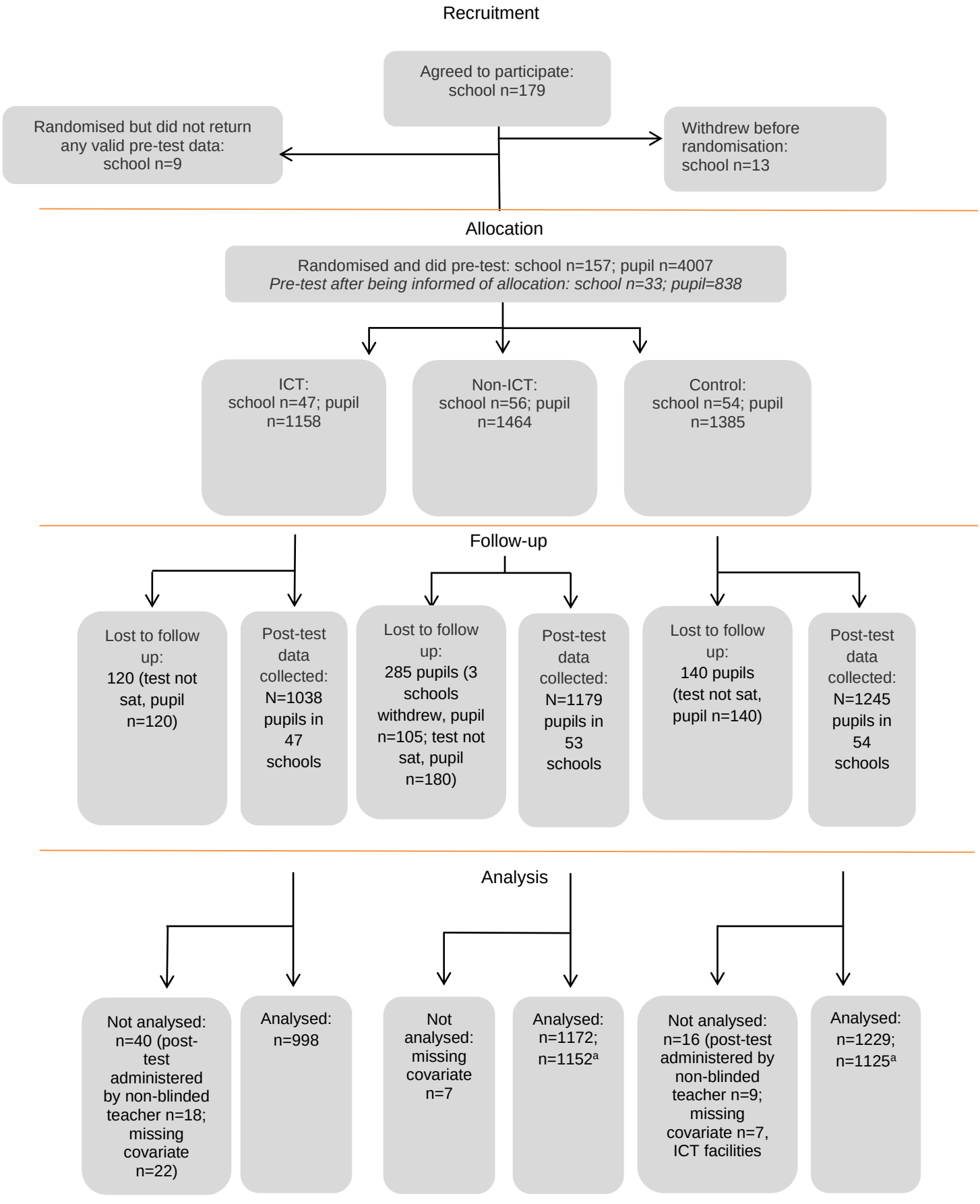
Table 9: Reasons for non-completion of the post test for the primary outcome

	ICT	Paper-based	Control
Reason for non-completion, n (%)			
School withdrew	0/120 (0)	105/285 (36.8)	0/140 (0)
Pupil absent	120/120 (100)	180/285 (63.2)	140/140 (100)

In summary, 3,462 (86.4%) randomised pupils completed both the pre-test and post-test for the primary outcome (ICT: n = 1038/1158, 89.6%; paper-based: n = 1179/1464, 80.5%; control: n = 1245/1385, 89.9%). Due to teachers wishing to conduct the PiRA for their own purposes prior to the post-test being scheduled, 18 pupils from one school randomised to the ICT arm and nine pupils from one school randomised to the control arm had their PiRA post-test facilitated and marked by a teacher at the school rather than by a research assistant blinded to group allocation. These 27 pupils were excluded from the primary analysis models and only included in a post-hoc sensitivity analysis looking at the impact of including their data on the primary analysis models results. A further 36 pupils could not be included in the primary analysis as they were missing the EYFSP Good Level of Development (FSP_GLD) variable from the NPD, which was included as a covariate in the models. This resulted in 3,399 analysable pupils—ICT: 998/1158 (86.2%); paper-based: 1172/1464 (80.1%); and control: 1229/1385 (88.7%). Therefore, of the 2,849 pupils randomised to paper-based or control and also pre-tested, 2,401 (84.3%) were included in the primary analysis comparing the paper-based intervention (1172/1464, 80.1%) with control (1229/1385, 88.7%).

Of the 4,007 randomised pupils who did the pre-test, 168 (4.2%) were from schools with no ICT facilities, therefore, 3,839 pupils were from schools with ICT facilities. Of these, 3,275 (85.3%) were included in the primary analysis model—ICT: 998/1158 (86.2%); paper-based: 1152/1417 (81.3%); control: 1125/1264 (89.0%). These groups form the comparisons between the ICT intervention and paper-based, and between ICT and control.

Figure 3: Participant flow diagram



[a] included in analysis for comparison with ICT intervention group, as this analysis excludes schools without ICT facilities

Minimum detectable effect size

Table 10: Minimum detectable effect size at different stages for schools with ICT facilities

		Protocol		Randomisation		Analysis[1]	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES (Cohen's d)		0.20	0.23	0.21	0.28	0.20	0.24
Pre-test/post-test correlations	Level 1 (pupil)	0.45	0.45	0.45	0.45	0.64	0.60
Intracluster correlations (ICCs)	Level 2 (school)	0.15	0.15	0.15	0.15	0.17	0.11
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-sided	two-sided	two-sided	two-sided	two-sided	two-sided
Average cluster size		27	7	26	6	23	6
Number of schools	ICT	67	67	47	47	46	42
	Paper-based	67	67	56	56	51	48
	Control	67	67	54	54	48	42
	Total:	201	201	157	157	145[2]	132
Number of pupils	ICT	1809	452	1158	267	998	263
	Paper-based	1809	452	1464	338	1152	310
	Control	1809	452	1385	318	1125	256
	Total:	5427	1356	4007	923	3275	829

[1] Effect size at analysis stage calculated for pupils from schools with ICT facilities who returned valid pre- and post-test data and had data for covariates and so were included in the primary analysis model.
[2] There were 149 schools with ICT facilities, two of which were not included in the primary analysis due to the test being marked by a teacher at the school, and two not being included having withdrawn before post-testing. There were three withdrawals before post-testing, two of which were from schools with ICT facilities and one from a school without ICT facilities.

Attrition

For the 3,839 pupils randomised from schools with ICT facilities, 564 (14.7%) were excluded from the primary analysis. Further information on attrition is given in Tables 11 and 12.

Table 11: Pupil-level attrition from the trial (primary outcome) for schools with ICT facilities

		ICT	Paper-based	Control	Total
Number of pupils	Randomised	1158	1417	1264	3839
	Analysed	998	1152	1125	3275
Pupil attrition (from randomisation to analysis)	Number	160	265	139	564
	Percentage	13.8	18.7	11.0	14.7

Table 12: Reasons for pupil-level attrition from the trial for schools with ICT facilities (reasons presented overall to avoid disclosure)

Reason for not being included in analysis	Total (n = 564) Count (%)
Did not take post-test	501 (88.8)
Missing FSP_GLD covariate	36 (6.4)
PiRA completed by teachers rather than blind assessors	27 (4.8%)

For the 2,849 pupils randomised to paper-based or control, 448 (15.7%) were excluded from the primary analysis. Further information on attrition is given in Table13 and 14.

Table 13: Pupil-level attrition from the trial for those randomised to paper-based or control (primary outcome)

		Paper-based	Control	Total
Number of pupils	Randomised	1464	1385	2849
	Analysed	1172	1229	2401
Pupil attrition (from randomisation to analysis)	Number	292	156	448
	Percentage	19.9	11.3	15.7

Table 14: Reasons for pupil-level attrition from the trial for schools randomised to paper-based or control (reasons presented overall to avoid disclosure)

Reason for not being included in analysis	Total (n = 448) Count (%)
Did not take post-test	425 (94.9)
Missing FSP_GLD covariate	14 (3.1)
PiRA completed by teachers rather than blind assessors	9 (2.0)

Pupil and school characteristics

Table 15 summarises baseline characteristics by randomised group for the schools with ICT facilities. The mean percentage of FSM pupils eligible for FSM at the school level was 31.9 (SD 18.4). There were 117 (78.5%) schools from an urban setting. The majority of schools were community schools (n = 61; 40.9%) and the most common Ofsted rating was 'good' (n = 88; 59.1%). There were 2,631 (68.5%) pupils who achieved a good level of development on the EYFSP G09.

There was a higher proportion of academy-converter mainstream schools, and a lower proportion of community schools, in the ICT group than in the paper-based and control groups. There was also a lower proportion of schools rated 'outstanding' by Ofsted and a higher proportion with no Ofsted rating in the ICT group than in the other two groups. There was a slightly lower proportion of schools in an urban setting in the paper-based group than in the other two groups. The groups were roughly comparable across all other measured characteristics.

There was no difference between groups in the PiRA raw score (ICT versus control: -0.01, 95% CI: -0.19, 0.18; paper-based versus control: -0.03, 95% CI: -0.24, 0.18; ICT versus paper-based: 0.02, 95% CI: -0.18, 0.22), nor were there differences between groups in the PiRA age-standardised score (ICT versus control: -0.01, 95% CI: -0.20, 0.17; paper-based versus control: -0.04, 95% CI: -0.25, 0.18; ICT versus paper-based: 0.03, 95% CI: -0.18, 0.23). The PiRA age-standardised score was normally distributed (Appendix D, Figure 2).

Table 15: Baseline characteristics of groups as randomised for the schools with ICT facilities

School level (categorical)	ICT		Paper-based		Control		
	n/N (missing)	Count (%)	n/N (missing)	Count (%)	n/N (missing)	Count (%)	
School type							
Academy: converter mainstream	19/47 (0)	19 (40.4)	14/53 (0)	14 (26.4)	11/49 (0)	11 (22.4)	
Academy: sponsor led mainstream	4/47 (0)	4 (8.5)	5/53 (0)	5 (9.4)	6/49 (0)	6 (12.2)	
Community school	13/47 (0)	13 (27.7)	25/53 (0)	25 (47.2)	23/49 (0)	23 (46.9)	
Foundation/ voluntary aided/voluntary controlled	11/47 (0)	11 (23.4)	9/53 (0)	9 (17.0)	9/49 (0)	9 (18.4)	
Ofsted rating							
Outstanding	3/47 (0)	3 (6.4)	8/53 (0)	8 (15.1)	8/49 (0)	8 (16.3)	
Good	28/47 (0)	28 (59.6)	31/53 (0)	31 (58.5)	29/49 (0)	29 (59.2)	
Requires improvement	3/47 (0)	3 (6.4)	6/53 (0)	6 (11.3)	5/49 (0)	5 (10.2)	
Inadequate	0/47 (0)	0 (0)	0/53 (0)	0 (0)	0/49 (0)	0 (0)	
No Ofsted	13/47	13 (27.7)	8/53	8 (15.1)	7/49	7 (14.3)	

	(0)		(0)		(0)		
School setting							
Urban	38/47 (0)	38 (80.9)	39/53 (2)	39 (73.6)	40/49 (1)	40 (81.6)	
Rural	9/47 (0)	9 (19.1)	12/53 (2)	12 (22.6)	8/49 (1)	8 (16.3)	
ICT facilities							
Yes	47/47 (0)	47 (100)	53/53 (0)	53 (100)	49/49 (0)	49 (100)	
No	0/47 (0)	0 (0)	0/53 (0)	0 (0)	0/49 (0)	0 (0)	
Location							
East Midlands	10/47 (0)	10 (21.3)	11/53 (0)	11 (20.8)	11/49 (0)	11 (22.4)	
Manchester	9/47 (0)	9 (19.1)	14/53 (0)	14 (26.4)	12/49 (0)	12 (24.5)	
Newcastle	13/47 (0)	13 (27.7)	14/53 (0)	14 (26.4)	13/49 (0)	13 (26.5)	
Teesside	7/47 (0)	7 (14.9)	5/53 (0)	5 (9.4)	5/49 (0)	5 (10.2)	
West Midlands	8/47 (0)	8 (17.0)	9/53 (0)	9 (17.0)	8/49 (0)	8 (16.3)	
School level (continuous)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	
School size	47 (0)	285.3 (129.0) 248 (208, 329)	>50 (<3)	301.4 (136.7) 274 (211, 416)	>46 (<3)	354.6 (205.9) 347 (210.5, 434)	
% eligible for FSM	47 (0)	32.6 (19.5) 26.7 (15.1, 48.6)	>50 (<3)	32.4 (17.9) 30.1 (16.6, 45.0)	>46 (<3)	30.7 (18.0) 28.9 (17.9, 45.35)	
% pupils whose first language is not English	47 (0)	11.6 (16.7) 4.2 (1.2, 16.8)	>50 (<3)	18.2 (26.0) 4.1 (1.5, 27.5)	>46 (<3)	10.3 (14.4) 3.7 (1.7, 12.8)	
% SEN or EHC	47 (0)	1.0 (1.3) 0.4 (0, 1.4)	>50 (<3)	0.9 (1.1) 0.8 (0.2, 1.2)	>46 (<3)	1.0 (0.8) 0.9 (0.3, 1.4)	
Pupil level (categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)	n/N (missing)	Count (%)	
Sex, male	615/1158 (0)	615 (53.1)	722/1417 (0)	722 (51.0)	677/1264 (0)	677 (53.6)	
Ever FSM							
Yes	310/1158	310 (26.8)	362/1417	362 (25.5)	308/1264	308 (24.4)	

	(7)		(57)		(2)		
No	841/1158 (7)	841 (72.6)	998/1417 (57)	998 (70.4)	954/1264 (2)	954 (75.5)	
EYFSP GLD							
Yes	809/1158 (34)	809 (69.9)	931/1417 (66)	931 (65.7)	891/1264 (12)	891 (70.5)	
No	315/1158 (34)	315 (27.2)	420/1417 (66)	420 (29.6)	361/1264 (12)	361 (28.6)	
Pupil level (continuous)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	Hedges' g effect sizes (95% CI)
PiRA raw-standardised score	1158 (0)	12.0 (6.5) 12 (7, 17)	1417 (0)	11.7 (6.8) 12 (6, 17)	1264 (0)	12.2 (6.7) 12 (7, 18)	ICT vs control: -0.01 (-0.19, 0.18); paper-based vs control: -0.03 (-0.24, 0.18); ICT vs paper-based: 0.02 (-0.18, 0.22)
PiRA age-standardised score	1157 (1)	99.8 (13.1) 100 (91, 109)	1417 (0)	99.0 (13.9) 100 (90, 108)	1264 (0)	100.0 (13.6) 101 (91, 109)	ICT vs control: -0.01 (-0.20, 0.17); paper-based vs control: -0.04 (-0.25, 0.18); ICT vs paper-based: 0.03 (-0.18, 0.23)

Table 16 summarises baseline characteristics by randomised group for the schools randomised to paper-based or control.

The mean percentage of FSM pupils eligible for FSM was 32.0 (SD 18.5). There were 85 (77.3%) schools from an urban setting. The majority of schools were community schools (n = 52; 47.3%) and the most common Ofsted rating was 'good' (n = 65; 59.1%). There were 1,918 (67.3%) pupils who achieved a good level of development on the EYFSP G09.

Schools were well balanced across the two groups with the exception that there was a slightly higher proportion of urban schools in the control group than in the paper-based group, the mean school size was higher in the control group, and there was a higher proportion of EAL pupils in the paper-based group.

There was no difference between groups in the PiRA raw score (paper-based versus control: -0.02, 95% CI: -0.22 to 0.19), nor were there differences between groups in the PiRA age-standardised score (paper-based versus control: -0.02, 95% CI: 0.23 to 0. 18).

Table 16: Baseline characteristics of groups as randomised for the schools allocated to paper-based or control

School level (categorical)	Paper-based		Control		
	n/N (missing)	Count (%)	n/N (missing)	Count (%)	

School type					
Academy: converter mainstream	14/56 (0)	14 (25.0)	13/54 (0)	13 (24.1)	
Academy: sponsor led mainstream	5/56 (8.9)	5 (8.9)	6/54 (0)	6 (11.1)	
Community school	27/56 (0)	27 (48.2)	25/54 (0)	25 (46.3)	
Foundation/ voluntary aided/voluntary controlled	10/56 (0)	10 (17.9)	11/54 (0)	11 (20.4)	
Ofsted rating					
Outstanding	8/56 (0)	8 (14.3)	8/54 (0)	8 (14.8)	
Good	33/56 (0)	33 (58.9)	32/54 (0)	32 (59.3)	
Requires improvement	7/56 (0)	7 (12.5)	5/54 (0)	5 (9.3)	
Inadequate	0/56 (0)	0 (0)	0/54 (0)	0 (0)	
No Ofsted	8/56 (0)	8 (14.3)	9/54 (0)	9 (16.7)	
School setting					
Urban	41/56 (2)	41 (73.2)	44/54 (1)	44 (81.5)	
Rural	13/56 (2)	13 (23.2)	9/54 (1)	9 (16.7)	
ICT facilities					
Yes	53/56 (0)	53 (94.6)	49/54 (0)	49 (90.7)	
No	3/56 (0)	3 (5.4)	5/54 (0)	5 (9.3)	
Location					
East Midlands	12/56 (0)	12 (21.4)	12/54 (0)	12 (22.2)	
Manchester	16/56 (0)	16 (28.6)	14/54 (0)	14 (25.9)	
Newcastle	14/56 (0)	14 (25.0)	14/54 (0)	14 (25.9)	
Teesside	5/56 (0)	5 (8.9)	5/54 (0)	5 (9.3)	
West Midlands	9/56 (0)	9 (16.1)	9/54 (0)	9 (16.7)	
School level (continuous)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	

School size	>53 (<3)	294.1 (137.1) 247 (199, 410)	>51 (<3)	351.2 (198.8) 340 (216, 434)	
% eligible for FSM	>53 (<3)	32.0 (18.1) 30.0 (18.1, 45.0)	>51 (<3)	32.0 (19.1) 30.7 (18.1, 46.2)	
% pupils whose first language is not English	>53 (<3)	17.4 (25.5) 4.7 (1.8, 18.2)	>51 (<3)	10.4 (14.4) 3.8 (1.8, 13.6)	
% SEN or EHC	>53 (<3)	1.0 (1.2) 0.8 (0.2, 1.3)	>51 (<3)	1.1 (1.1) 0.9 (0.3, 1.5)	
Pupil level (categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)	
Sex, male	755/1471 (0)	755 (51.3)	736/1385 (0)	736 (53.1)	
Ever FSM Yes No	 364/1464 (84) 1016/1464 (84)	 364 (24.9) 1016 (69.4)	 353/1385 (2) 1030/1385 (2)	 353 (25.5) 1030 (74.4)	
EYFS GLD Yes No	 943/1464 (93) 428/1464 (93)	 943 (64.4) 428 (29.2)	 975/1385 (12) 398/1385 (12)	 975 (70.4) 398 (28.7)	
Pupil level (continuous)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	Hedges' g effect sizes (95% CI)
PiRA raw- score	1471 (0)	11.7 (6.8) 12 (6, 17)	1385 (0)	12.1 (6.7) 12 (7, 18)	Paper-based vs control: -0.02 (-0.22, 0.19)
PiRA age-standardised score	1471 (0)	99.1 (13.9) 100 (90, 108)	1385 (0)	100.0 (13.5) 101 (91, 109)	Paper-based vs control: -0.02 (0.23, 0.18)

Table 17 summarises baseline characteristics by randomised treatment group for the schools included in the primary analysis model for schools with ICT facilities. There was no difference between groups in the PiRA raw score (ICT versus control: -0.01, 95% CI: -0.20 to 0.18; paper-based versus control: -0.07, 95% CI: -0.29 to 0.16; ICT versus paper-based: 0.06, 95% CI: -0.16 to 0.27), nor were there differences between groups in the PiRA age-standardised score (ICT versus control: 0.02, 95% CI: -0.21 to 0.18; paper-based versus control: -0.07, 95% CI: -0.29 to 0.15; ICT versus paper-based: 0.06, 95% CI: -0.16 to 0.27).

With the exception of school type, school setting and Ofsted rating (as with the 'as randomised participants'), baseline characteristics were well-balanced between groups.

Table 17: Baseline characteristics of groups as analysed for the schools with ICT facilities

School level (categorical)	ICT		Paper-based		Control	
	n/N (missing)	Count (%)	n/N (missing)	Count (%)	n/N (missing)	Count (%)
School type						
Academy: converter mainstream	19/46 (0)	19 (41.3)	13/51 (0)	13 (25.5)	11/48 (0)	11 (22.9)
Academy: sponsor led mainstream	4/46 (0)	4 (8.7)	5/51 (0)	5 (9.8)	6/48 (0)	6 (12.5)
Community school	13/46 (0)	13 (28.3)	24/51 (0)	24 (47.1)	22/48 (0)	22 (45.8)
Foundation/ voluntary aided/ voluntary controlled	10/46 (0)	10 (21.7)	9/51 (0)	9 (17.6)	9/48 (0)	9 (18.8)
Ofsted rating						
Outstanding	3/46 (0)	3 (6.5)	8/51 (0)	8 (15.7)	8/48 (0)	8 (16.7)
Good	27/46 (0)	27 (58.7)	30/51 (0)	30 (58.8)	28/48 (0)	28 (58.3)
Requires improvement	3/46 (0)	3 (6.5)	6/51 (0)	6 (11.8)	5/48 (0)	5 (10.4)
Inadequate	0/46 (0)	0 (0)	0/51 (0)	0 (0)	0/48 (0)	0 (0)
No Ofsted	13/46 (0)	13 (28.3)	7/51 (0)	7 (13.7)	7/48 (0)	7 (14.6)
School setting						
Urban	38/46 (0)	38 (82.6)	38/51 (2)	38 (74.5)	39/48 (1)	39 (81.3)
Rural	8/46 (0)	8 (17.4)	11/51 (2)	11 (21.6)	8/48 (1)	8 (16.7)
ICT facilities						
Yes	46/46 (0)	46 (100)	51/51 (0)	51 (100)	48/48 (0)	48 (100)
No	0/46 (0)	0 (0)	0/51 (0)	0 (0)	0/48 (0)	0 (0)
Location						
East Midlands	10/46 (0)	10 (21.7)	9/51 (0)	9 (17.6)	11/48 (0)	11 (22.9)
Manchester	9/46 (0)	9 (19.6)	14/51 (0)	14 (27.5)	11/48 (0)	11 (22.9)
Newcastle	13/46	13 (28.3)	14/51	14 (27.5)	13/48	13 (27.1)

	(0)		(0)		(0)		
Teesside	6/46 (0)	6 (13.0)	5/51 (0)	5 (9.8)	5/48 (0)	5 (10.4)	
West Midlands	8/46 (0)	8 (17.4)	9/51 (0)	9 (17.6)	8/48 (0)	8 (16.7)	
School level (continuous)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	
School size	46 (0)	287.8 (129.2) 248.5 (213, 329)	>48 (<3)	303.1 (138.7) 274 (212, 416)	>45 (<3)	352.9 (207.8) 341 (209, 434)	
% eligible for FSM	46 (0)	32.2 (19.5) 26.7 (15.1, 46.4)	>48 (<3)	32.6 (17.8) 30.1 (20.2, 44.8)	>45 (<3)	30.4 (18.0) 27.1 (17.5, 44.8)	
% pupils whose first language is not English	46 (0)	11.5 (16.9) 4.2 (1.6, 14.7)	>48 (<3)	17.7 (25.8) 4.1 (1.8, 18.2)	>45 (<3)	10.1 (14.5) 3.6 (1.6, 10.6)	
% SEN or EHC	46 (0)	0.9 (1.4) 0.4 (0, 1.3)	>48 (<3)	1.0 (1.1) 0.8 (0.2, 1.2)	>45 (<3)	1.0 (0.8) 0.8 (0.3, 1.5)	
Pupil level (categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)	n/N (missing)	Count (%)	
Sex, male	532/998 (0)	532 (53.3)	596/1152 (0)	596 (51.7)	599/1125 (0)	599 (53.2)	
Ever FSM							
Yes	263/998 (0)	263 (26.4)	310/1152 (0)	310 (26.9)	256/1125 (0)	256 (22.8)	
No	735/998 (0)	735 (73.6)	842/1152 (0)	842 (73.1)	869/1125 (0)	869 (77.2)	
EYFS GLD							
Yes	739/998 (0)	739 (74.0)	819/1152 (0)	819 (71.1)	833/1125 (0)	833 (74.0)	
No	259/998 (0)	259 (26.0)	333/1152 (0	333 (28.9)	292/1125 (0)	292 (26.0)	
Pupil level (continuous)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	Hedges' g effect sizes (95% CI)
PiRA raw-score	998 (0)	12.3 (6.3) 12 (7, 17)	1152 (0)	11.9 (6.8) 12 (6, 17)	1125 (0)	12.6 (6.6) 13 (7, 18)	ICT vs control: -0.01 (-0.20, 0.18); paper-based vs control: -0.07 (-0.29, 0.16); ICT vs paper-based: 0.06 (-0.16, 0.27);

PiRA age-standardised score	998 (0)	100.4 (12.7) 100 (93, 109)	1152 (0)	99.4 (13.9) 101 (90, 109)	1125 (0)	100.9 (13.4) 102 (92, 109)	ICT vs control: 0.02 (-0.21, 0.18); paper-based vs control: -0.07 (-0.29, 0.15); ICT vs paper-based: 0.06 (-0.16, 0.27)

Error! Reference source not found.8 summarises baseline characteristics by randomised treatment group for the schools included in the primary analysis model for schools randomised to paper-based or control. Baseline characteristics were well-balanced between groups.

Table 18: Baseline characteristics of groups as analysed for the schools randomised to paper-based or control

School level (categorical)	Paper-based		Control		
	n/N (missing)	Count (%)	n/N (missing)	Count (%)	
School type					
Academy: converter mainstream	13/53 (0)	13 (24.5)	13/53 (0)	13 (24.5)	
Academy: sponsor led mainstream	5/53 (0)	5 (9.4)	6/53 (0)	6 (11.3)	
Community school	25/53 (0)	25 (47.2)	24/53 (0)	24 (45.3)	
Foundation/ voluntary aided/ voluntary controlled	10/53 (0)	10 (18.9)	10/53 (0)	10 (18.9)	
Ofsted rating					
Outstanding	8/53 (0)	8 (15.1)	8/53 (0)	8 (15.1)	
Good	31/53 (0)	31 (58.5)	31/53 (0)	31 (58.5)	
Requires improvement	7/53 (0)	7 (13.2)	5/53 (0)	5 (9.4)	
Inadequate	0/53 (0)	0 (0)	0/53 (0)	0 (0)	
No Ofsted	7/53 (0)	7 (13.2)	9/53 (0)	9 (17.0)	
School setting					
Urban	39/53 (2)	39 (73.6)	43/53 (1)	43 (81.1)	

Rural	12/53 (2)	12 (22.6)	9/53 (1)	9 (17.0)	
ICT facilities (broad summaries given for paper-based arm to avoid disclosure)					
Yes	>50 (0)	>50	48/53 (0)	48 (90.6)	
No	≤3 (0)	≤3	5/53 (0)	5 (9.4)	
Location					
East Midlands	10/53 (0)	10 (18.9)	12/53 (0)	12 (22.6)	
Manchester	15/53 (0)	15 (28.3)	13/53 (0)	13 (24.5)	
Newcastle	14/53 (0)	14 (26.4)	14/53 (0)	14 (26.4)	
Teesside	5/53 (0)	5 (9.4)	5/53 (0)	5 (9.4)	
West Midlands	9/53 (0)	9 (17.0)	9/53 (0)	9 (17.0)	
School level (continuous)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	
School size	>50 (<3)	297.5 (139.6) 249 (211, 416)	>50 (<3)	349.6 (200.4) 333 (214, 425)	
% eligible for FSM	>50 (<3)	31.7 (18.0) 29.8 (15.9, 44.8)	>50 (<3)	31.7 (19.2) 28.9 (17.8, 46.05)	
% pupils whose first language is not English	>50 (<3)	17.1 (25.4) 4.1 (1.8, 18.2)	>50 (<3)	10.2 (14.4) 3.7 (1.7, 12.1)	
% SEN or EHC	>50 (<3)	1.0 (1.3) 0.8 (0.2, 1.3)	>50 (<3)	1.1 (1.1) 0.9 (0.3, 1.55)	
Pupil level (categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)	
Sex, male	612/1172 (0)	612 (52.2)	648/1229 (0)	648 (52.7)	
Ever FSM					
Yes	312/1172 (0)	312 (26.6)	292/1229 (0)	292 (23.8)	
No	860/1172 (0)	860 (73.4)	937/1229 (0)	937 (76.2)	

EYFS GLD Yes No	 831/1172 (0) 341/1172 (0)	 831 (70.9) 341 (29.1)	 909/1229 (0) 320/1229 (0)	 909 (74.0) 320 (26.0)	
Pupil level (continuous)	n (missing)	Mean (SD) Median (IQR)	n (missing)	Mean (SD) Median (IQR)	Hedges' g effect sizes (95% CI)
PiRA raw-score	1172 (0)	11.9 (6.8) 12 (6, 17)	1229 (0)	12.5 (6.6) 13 (7, 18)	Paper-based vs control: -0.05 (-0.27, 0.16)
PiRA age-standardised score	1172 (0)	99.6 (13.9) 101 (90, 109)	1229 (0)	100.8 (13.3) 102 (92, 110)	Paper-based vs control: -0.06 (-0.16, 0.27)

Outcomes and analysis

Primary analysis

In total, 1,494 pupils (37.3%) achieved an above average age-standardised PiRA score at post-testing (>100, relative to PiRA's national standardisation sample; ICT: n = 415, 35.8%; paper-based: n = 531, 36.3%; control: n = 548, 39.6%). The correlation between the PiRA pre-test and post-test scores was 0.63 (95% CI: 0.61 to 0.65). The ICC for the pre-test PiRA age-standardised score was 0.23 (95% CI: 0.19 to 0.28).

Table 19 gives the results of the primary analysis. There was evidence of a difference between the ICT and paper-based arm (Hedges' g: -0.14, 95% CI: -0.28 to 0.00, p = 0.04) and between the paper-based and control arms (Hedges' g: 0.15, 95% CI: 0.01 to 0.30, p = 0.04), both in favour of the paper-based arm. There was no evidence of a difference between the ICT and control arms (Hedges' g: 0.00, 95% CI: -0.14 to 0.14, p = 0.99).

The R-squared from model 1 was 45.7%, and from model 2 was 46.8%.

Table 19: Primary outcome analysis results

	Raw means				Effect size		
Outcome	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Total n	Hedges' g (95% CI)	p-value
	ICT		Control				
PiRA	1020 (138)	99.8 (11.2)	1132 (132)	100.9 (10.9)	ICT 998; control 1125	0.00 (-0.14, 0.14)	0.99
	Paper-based		Control				
PiRA	1179 (285)	101.0 (11.1)	1236 (149)	100.8 (10.9)	Paper-based 1172; control 1229	0.15 (0.01, 0.30)	0.04
	ICT		Paper-based				
PiRA	1020	99.8	1159	101.1	ICT 998;	-0.14	0.04

	Raw means				Effect size		
Outcome	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Total n	Hedges' g (95% CI)	p-value
	ICT		Control				
	(138)	(11.2)	(258)	(11.1)	paper-based 1152	(-0.28, 0.00)	

Sensitivity analyses

Sensitivity analysis of primary outcome to results of pupils who completed the pre-test after their school was informed of their trial allocation

Table 20 summarises descriptively the age standardised PiRA pre-test score stratified by whether the school of the pupil taking the pre-test was informed of their trial allocation before or after pre-testing. There was no visible difference in age standardised PiRA pre-test score between these two groups.

Table 20: Age-standardised PiRA pre-test score summarised descriptively by whether the pupil's school was informed of the trial allocation before or after the pupil took the pre-test.

	Informed before pre-test (n = 810)	Informed after pre-test (n = 3029)
Age-standardised PiRA pre-test score		
n (%)	810 (100)	3028 (99.9)
Mean (SD)	99.8 (13.3)	99.6 (13.6)
Median (IQR)	101 (91, 109)	100 (90, 109)

For schools with ICT facilities, 3,029 pupils (78.9%) were from schools that completed pre-testing before being informed of their trial allocation (ICT: n = 972, 83.9%; paper-based: n = 1177, 83.1%; control: n = 880, 69.6%). For schools randomised to paper-based or control, 2,197 pupils (77.1%) were from schools that completed pre-testing before being informed of their trial allocation (paper-based: 1224, 83.6%; control 973, 70.3%). When the pupils who completed pre-testing after their school had been informed of their trial allocation were removed from the primary analysis models, there was very little difference in the effect sizes or 95% confidence intervals (confidence intervals were marginally wider) when compared to the primary results (Table 21). However, the p-value for the ICT versus paper-based comparison increased from 0.04 to 0.07 and for the paper-based versus control comparison from 0.04 to 0.20. This is likely due to the reduced sample size leading to a loss of power. There was very little difference in PiRA pre-test scores between pupils who were pre-tested before or after their school was informed of their trial allocation (mean (SD) 99.8 (13.3) and 99.6 (13.6), respectively) so differences are less likely to be due to an imbalance in baseline scores, but this can be a risk when pre-testing occurs after randomisation.

Table 21: Information on the primary analysis repeated in pupils who did their pre-test before their school was informed of its allocation

	Raw means				Effect size		
Outcome	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Total n	Hedges' g (95% CI)	p-value
	ICT		Control				
PiRA	869 (103)	100.0 (11.2)	788 (92)	100.9 (11.2)	ICT 850; control 783	-0.02 (-0.17, 0.14)	0.83
	Paper-based		Control				
PiRA	962 (262)	100.6 (11.1)	866 (107)	100.9 (11.3)	Paper-based 958; control 861	0.11 (-0.06, 0.27)	0.20
	ICT		Paper-based				

	Raw means				Effect size		
Outcome	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Total n	Hedges' g (95% CI)	p-value
	ICT		Control				
PiRA	869 (103)	100.0 (11.2)	942 (235)	100.7 (11.1)	ICT 850; paper-based 938	-0.13 (-0.28, 0.01)	0.07

Sensitivity analysis of primary outcome analysis to replacing pre-test PiRA covariate with EYFSP result in G09 for literacy (reading) score

When the primary analysis models were repeated replacing the age-standardised pre-test PiRA score with the EYFSP result in G09 for literacy (reading) score as a covariate (Table 22), there was very little change between the results found here and those found in the primary analysis for the ICT versus control comparison (Hedges' g: 0.01, 95% CI: -0.15 to 0.14, p = 0.93) and the paper-based versus control comparison (Hedges' g: 0.14, 95% CI: 0 to 0.28, p = 0.05). Similarly, results are similar for the ICT and paper-based comparison, but the confidence interval is a little wider and the p-value higher (p = 0.12).

Table 22: Information on the sensitivity analysis of the primary outcome replacing pre-test PiRA covariate with EYFSP result in G09 for literacy (reading) score.

	Effect size		
Outcome	Total n	Hedges' g (95% CI)	p-value
ICT vs control	ICT 995; control 1123	0.01 (-0.15, 0.14)	0.93
Paper-based vs control	Paper-based 1172; control 1227	0.14 (0.00, 0.28)	0.05
ICT vs paper-based	ICT 995; paper-based 1152	-0.12 (-0.28, 0.03)	0.12

Selection of pupils to receive programme

Of the 157 randomised schools that carried out pre-testing, 33 (21.0%) requested some of their pupils be deselected from the programme after taking the pre-test, of which 28 (84.8%) underwent random deselection. In total, 680 (17.0%) pupils were deselected, of which 548 (80.6%) were randomly deselected.

Table 123 gives information on the sensitivity analysis repeating the primary analysis excluding the 548 pupils who were randomly deselected. The results of the primary analysis were robust to the change tested in this sensitivity analysis.

Table 1: Information on the sensitivity analysis repeating the primary analysis excluding pupils who were randomly deselected from the programme.

	Raw means				Effect size		
Outcome	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Total n	Hedges' g (95% CI)	p-value
	ICT		Control				
PiRA	707 (105)	99.9 (11.1)	1132 (132)	100.9 (10.9)	ICT 690; control 1125	-0.01 (-0.15, 0.13)	0.87
	Paper-based		Control				

	Raw means				Effect size		
Outcome	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Total n	Hedges' g (95% CI)	p-value
PiRA	1020 (242)	101.4 (11.1)	1236 (149)	100.8 (10.9)	Paper-based 1013; control 1229	0.17 (0.02, 0.31)	0.03
	ICT		Paper-based				
PiRA	707 (105)	99.9 (11.1)	1000 (215)	101.4 (11.1)	ICT 690; paper-based 993	-0.17 (-0.31, -0.03)	0.02

Small group teaching

Information on small group teaching was provided by 19 schools (12.1%) for 70 pupils (1.7%). Due to the low return of data on small group teaching, the planned sensitivity analysis looking at the impact of small group teaching was not carried out.

Analysis in the presence of non-compliance

Of the 2,622 pupils allocated to either of the programme groups who did a pre-test, registry data was received for 1,292 (49.3%). Descriptive summarises of the number of sessions completed in each group are summarised in Table 4. The median number of sessions completed was slightly higher in the paper-based group compared to the ICT group.

Table 24: Number of programme sessions completed summarised descriptively by group

	ICT (n = 1158)	**Paper-based (n = 1464)**
Number of sessions completed		
Number with data (%)	574 (49.7)	719 (49.0)
Mean (SD)	66.6 (18.4)	72.1 (11.1)
Median (IQR)	75 (63, 78)	76 (72, 79)

Due to the high percentage of registry data not received, it was decided it was not appropriate to carry out a formal CACE analysis. We did not specify in the published protocol or SAP how we would deal with missing registry data or what proportion of data was required in order for the analysis to be carried out; however, the amount of missing registry data here is large (approximately half of those allocated to the programme groups). Therefore, it did not seem reasonable to carry out the CACE analysis in this instance since all schools with missing data would have been considered as non-compliers by default, which is unlikely to be true, and this would lead to unreliable CACE analysis results.

Missing data analysis

The following baseline variables were found to be predictive of whether the post-test PiRA score was missing:

- the percentage of FSM children in the year—adjusted odds ratio (AOR): 1.02, 95% CI: 1.01 to 1.03, $p < 0.01$, n = 3768;
- whether the pupil had obtained a good level of development on the EYFSP—AOR: 0.48, 95% CI: 0.35 to 0.64, $p < 0.01$, n = 3768; and
- the pupil's age-standardised PiRA pre-test score—AOR: 0.98, 95% CI: 0.97 to 0.99, $p < 0.01$, n = 3768.

However, the size of the AOR was non-negligible only for whether the pupil had obtained a good level of development, with pupils with a good level of development being 0.48 times more likely to not provide a post-test PiRA score. Treatment allocation was not predictive of whether the post-test PiRA score was missing (ICT versus control AOR: 0.81, 95% CI: 0.50 to 1.32; p = 0.40; paper-based versus control AOR: 0.93, 95% CI: 0.56 to 1.55, p = 0.78).

Repeating the primary analyses using multiple imputation by chained equations, results were as follows:

- ICT versus control—0.00, 95% CI: -0.15 to 0.13, p = 0.89;
- paper-based versus control—0.15, 95% CI: 0.00 to 0.29, p = 0.05; and
- ICT versus paper-based— -0.15, 95% CI: -0.29 to -0.02; p = 0.03.

In general, the results of the primary analysis models were unchanged when missing data was imputed.

EverFSM subgroup analysis

There was no evidence of an interaction between EverFSM status and randomised treatment for the schools with ICT facilities (ICT #EverFSM 0.08, 95% CI: -0.05 to 0.21, p = 0.22; paper-based #EverFSM -0.05, 95% CI: -0.17 to 0.07, p = 0.41). The total number of pupils included in this analysis was 3,275.

There was also no evidence of an interaction between EverFSM status and randomised treatment in the schools randomised to paper-based or control (Paper-based#EverFSM -0.05, 95% CI: -0.16 to 0.06, p=0.37). The total number of pupils included in this analysis was 2401.

Table5 gives the results of the primary analysis repeated in students ever eligible for FSM. There was evidence that, within the FSM subgroup, those allocated to the ICT group did not make any additional progress in reading compared to the control group (Hedges' g: 0.03, 95% CI: -0.16 to 0.22, p = 0.75). There was evidence that those allocated to the paper-based group made two months' additional progress relative to the control group and one month of progress relative to the ICT group, but the 95% CI for both effect sizes included a possible negative effect.

Table 25: Results of the primary analysis repeated in pupils eligible for FSM

	Effect size		
Outcome	Total n	Hedges' g (95% CI)	p-value
ICT vs control	ICT 263; control 256	0.03 (-0.16, 0.22)	0.75
Paper-based vs control	Paper-based 312; control 292	0.12 (-0.05, 0.29)	0.16
ICT vs paper-based	ICT 263; paper-based 310	-0.09 (-0.26, 0.09)	0.33

Post-hoc sensitivity analysis assessing the impact of pupils whose PiRA post-test was facilitated and marked unblinded to group allocation on the primary analysis models

There were 27 pupils (ICT 18; control 9) from two schools whose PiRA post-tests were facilitated and marked unblinded to allocation. The mean of the unblinded scores was higher than the mean of the blinded scores in both the 18 pupils in the ICT group (mean 108.1; SD 9.5) and the nine pupils in the control group (mean 113.7; SD 13.7). Table 26 gives the results of the sensitivity analysis repeating the primary analysis with the inclusion of these pupils whose PiRA post-test was facilitated and marked unblinded to allocation. The effect sizes and confidence intervals were not materially changed in this sensitivity analysis to the primary analysis, but the confidence intervals for the ICT versus paper-based comparison and for the paper-based versus control group comparison extended into the negative range.

Table 26: Results of the primary analysis repeated with the inclusion of pupils whose post-test PiRA was marked unblinded by a teacher at their school

	Effect size		
Outcome	Total n	Hedges' g (95% CI)	p-value
ICT vs control	ICT 1016; control 1133	-0.02 (-0.16, 0.13)	0.81
Paper-based vs control	Paper-based	0.12	0.12

	1172; control 1237	(-0.03, 0.28)	
ICT vs paper-based	ICT 1016; paper-based 1152	-0.13 (-0.26, 0.01)	0.07

Secondary outcome analysis

Table27 gives the results of the analysis of the secondary outcomes: ability to read exception, regular, and non-words as measured by the DTWRP, ability to sound out single letters and letter combinations as measured by the LeST, and reading attitude as measured by the RAQ. There was evidence that pupils allocated to either the ICT or paper-based groups made between one- and two-months' additional progress, relative to the control group, on both the DTWRP and LeST. For the RAQ, there was no difference between the ICT and control groups, but pupils in the paper-based group made two months' additional progress relative to control. However, for all these results (except the DTWRP between the paper-based and control groups) the 95% CIs extend into the negative range.

Relative to the ICT group, the paper-based group made one to two months' additional progress on the DTWRP and RAQ, but no progress on the LeST. For all these results the 95% CIs extend into the negative range.

The R-squared for 'model 1' analysing the DTWRP, LeST, and RAS were 40.3%, 28.6%, and 3.9% respectively. The R-squared for 'model 2' analysing the DTWRP, LeST and RAS were 40.7%, 28.5% and 4.0% respectively.

Table 27: Secondary outcome analysis results

	Raw means				Effect size		
Outcome	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Total n	Hedges' g (95% CI)	p-value
	ICT		Control				
DTWRP	405 (31)	108.7 (15.3)	446 (33)	107.9 (14.8)	ICT, 395; control, 441	0.09 (-0.07, 0.24)	0.27
LeST	404 (32)	1.43 (0.89)	446 (33)	1.32 (0.93)	ICT, 394; control, 441	0.15 (-0.01, 0.32)	0.06
RAQ	358 (78)	71.4 (9.4)	378 (101)	71.8 (8.9)	ICT, 350; control, 373	-0.01 (-0.15, 0.14)	0.94
	Paper-based		Control				
DTWRP	484 (24)	109.7 (15.3)	490 (35)	108.2 (15.0)	Paper-based, 480; control, 485	0.17 (0.05, 0.30)	0.01
LeST	484 (24)	1.39 (0.93)	490 (35)	1.33 (0.93)	Paper-based, 480; control, 485	0.13 (-0.02, 0.27)	0.09
RAQ	425 (83)	72.4 (9.0)	421 (104)	71.6 (9.0)	Paper-based, 421; control, 416	0.11 (-0.05, 0.27)	0.19
	ICT		Paper-based				
DTWRP	405 (31)	108.7 (15.3)	465 (24)	109.8 (15.3)	ICT, 395; paper-based, 461	-0.09 (-0.24, 0.06)	0.24
LeST	405 (31)	1.43 (0.89)	465 (24)	1.41 (0.92)	ICT, 395; paper-based, 461	0.01 (-0.14, 0.18)	0.84
RAQ	358 (78)	71.4 (9.4)	406 (83)	72.6 (9.0)	ICT, 350; paper-based, 402	-0.10 (-0.27, 0.07)	0.26

Secondary outcome sensitivity analyses

Sensitivity analysis of secondary outcomes excluding pre-test PiRA score covariate

Table28 gives information on the secondary outcome analyses repeated with the exclusion of the pre-test PiRA as a covariate. The interpretation of the secondary outcome analyses were unchanged by this sensitivity analysis.

Table 28: Secondary outcome analysis results repeated with the exclusion of the pre-test PiRA score as a covariate

	Effect size		
Outcome	Total n	Hedges' g (95% CI)	p-value
ICT vs control			
DTWRP	ICT, 395; control, 441	0.07(-0.08, 0.23)	0.37
LeST	ICT, 395; control, 441	0.14(-0.01, 0.31)	0.08
RAQ	ICT, 350; control, 373	0.00(-0.15, 0.15)	0.96
Paper-based vs control			
DTWRP	Paper-based, 480; control, 485	0.16(0.02, 0.29)	0.02
LeST	Paper-based, 480; control, 485	0.12(-0.03, 0.27)	0.14
RAQ	Paper-based, 421; control, 416	0.11(-0.05, 0.27)	0.18
ICT vs paper-based			
DTWRP	ICT, 395; paper-based, 461	-0.08 (-0.25, 0.08)	0.34
LeST	ICT, 395; paper-based, 461	0.02 (-0.14, 0.20)	0.79
RAQ	ICT, 350; paper-based, 402	-0.10 (-0.27, 0.07)	0.26

Sensitivity analysis of secondary outcomes replacing pre-test PiRA score covariate with the EYFSP result in G09 for literacy (reading) score

Table29 gives information on the secondary outcome analyses replacing the pre-test PiRA score covariate with the EYFSP result in G09 for literacy (reading) score. The interpretation of the secondary outcome analyses were unchanged by this sensitivity analysis.

Table 29: Secondary outcome analysis results repeated with replacing the pre-test PiRA score covariate with the EYFSP result in G09 for literacy (reading) score

	Effect size		
Outcome	Total n	Hedges' g (95% CI)	p-value
ICT vs control			
DTWRP	ICT, 395; control, 440	0.07 (-0.08, 0.22)	0.34
LeST	ICT, 395; control, 440	0.14 (-0.02, 0.30)	0.09
RAQ	ICT, 350; control, 372	0.01 (-0.16, 0.14)	0.92
Paper-based vs control			
DTWRP	Paper-based, 480; control, 484	0.18 (0.05, 0.30)	0.01
LeST	Paper-based, 480; control, 484	0.12 (-0.03, 0.27)	0.12
RAQ	Paper-based, 421; control, 415	0.11 (-0.05, 0.27)	0.19
ICT vs paper-based			
DTWRP	ICT, 395; paper-based, 461	-0.08 (-0.25, 0.08)	0.32
LeST	ICT, 395; paper-based, 461	0.02 (-0.15, 0.19)	0.84
RAQ	ICT, 350; paper-based, 402	-0.10 (-0.27, 0.07)	0.26

Implementation and process evaluation results

The implementation and process evaluation (IPE) was cross-sectional and longitudinal, encompassing all aspects of the programme. The purpose of the IPE was to explore the cascade model of training (train-the-trainers approach), assess the fidelity of the programmes, consider the views of various stakeholders, and identify the barriers and facilitators to successful implementation. We aimed to achieve this through a combination of data collection techniques including surveys, observational work, interviews, and focus groups. A full overview of the varied data collected in relation to each of the IPE objectives is presented in Tables 5 to 7 (Methods section).

We interviewed school staff and observed 25 sessions in 20 selected programme schools and six comparative teaching sessions in five control schools. We conducted focus groups with Year 1 pupils in receipt of the programme at six schools. We observed all stages of the training model and conducted surveys with regional trainers and school staff. We also interviewed the developers of the RUKS programme.

The findings from the process evaluation are presented in relation to specific research questions as detailed in the Methods section above.

Defining usual practice

RQ1 What does baseline practice in participating schools look like (control and programme) in terms of teaching or interventions targeted at improving literacy in Year 1?

Summary

- The use of ICT facilities was commonplace in Key Stage 1 teaching.
- Two-thirds of the TAs already delivered decoding or reading interventions.
- Schools typically employed a mix of group and individual approaches.
- TAs were present in literacy lessons in almost all schools, generally working with small groups of children on a regular basis.
- Year 1 pupils tended to be grouped by ability in literacy lessons.
- Year 1 classes receive around six hours of specific literacy lessons each week.
- Around half of the schools used small group teaching.

Data regarding usual practice in schools was collected primarily through surveys, completed at the start and end of the trial. These were delivered to both programme (ICT and paper-based) and control schools, though the questions varied slightly for the end of trial survey depending on allocation. Both surveys are presented in Appendices I and K.

All schools were asked to complete the baseline survey prior to randomisation with 501 members of staff from 150 schools providing data. This allowed us to establish a clear idea of existing practices and the schools' plans for programme delivery if randomised to receive either version of the programme.

Almost half the respondents were KS1 classroom teachers (46%) while a further 22% were TAs. Around a third (32%) were the reading support leads within their school. Just over half (52%) said they would personally attend the training and be responsible for programme delivery within their school if randomised to either of the programme arms. None had ever used the ABRA software or the paper-based version before.

Approximately three-quarters of schools anticipated that the programme would be delivered by Year 1 class teachers (77%) and/or Year 1 TAs (78%) if they were randomised to the programme. Individuals who identified themselves as likely deliverers of the programme reported considerable use of ICT, typically using it every day in their jobs (72%). The majority reported using ICT in a professional capacity for more than five years (71%). Seventy-two percent said they used iPads or tablets as part of KS1 teaching, 49% used laptops, and 48% used desktop PCs. Year 1 teachers reported a similar pattern for pupil usage, with 76% using iPads or tablets and just over a third using laptops and desktops. Most respondents felt very comfortable (35%) or comfortable (37%) using ICT for delivering teaching. Prior to implementation,

there was a spread of preferences in terms of how they would like to deliver the programme: 32% favoured ICT, 26% a paper-based approach, and 31% had no preference.

Two-thirds of the TAs (68%) already delivered decoding or reading programmes, with smaller proportions working with speech and language (19%) or other language and literacy programmes (15%). They delivered these mainly to Year 1 (60%) and to a lesser extent to Year 2 (33%) and reception classes (17%). They were employing a mix of group and individual approaches, most commonly working one to one (58%) or with groups of six to ten (40%). Around one in three (29%) reported working with groups of four or five, as encouraged in the programme guidance. Sixty-two percent of TAs were in a delivering or facilitating role, and almost half were involved in planning (46%) and/or assessing (46%).

Nearly all the KS1 teachers in the survey said they had TA support in their Year 1 literacy lessons either most or all of the time (72%) or some of the time (19%). Echoing findings from the TAs, this support took different forms: in 71% of cases the TA worked with a small group of children on a regular basis; in 38% of cases they provided one to one support. The majority of TAs were providing support as required in response to teacher direction (83%) and in 45% of cases they delivered planned, structured interventions with prepared materials. Most teachers reported that Year 1 pupils were grouped by ability in literacy lessons (60%) compared to 21% delivering to mixed-ability groups and 10% using both at different times.

Data collected through the post-implementation survey indicated that Year 1 classes receive around six hours of specific literacy lessons (that is, not within another subject) each week. The figure was higher in the paper-based arm (6.5 hours) than the ICT or control arms (5.8 and 5.9 hours respectively). Most teachers reported that literacy was usually delivered through whole-class teaching (93% overall), with around half also using small groups (50%). Some reported one to one delivery (22%), though this was more common in the control (28%) than in the ICT (16%) or paper-based (19%) schools.

An interactive whiteboard (IWB) was the most common type of technology used by teachers to support their literacy teaching (92% total), followed by tablets or iPads (44%). Whiteboards without interactive features (31%) and laptops (30%) were less common.

Programme fidelity

RQ2 To what extent do the schools and teachers implementing the programmes adhere to the intended model and dosage?

a) How effectively has the training provided to the trainers been cascaded to the school staff?

Summary

- A train-the-trainers model was used to cascade training from the developers down to school staff.
- The model worked well, regional trainers felt well prepared to deliver training, and schools felt well prepared to deliver the programme.
- Regional trainer training could be shortened.
- Regional trainers could be more proactive with checking in on schools in between visits.

Profile of regional trainers

The implementation team initially recruited eleven regional trainers through academic contacts and recommendations from the National Literacy Trust. Nearly all had a teaching or training qualification and previous experience in literacy training initiatives or programmes. Around half reported substantial experience in delivering literacy-focused CPD training. Most rated themselves highly on knowledge of phonics in the context of teaching beginning readers, but more detailed knowledge of the KS1 reading curriculum and specific interventions was patchy.

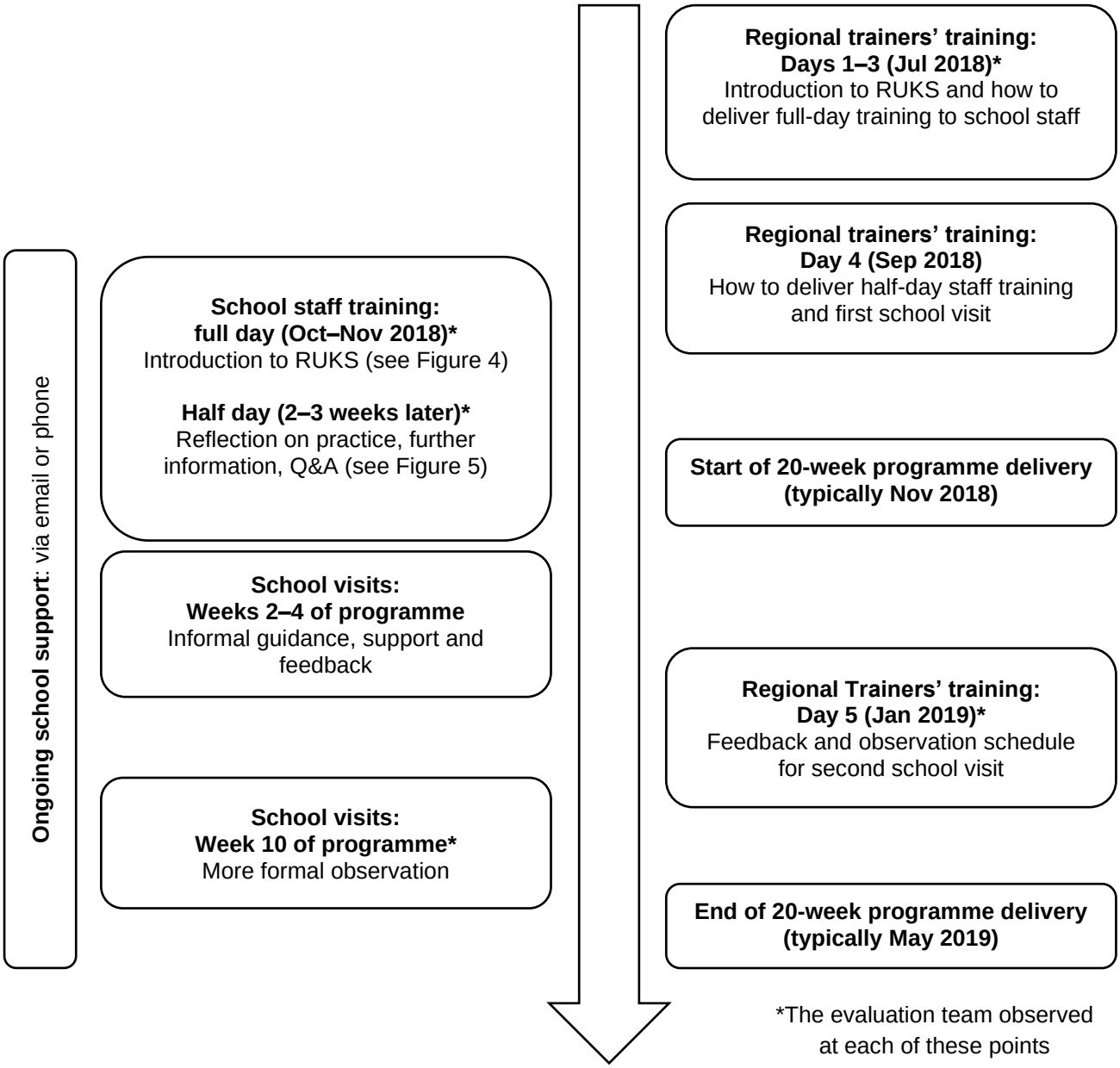
One had to drop out early in the process due to illness and the replacement was given two days of bespoke training to bring them up to speed. Another left during the project and their workload was shared out between some of the other

regional trainers. The regional trainers generally worked with five to nine schools, and all dealt with a mix of ICT and paper-based schools.

The training model

A 'train the trainers' or 'cascade' model was adopted, whereby regional trainers were recruited and trained by the implementation team who, in turn, were responsible for delivering continuing professional development training, scheduled visits, and just-in-time support to members of staff from schools randomised to deliver the programme (see Figure 3).

Figure 3: Timeline of training programme for regional trainers and school staff



Stage one: training of the regional trainers

The training of the regional trainers comprised one three-day course ('initial training') followed by two one-day courses ('follow-up training').

Initial training

The initial three-day training course took place in July 2018. Members of the evaluation team attended various training sessions to observe. The first and last days (days one and three) focused on going through the full-day training they would be delivering to school staff for the paper-based and ICT delivery models respectively (see Figure 4). Regional trainers observed a presentation with interactive segments providing opportunities for role play and exploring resources in smaller groups in a way which modelled how they were expected to train the school staff. Observers from the evaluation team (HA, LG, KW) noted that there was a strong emphasis on preserving the content but having some flexibility in delivery. The middle day (day two) focused on introducing the ABRA ICT software. This included an overview of the evidence base for the activities within the software and previous research on ABRA as well as a chance to practise navigating and using the software. Additionally, regional trainers also received guidance on keeping in touch with schools, providing just-in-time support via email and mobile phone, and the scheduled visits during Weeks 2 to 4 and in Week 10 (though these visits were covered fully on training days four and five). If a new member of school staff started delivering the programme, the regional trainers, rather than school colleagues, were expected to provide the necessary training. To our knowledge, this process was followed, and regional trainers visited four schools to provide training.

Potential challenges and concerns that school staff may face, as well as how to overcome them, were discussed on each day. Common issues and questions that had arisen from schools during the efficacy trial were highlighted and regional trainers were also given time to anticipate and discuss questions that school staff may have.

Three key potential issues that were highlighted in the efficacy trial were raised to enable trainers to better prepare schools for challenges they may face:

- first, understanding how to complete the progression sheets to monitor children's progress and ensure they were working at the appropriate level: it was recognised that school staff would need some support with this;
- second, concerns about the Canadian pronunciation in the ABRA software (used in the ICT arm) were highlighted: the guidance was to advise staff to provide the British pronunciation as appropriate (for example, to explain 'we say it like this'); and
- third, the likely challenge to staff of pacing activities and following the strict session plans was discussed; in addition, the importance of preparation time for school staff was emphasised and it was noted that schools would probably vary in the time allocated for this.

Regarding the ICT delivery model, regional trainers were asked to stress that children should actively use the ABRA software, with guidance from the facilitator, rather than being passive recipients of the programme. Strategies to encourage turn-taking and deal with dominant children who may want to takeover using the ABRA software were also discussed.

Regional trainers were given a guide for themselves and teachers, which linked the ABRA software to the U.K. curriculum and provided guidance on using ABRA.

There was some repetition between the second and third training days, which both focused on ICT delivery. Observers from the evaluation team (HA and KW) found that, at times, the pace was slow on the second day, so these days could possibly have been condensed into one.

Follow-up training

After a chance to practice, a further training day (day four) was held in September 2018, several weeks after the initial three-day course. During an interview with the implementation team, this session was described as 'a bit of a show and tell', designed to allow regional trainers the opportunity to raise questions. The session also provided guidance around

how to deliver the second half-day training for teachers (see Figure 5) as well as including further discussion of the progression rules, programme extensions, and the Week 2 to 4 observation visit to schools.

The fifth and final training day, held in January 2019, focused on trialling and calibrating the Week 10 observation schedule. Regional trainers watched videos of role-played programme sessions and used the sheet to evaluate them. They discussed their individual judgements in small groups before coming together to reach a consensus on scoring different scenarios. Some wording was subsequently changed for greater clarity. Regional trainers also topped-up knowledge of the rules governing progression to longer comprehension activities at Week 12 and discussed how school monitoring was proceeding.

Although a few regional trainers proactively contacted schools to check on progress, most appeared to wait to be approached. This was identified in the end-of-trial interview with the implementation team, with one interviewee saying:

> *'And I think that was the critical factor for me, the ones that I had confidence in were the ones who, when I saw them, when I spoke to them, they could talk to me about their schools and they clearly knew what was going on.'*

Stage two: training of school staff

The evaluation team carried out six visits to the school staff training sessions, observing four full day sessions (two ICT; two paper-based) and two half day sessions (one ICT; one paper-based). In total, five regional trainers were observed across four venues.

The content of the full day and half day sessions are shown in Figures 4 and 5. The presentation slides prepared by the implementation team for the ICT and paper-based delivery models followed the same structure, differing only where necessary to explain how to complete activities using the ABRA software versus paper resources, as applicable.

Figure 4: Overview of the full day school staff training for the ICT and paper-based delivery models

1	Brief overview of the project
2	Research evidence on learning and teaching word reading, comprehension, and reading fluency
3	Introduction to activities (paper-based or ICT activities, as appropriate)
4	20-week programme overview and getting to know the lesson plans
5	Introduction to the ABRA software (ICT delivery model training only)
6	Practising and exploring activities (role play delivering activities in small groups)
7	Progression rule, recording attendance (for evaluation purposes) and grouping children
8	Homework (review 20-week plan and resources; practise lessons with reception or Year 2 children

Figure 5: Overview of the half-day school staff training for the ICT and paper-based delivery models

1	Discussing how the homework went (reviewing the 20-week plan and resources; practising lessons with reception or Year 2 children)
2	Practising further activities to develop word reading, comprehension, and fluency

3	Overview of programme extensions and how to record these
4	Group-led time (for example, discussion, troubleshooting, review of materials, practising activities)

Overall, the training was of good quality and the content was generally well-delivered. Regional trainers stayed faithful to the presentation slides they had received. However, sessions tended to run faster than anticipated: the full day sessions finished between 30 minutes to two hours early and the half day sessions finished between 30 to 45 minutes early. In some sessions, there was a sense that school staff might have benefited from more time going over the lesson plans, further practice with the activities, or (in one case in particular) to have been given a more thorough introduction to the trial.

Some regional trainers modified how the progression rule for blending, word changing, and basic decoding activities was introduced (for example, by providing an additional handout or introducing this earlier than planned). Also, some regional trainers did not appear confident in their explanation and there was some inconsistency about whether the marking of an individual item as correct or incorrect was per pupil or as a judgement based on the majority of the group. Consequently, school staff were applying different rules. The evaluation team sought advice from the EEF and the implementation team subsequently provided clarification to all regional trainers (each item marked should correspond to one pupil's response to a unique word) and this was communicated to each school at the Week 2 to 4 observation visit. Most interviewees said these problems had been ironed out during the trainer's first, or sometimes second, visit. This will be described further later in this section.

In all sessions, school staff were given time to practise activities and ask questions, most of which were answered well, although sometimes regional trainers could have projected more confidence. Issues such as Canadian pronunciation and clashes with the school's overall phonics approach (for instance, the order in which skills are taught) arose during discussions.

To some extent, the regional trainers' background seemed to affect their delivery, with those who had worked in schools themselves being more aware of practical considerations, such as finding somewhere to deliver the sessions, and other regional trainers being stronger on the theoretical and research background. All regional trainers tended to adapt their communication to the group as appropriate, having established the background of attendees. The majority of school staff had arrived at the training knowing little about the programme or the trial and many were surprised by the intended intensity of the programme. Several were concerned about whether they would have the capacity to deliver it, especially to the cohort size their school had initially committed to.

Attendance at the training sessions was good. Most schools sent at least one staff member (often two or more) to the full day training (47/51 ICT and 55/57 paper-based), with slightly fewer attending the half-day follow-up (43/51 and 42/57 respectively). One school in each arm received training via a school visit instead. Other schools that did not attend the full day training had chosen to withdraw.

Feedback from training

Overall, the training provided through the 'train the trainers' model seemed to represent a good underpinning for implementing both versions of the programme in schools. Feedback forms were received for nine of the ten regional trainers that attended the initial three-day national training sessions (see Appendix P for an example form). A feedback form was also returned by a regional trainer who was unable to attend the training dates and instead received separate training from the delivery team. The feedback, both from regional trainers (nine of ten respondents after their three-day session) and school staff (223 respondents following the half-day session), was very positive. At least nine of the ten regional trainers attending the three-day course agreed or strongly agreed with each of the 17 positive statements about the training events (such as training goals had been met and they were satisfied with their increased understanding of the topic). For school staff, the equivalent figures ranged from 87% to 99%. No regional trainers agreed with the negative statement ('the training was too technical and difficult to understand') and only 9% of school staff did—mostly those undergoing training in the ICT programme (14% in the ICT arm; 2% in the paper-based arm).

Six of the eight regional trainers who completed the post-implementation survey (Appendix K) felt the training had prepared them very well for their role and five felt they were left with a very good understanding of the programme although two said they were left unsure about how the progression rules worked. Six regional trainers found the Week 10 observation form very easy to use with one reporting, 'The format was easy to follow and helped structure conversations.' All schools were provided with the contact details for their assigned regional trainer and were encouraged to contact them with any queries. Schools were also given a manual detailing all processes associated with programme delivery, including the progression rules.

The implementation team had also monitored regional trainers' delivery to school staff. Their only concerns were one trainer being too honest about their lack of background in reading theory or teaching, losing the confidence of some schools, and another who was so poorly prepared on their first training day that they had to keep checking for clarification. Illness led to the last-minute substitution of one regional trainer, but their replacement was given one to one training and the evaluation team observed their subsequent delivery to school staff and found it unproblematic.

Most regional trainers reported that they delivered the training without making any adjustments to how it had been presented to them, but two restructured the presentation and one re-wrote it. All regional trainers reported being very or fairly confident delivering to staff in paper-based programme schools and seven of the eight when delivering to staff in ICT programme schools. In two cases, regional trainers were unconvinced about the benefits of an ICT version and two regional trainers were initially apprehensive of possible IT issues.

In the post-implementation survey of school staff, 80% of those who had delivered the programme reported attending both days of training (86% of those in the ICT arm and 75% of those in the paper-based arm). Small percentages attended only one of the two sessions (8% total). Around 11% reported that they did not attend any training at all. More deliverers who were Year 1 teachers (92%), rather than other roles (75%), reported attending both training sessions.

Reflecting on the training at the post-implementation stage, over half the respondents thought it had prepared them 'extremely' or 'very' well (57%) to deliver the programme and another 25% said 'moderately' well. There was no discernible difference between ICT and paper-based training. This endorsed what emerged in the in-school interviews, with deliverers saying they appreciated the chance to practise (especially those attending the ICT version), role play, and to see demonstrations. Only one interviewee was dissatisfied, saying that the regional trainer did not seem very knowledgeable and the session was not well organised or informative with lots of confusion and questions.

Ongoing support in school

Post-implementation survey data collected from 142 of 157 schools (90%) indicated that, in general, school staff felt supported by their respective school, with 74% reporting that they received 'a great deal' or 'a lot' of support. The proportion of ICT and paper-based deliverers choosing these options was similar overall, although those in the ICT arm were more likely to select higher levels of support (45% versus 22%). More of the staff in 'other' roles (36%) reported being supported 'a great deal' than teachers (24%), perhaps reflecting support that TAs were receiving from the Year 1 teachers. Deliverers were also positive about how well their regional trainer had supported them to deliver the programme, with 62% saying 'extremely' or 'very' well.

Seven of the eight regional trainers responding to the survey had found it fairly or very easy to arrange visits to schools, with the remaining one finding it fairly difficult. They all described teachers and TAs as very responsive to their feedback and guidance. Five regional trainers reported that they were typically contacted proactively by staff at each school two or three times each term. The others had a lower level of contact, only once or twice per school across the implementation period. The most common reasons for schools being in touch were reported as clarification of how to use the progression tool (three regional trainers), IT issues (two regional trainers), and issues with differentiation and ensuring that children reached their potential (two regional trainers).

Although they had contact details for the regional trainers, few of the deliverers interviewed said they had used them with one deliverer remarking that they forgot that contacting trainers was an option: 'I did forget about that aspect actually.' However, one said they had sent several emails about problems and always got a response. Two interviewees would have welcomed more proactivity from the regional trainers, such as emailing to check how they were doing and prompting them to raise issues.

Instead, judging from the interviews, most stored up questions for the regional trainers' visits to the school. Two specifically said they were waiting for the forthcoming regional trainer visit to clarify an issue. This was also when errors, especially in following the progression rule and completing the record correctly, had been picked up by regional trainers. Consequently, regional trainers' visits were very valuable for advice and clarification, for instance about filling in the progression record, timings, tips on how to extend an activity, and help with IT issues such as providing a word list to use when the programme was downloading too slowly.

Although deliverers were generally satisfied with the regional trainers' visits, there were two negative comments both related to flexibility. One felt there had been a lack of recognition that the realities of school demanded adaptability. Another said that on the first visit, the regional trainer had been 'really strict' about delivering to the whole class in mixed ability groups but by the second visit:

> *'She completely changed her tune and … the fact that there's someone who's lower ability … you can progress, just sort of, not ignore that child, but don't mark them down' (paper-based teacher).*

This made the programme easier to use because they could move on groups whose progress had stagnated but left the deliverer confused and frustrated.

b) What variability in implementation exists across different participating settings? Are there any barriers or adaptations?

Summary

- Deliverers adhered relatively closely to the manual.
- Deviations were more likely in the non-ICT arm compared to the ICT arm.
- Deviations were most likely to be made to support struggling pupils.
- Sessions were most likely to be delivered by a TA in small groups.
- The small-group nature of the programme was particularly valued by both staff and children alike.
- Sessions were typically conducted in the afternoon, outside of the classroom environment.
- Schools found timings difficult to adhere to and felt the programme was too time consuming to deliver to the entire Year 1 cohort.
- Issues with ICT facilities were a key barrier for successful implementation of the ICT version of the programme. Four in five of survey respondents reported some level of ICT issues

Adaptations

Responses to the post-survey (n = 130 deliverers: 72 paper-based and 58 ICT) indicated that programme deliverers complied relatively closely with the guidance in the manual: 37% never deviated and 50% sometimes deviated compared with 10% deviating about half of the time or more often. Those in the paper-based arm were more likely to report deviating from the guidance (14% deviating at least half the time compared with 3% of the ICT arm). In the paper-based arm, changes related to meeting the needs of specific children or groups of children (such as additional explanation or simplification, altering activities, or providing extension activities for the more able) and constraints around timing (for example, running over time, extending discussions). For the ICT deliverers, the most common adaptations related to ICT not performing as required. This was reinforced in the interviews where two ICT deliverers held 'back up' paper resources, such as paper-based blending and decoding activities, in case the technology was slow on the animation slides.

When asked about the manual in the interviews, the feedback reflected these findings. A number of respondents likened it to a Bible or equivalent:

> *'We follow that like, gospel. And we always have that out next to us, because you can just easily go wrong with [the programme]' (paper-based Year 1 teacher).*

They embraced the scripted nature of the programme, with one TA saying that a teacher may choose to deviate from the plan, but she had stuck to it:

'I like it to be prescriptive, but to make little adjustments, like with time, but obviously that would be after this programme has finished … after this trial has finished' (ICT TA).

About half the interviewees reported lengthening or shortening components within a sessions:

'Because sometimes you'll get into a conversation, sort of with animated alphabet, you might get into a conversation of different words beginning with that letter … I manage to do it in that 15 minutes, but, yeah, sometimes I have to jiggle the timings on the different activities' (ICT LSA).

Two case study schools had tweaked the order of elements within the session, judging that this made it flow better for pupils, for instance, extending the story section to repeat previous pages as a reminder of what has happened or merging two separate sessions of decoding into one longer one.

Some schools had adapted peripheral aspects of the programme to fit their school system, for instance, incorporating good performance in sessions in the existing school reward scheme and using the school group or 'carpet' rules rather than ABRA STAR rules. Adopting the school rewards system may have contributed to the success of the programme by incentivising children to take part during the sessions. It is unlikely that using existing school rules rather than the STAR rules would have made an impact as these are likely to be similar in meaning.

Respondents to the survey expressed a range of views about the flexibility of the programme. More paper-based deliverers thought it was very/somewhat flexible (55%) than very/somewhat inflexible (21%). The pattern was very different for ICT deliverers, with 36% considering it very/somewhat flexible and 41% finding it very/somewhat inflexible.

Lack of time meant most interviewees had not done any extension activities. A few said they might in future, but there was a sense that this was said just to please the interviewer. One paper-based deliverer reported that she brings in pictures or uses her iPad to show pupils what something means. For instance, when doing the *How Beans Grow* book she discovered that her pupils knew nothing about plants, so she had brought a real one in:

'So they didn't know anything about a plant, leaf, or petal and I brought in a real plant to show its root … When the book shows the bean growing it shows the root but I showed them it's really below the ground, deep, then I scraped away the soil to show the root' (paper-based TA).

Logistics

In the post-implementation survey, 85% of teachers reported that the programme was delivered by a TA, with delivery by a Year 1 teacher reported less than half as often (40%). In a minority of cases, other staff were delivering sessions (for example, reading specialists).

Nearly all respondents (99%) said they delivered sessions in small groups: only one reported delivering to a whole class, and three in a 'combination of ways'. Across the lesson observations, the average group size was slightly smaller in the ICT schools than in paper-based schools (3.8 versus 4.2). In one focus group, children in an ICT school had occasionally merged with another group because of time issues. This was recognised to have benefits (being with friends, helping each other) and drawbacks (noisy, waiting longer for their turn). Three regional trainers responding to the survey identified group working as a positive feature of the programme.

Working in small groups was, on the whole, felt to be a real benefit for the pupils and one that they enjoyed. When regional trainers were asked what the programme's main successes had been, group work teaching was a commonly-mentioned element (one for the ICT version, three for paper-based). During the school visits, deliverers mentioned that it was an appealing novelty for pupils, giving them a rare opportunity to listen and contribute in a small group. Others talked about pupils helping, supporting, and encouraging each other. In two instances it had given the deliverer an opportunity to encourage contributions from those who were generally more difficult to engage or quieter in class:

'I think even for the quietest children, in a small group it helps them to interact and discuss because, like with the prediction, there's lots of talk that they have to do during a session and it does give them that confidence, especially if they're in a quiet place like this' (ICT teacher).

Most of the deliverers who responded to the post-survey judged that the groups worked together 'extremely' (23%) or 'very' (54%) well. There was little difference between the two trial arms, but teachers were slightly less likely than those in other roles to report that the groups worked extremely or very well together (69% and 79% respectively). The data cannot tell us whether this reflects an objective difference or a difference in perception or expectation.

For some respondents, having mixed ability groups had caused issues around balancing the differing needs of the children, for example, 'stronger readers became frustrated waiting for a turn' (ICT Year 1 teacher). Nine survey respondents reported grouping children into ability groups as a reason for the group working well together. The composition of the groups seen in the school visits varied, not just in terms of whether they were mixed or grouped by ability but also whether personalities and behaviour had been considered when forming the groups. Four of the deliverers who were interviewed (three paper-based, one ICT) talked about having to manage contributions to the group, for instance, encouraging quieter members and controlling the more dominant. One interviewee had, in their words, 'a very, very, very low ability' child who found it impossible to keep up, but she encouraged the others to help him and gave him tasks 'like pressing the little bird to listen to the sentences' (ICT Year 1 teacher). Clearly, the ABRA programme as implemented in this example is not meeting the child's needs but giving the child minor tasks allowed them to be included in the sessions.

Over half the respondents (62%) reported that sessions took place in a separate room, with an open area the next most common response (25%) followed by in the classroom with the rest of the class (15%). Although teachers and other staff were equally likely to say the sessions happened in a separate room, teachers were far more likely to report them being run in the main classroom (39% versus 8% for other staff). One of the implications might be that teachers could not always escape their classroom as they might wish when the session was due to be delivered, raising the potential for distraction from other pupils. Across the 130 respondents, 70% were extremely or somewhat satisfied with the space for delivering sessions.

The variation in venues was also observed on school visits. Some deliverers talked about how the lack of space meant they delivered in non-ideal conditions, such as the main classroom with all its distractions for pupils and deliverers. However, most sessions observed were carried out in quiet and calm surroundings, although it sometimes became apparent that the location had been changed specifically to improve conditions for the observer. For instance, one session was held in the library because the usual room was regularly disturbed by parents collecting coats; another normally shared with a second group.

The majority of ICT deliverers reported usually using an IWB (52%) or laptop (50%) for the sessions with only a minority saying iPad (12%). Some schools used a mixture, hence values do not sum to 100%. Most ICT deliverers found the software extremely (33%) or somewhat (41%) easy to use with only 3% (two respondents) saying it was difficult. Similarly, most of them judged that the children found it extremely or somewhat easy to use (28% and 47% respectively) with just four saying children found it difficult.

Altogether, 28% of respondents had not completed all the sessions within the 20- to 22-week target. This was higher among teachers (38%) than those in other roles (24%), and among ICT (36%) than paper-based (21%) deliverers. The main reasons for non-completion included scheduling difficulties and finding the staff time required to run it, as well as extracurricular activities (particularly around the lead-up to the Christmas holidays). The staff interviews confirmed that, where the programme was running behind, it was often due to Christmas activities (one TA had avoided this by using her own planning time to deliver instead) and, to a lesser extent, illness (their own or having to provide cover rather than delivering programme sessions) and school trips.

Timings

When asked how easy they found it to keep to session timings, two of the 13 deliverers interviewed said that it was enforced by strict school timetabling, but five interviewed (three ICT, two paper-based) admitted to struggling:

> *'That is the hardest thing, the timings, definitely. Fifteen minutes is not long enough. I prefer to have a bigger group and have a bit longer' (ICT teacher).*
>
> *'I think sometimes when we're reading a story we could do with a little bit more time and sometimes I am a little bit ... if I go 30 seconds over I'm, like, well, at least we've explored that bit a little bit more in-depth rather than just racing through, reading it and not talking about the words in it' (ICT TA).*

The majority of deliverers found it particularly challenging to squeeze everything into the time available, though two did report failing to fill the allotted slot on occasion. Several pointed out that the appropriate timing varied depending on the children in the session. Lower ability children and EAL pupils with limited English often needed more than 15 minutes, whereas faster readers could get through the content much more quickly.

There were numerous comments that three weeks was too long to spend on one story with some children becoming bored, and this was occasionally mentioned in the pupil focus groups as well. To a lesser extent, a similar criticism was levelled at the games and activities becoming tedious as they were repeated around the group. Deliverers were not, however, unaware of the potential benefits of repetition:

> *'But, in fact, one of the things that makes the scheme quite laborious is that you're constantly going over, and over, and over. Which, the children then get bored. But the other end of it is, that you need to go over, and over, and over, to make sure they've understood every phrase and they've got the vocabulary to grasp everything. … Swings and roundabouts, really' (paper-based teacher).*

One of the main themes elicited from the interviews was that the programme was too time-consuming with four sessions a week, especially for the schools that delivered as requested to all pupils in the class. They believed it would be more acceptable and manageable as an intervention for particular pupils. In some cases, this belief was exacerbated because deliverers felt that low ability pupils were holding others in the group back and two regional trainers suggested that the paper-based programme could have been improved by enabling better support, or faster progress, for more able children.

Issues with technology

Issues with technology was identified as a substantial issue faced by schools. When regional trainers completing the post-implementation survey were asked what challenges they had faced, there was a clear view that ICT schools had faced more difficulties than paper-based schools. ICT facilities (and finding space in which to use them) was the most common problem (mentioned by six out of the eight regional trainers responding):

> *'Internet issues or glitches within the programme … Good when they had an interactive whiteboard but some were sat around a laptop or even a mid-sized iPad.'*

According to responses garnered through the post-survey of school staff, problems with the technology had been encountered by 79% of deliverers in the ICT arm (44 of 56 respondents). Most of the problems (buffering, failure to load, elements not working, and slow speeds) could have been due to poor internet connections or low capacity. Although this is not an issue with the programme itself, the software has to operate in the context of schools' technological facilities. Sometimes problems were encountered on the evaluation team's school visits, including having to switch from an IWB to tablets when the technology failed.

Table 30: Technological issues with using the ICT-version of the programme (deliverer survey)

Technological issue	**Number of mentions**	**Example quote**
Buffering/freezing/poor connection	14	*It froze a lot especially in animated alphabet*
Programme/activities won't load	12	*arrows on alphabet sounds sometimes didn't load*
Errors in content	8	*Sometimes [pupils] would choose the correct answer but it would tell them that it wasn't correct*
Elements (e.g. sounds) not working	7	*Some of the icons didn't work*
Slow speed/lag	7	*Running slow, lagging, sometimes new words appeared over the top of old words rather than replacing them*
Incorrect routing in programme	3	*Sometimes we would swipe and end up on the previous page*

Likewise, in their interviews, there was widespread discontent among deliverers about the IT being slow or not working. Pupils occasionally mentioned this as well:

> *'And sometimes the whiteboard doesn't work and yesterday it wouldn't work so we had to go back to classroom, then Miss [X] turned it on, then it wasn't typing abracadabra, then we couldn't do that at all' (ICT pupil FG).*

Some deliverers found it challenging to secure equipment for their sessions or to read the small screens.

Some pupils in the ICT arm had been able to compare delivery on an IWB and a tablet because on occasion the IWB failed to work or was unavailable. Their preferences were mixed: some found an iPad easier to see and use, others said the same about the IWB. Others said they would prefer paper books because of issues seeing the board or screen.

c) How well have components of the programmes been delivered and how well have pupils and school staff engaged with it?

Summary

- Observed sessions in both arms had elements that were high quality but tended to struggle on pacing and timing.
- More adaptations were made in the non-ICT arm to cater to varying levels of pupil understanding.
- Deliverers in the non-ICT arm spent longer on average undertaking preparation activities than deliverers in the ICT arm.
- Some schools reported struggling with progression monitoring and were confused by progression rules.
- Pupil engagement was good, though pupils were viewed to engage more with the ICT programme than the non-ICT programme.

Delivery

ICT and paper-based lessons observed by the evaluators were rated almost identically, achieving on average around three-quarters of the maximum available score summed across the nine dimensions (see Appendix J). Three in five of the lessons were judged to be 'good' or 'excellent' overall. This proportion was the same for ICT and paper-based, although the paper-based lessons were more likely to be rated 'excellent' and the ICT ones 'good'.

The dimension receiving the highest ratings was preparation. As all observations were announced at least several days in advance, this is perhaps not surprising. Around three-quarters of the lessons were rated excellent for exposure to programme activities, and the instructional guidance given to pupils (for instance, explanations about what activity they were about to do, how to do it, and subsequent corrective feedback) was also judged favourably. All but one of the lessons were rated above 'adequate' for behaviour.

Group cohesion related to how effectively the facilitator encouraged pupils to work together, avoiding dominance or lack of contribution by enforcing turn-taking and sharing. In most cases this was not an issue but on average the paper-based sessions were rated slightly higher than ICT ones, two of which were judged poor. In one of these cases, pupils had no control of the tablet, there was no groupwork, and one or two group members dominated; in the other case, the deliverer did not encourage the group to engage with the activities and only one child spoke.

Paper-based lessons scored more highly on 'opportunities to succeed' than ICT lessons. This was often due to ICT deliverers simply letting the programme run rather than trying to intervene and target questions at specific pupils or including additional material to cater for the range of abilities. Paper-based deliverers showed a greater propensity to adapt the materials used to cater for all group members.

Neither the ICT nor paper-based sessions scored very highly for 'pacing and efficient use of time'. On average, the sessions lasted about 18 minutes rather than the intended 15 although the length varied from 13 to 24 minutes (similar between programme arms).

It proved difficult to assess the degree to which the programme and classroom literacy learning were integrated based on the observations, there being no evidence either way in most cases. More data surrounding this was generated through the survey as described later.

Preparation

On average, in the post-intervention survey, deliverers reported spending 25 minutes per week preparing for the sessions. This was higher for paper-based respondents (29 minutes) than ICT respondents (21 minutes). Teachers reported spending more time (37 minutes) than those in other roles (21 minutes) on preparation. Possible explanations for this include a higher proportion of the teachers delivered the paper-based version, which involved more preparation on average, teachers may also be supporting other staff to deliver sessions, they may have been including time as the project lead if that was their role, or the other staff (for example, TAs) may have had more experience preparing and delivering small group sessions so may be quicker at it.

The majority of ICT and paper-based deliverers (85% overall) said that they read the manual during preparation time. Understandably, those in the paper-based arm were much more likely than those in the ICT arm to have spent time creating resources (83% versus 9%) and photocopying (83% versus 16%). It is unclear from the data whether these were core resources or additional resources. Other tasks included setting up the room, the resources, and the technology as relevant.

The in-school interviews reinforced the finding that the ICT programme took less preparation than the paper-based version, although a couple found the technological set-up time-consuming:

> *'By the time you turn the laptop on and found out where you are and half the time, the laptop there will then die' (ICT Year 1 teacher).*

The paper-based interviewees found planning more onerous, although the reported preparation time varied and often proved difficult to estimate because it was fitted in around other duties. Five of the six paper-based deliverers mentioned preparing outside usual working hours (before or after school, during the dinner break) and one teacher prepared the next few weeks' sessions during her holidays. A few deliverers from both arms of the programme commented that preparation got quicker as they became more familiar with the programme, and if they were delivering sessions more than once.

Progression monitoring

Although the use and accuracy of completion of the progression records was part of the observation schedule, it proved a difficult dimension to rate. In many cases it was impossible to see how the sheet was being filled in because the observer did not want to disturb the session unduly so was often sitting too far from the deliverer to observe the detail, or the deliverer was completing it under their desk to hide it from the children. In other cases, the sheet was not completed until after the session. In both these situations the observer did not have adequate knowledge of individual children's names, or sufficiently fine-grained recall, to determine post-hoc whether the logs had been filled in accurately.

Although 65% of all deliverers in the survey claimed to find the progress records 'extremely' or 'somewhat' easy to complete, a large minority had struggled, with 18% overall saying it had been 'somewhat' or 'extremely' difficult. There was evidence that ICT deliverers (60% extremely/very easy, 22% extremely/somewhat difficult) struggled more than paper-based deliverers (70% and 14% respectively).

Responses in the interviews suggested that difficulties related primarily to failing to understand the rules for progression or how to fill in the record, and confusion over having different levels for the different activities (blending, decoding, and word changing):

> *'Because you're starting with one perception of … so it's like they started with blending and then it moved onto the different … and you weren't sure which week you were. So you think you get it and then you're like, they were like, we're on Level 3 and then you're like, "Oh, for this you're on Level 1." So they have to like … you just have to keep track of everything. Very brain consuming' (ICT Year 1 teacher).*

Most interviewees said these problems had been ironed out during the trainer's first (or sometimes second) visit. There was little evidence that staff had proactively sought help so it was hard to gauge how long inaccurate recording and therefore potentially inappropriate rates of progression had been continuing. Two teachers said they had provided

guidance to their TAs who had been struggling. Four of the 12 interviewees (one ICT, three paper-based) were not aware of any problems with the progression rules.

In the interviews, four deliverers (three ICT, one paper-based) mentioned struggling to complete the progression record in the session itself. They caught up straight afterwards or, in one case, later in the day because the turnaround between sessions was too tight.

Survey respondents were also asked how appropriate they found the progression rules in terms of how quickly the children moved up the levels. More than half (63%) of those in the paper-based arm reported that it was 'about right for most of them' compared to only 33% of those in the ICT arm. Where deliverers felt the speed of progress was inappropriate, this tended to be because they judged that children were being moved up too slowly (26%: 40% ICT, 15% paper-based) rather than too quickly (6%: 9% and 4% respectively). Deliverers in about half the interviews said they had issues with the speed of progress. Most of these felt that higher ability children were not being challenged sufficiently. One thought it was too hard for her pupils, many of whom were EAL children at an early stage of English language learning, and there was some concern about individual children who were falling behind the group. One ICT deliverer had used their own judgement to progress faster but was subsequently told by the regional trainer to stick to the rules while another, on raising the issue of it being too easy with their trainer, had been advised to adjust accordingly*:* 'someone came to watch us, and talk to us, and they advised, just skipping some weeks, to find some more challenge, which I did … so that made it a bit better' (paper-based teacher).

Deliverers expressed confidence in knowing when to move the children up. Reponses to the survey showed 79% felt 'extremely' or 'very' confident. The level of confidence was higher among teachers than those in other roles: they were more likely to opt for 'extremely' rather than 'very' confident (46% 'extremely', 35% 'very'), whereas the reverse was true for other staff (20% and 58% respectively).

Pupil engagement

Analysis of pupil interactions showed that pupils in the focus groups were more than twice as likely to be positive than negative about the programme (either version). However, children may have thought this was more acceptable to the interviewer (and to their deliverer, who often sat in on the groups). Several pupils (particularly in the ICT arm) identified specific games and activities within the programme that they liked. These covered a wide range from core features such as word changing, animated alphabet, and sequencing to more entertainment-related peripherals like the hockey game and pinata. A recency effect was apparent, with children more inclined to mention elements of the session they had just attended.

> *I: 'Why do you like [word changing activity]?'*
>
> *C: 'Because you get to like have a guess and choose a letter that's supposed to go in' (ICT Pupil FG).*
>
> *C: 'I enjoy it because it like helps you with words if you don't know some, and then you can just have a go at it and it doesn't matter if you get it wrong' (ICT pupil FG).*
>
> *I: 'Okay. Why do you like [sequencing activity]?'*
>
> *C: 'Because it helps you.'*
>
> *I: 'What does it help you do?'*
>
> *C: 'It helps your brain' (paper-based pupil FG).*

Half of those responding to the regional trainers' survey identified good pupil engagement in the ICT programme as its main success but none of them gave that response to the same question about the paper-based version. Similarly, more ICT deliverers mentioned that the pupils had been engaged by the ABRA software than their paper-based counterparts. It was seen as novel, providing a contrast to other schoolwork, with a heavy emphasis on interactivity. They also found the games, characters, and catchphrases engaging for the pupils:

> *'I think it's a different way for them to be able to do reading and to do those activities because there's a lot of paper-based reading that we do throughout the day in school with phonics lessons and just general reading of … doing different lessons' (ICT Year 1 teacher).*

In the interviews, only one paper-based deliverer mentioned any problems with engagement and that was a couple of pupils who were more difficult to engage more generally. Three ICT deliverers said children's attention could wander

during certain activities but they usually recovered focus when the activity changed. Only two mentioned that they actively used the group (STAR) rules and pupils at another school talked about how these rules had disappeared from the wall:

> *'We used to have them in the computer room, in the ICT room, but I don't know where they've gone now, because they're gone' (ICT pupil FG).*

Although there were no specific questions about behaviour in the survey, such issues were rarely mentioned in answer to other potentially relevant open-ended questioning, for instance about how well groups worked together and reasons for how likely they would be to use ABRA again.

One paper-based deliverer believed that her own efforts had transformed the 'daily grind' of the programme into something her pupils enjoyed, something she felt a less experienced member of staff would be unable to do:

> *'And to be honest, if it wasn't for me, really zhooshing it up, sort of, trying to make it, "Oh aren't we having fun, yeah." If it wasn't for me doing that, I don't think they'd enjoy it' (paper-based teacher).*

In a similar vein, the implementation team picked up informal feedback from schools suggesting that paper-based deliverers had been expecting something slicker:

> *R2: 'More jazz hands, and more sophisticated. Certainly, you know, something that was maybe more glossy and better produced.'*
>
> *R1: 'More commercial looking' (implementation team).*

Key Stage 1 literacy practices within evaluation schools

RQ3 During the intervention period what other practices do evaluation schools use that focus on improving literacy at KS1?

Summary

- Just under half of the schools in the control arm delivered literacy interventions to the pre-selected pupils; these were generally delivered by a TA.
- Just under half of the schools across the programme arms (ICT and non-ICT) reported that they had delivered other structured literacy interventions to Year 1 pupils during the evaluation period, mostly phonics or reading interventions.
- Many schools in all arms of the trial reported setting pupils by ability for literacy.
- Around a quarter of schools reported that pupils missed literacy-related lessons (such as guided reading) to attend their ABRA sessions.
- 66% of deliverers in the non-ICT arm and 53% in the ICT arm reported stopping other activities, for example, guided reading, phonics, reading, and handwriting, in order to deliver the RUKS programme.

a. What small group interventions have been used in control schools and/or did any compensatory activities occur?

Only about half (49%) the 65 survey respondents from the control group indicated that their school had, as requested at the start of the study, delivered a literacy intervention to four pre-selected pupils. Of these, 75% claimed to have delivered to a small group and 31% on a one to one basis. The interventions chosen mostly related to phonics or reading. In almost all cases, these interventions were delivered by a TA (94%) and a Year 1 teacher also delivered the intervention in nearly a third of cases (31%).

As context, the majority of respondents in control schools (85%) said they set pupils by ability for phonics teaching in their usual literacy teaching with Letters and Sounds (55%), Read Write Inc. (39%), and Jolly Phonics (20%) the most commonly used programmes.

b. Have programme schools used any other literacy-focused interventions, or small group teaching, in addition to the programmes?

Information concerning additional literacy-based programmes and interventions delivered in programme schools was collected as part of the post-implementation survey, which was completed by 87 Year 1 teachers. Just under half of these (47%) reported that they had delivered other structured literacy interventions to Year 1 pupils during the evaluation period, mostly phonics or reading interventions. Like the control group, most schools (80%) reported setting pupils for phonics teaching (ICT 86%; paper-based 75%). The main programmes used by the schools were also similar: Letters and Sounds (55%), Read Write Inc. (40%), and Jolly Phonics (9%). This pattern was reflected in the schools visited for the evaluation, where Letters and Sounds (or an adaptation) and Read Write Inc. were the phonics approaches being followed most often.

c. Have the programmes been delivered in addition to usual literacy teaching, or to what extent have the programmes been substituted for usual literacy teaching?

Sessions took place in afternoon lesson time (62% of survey responses) more often than morning lessons (35%) with 15% giving another response (figures do not sum to 100% because of multiple responses). These figures include 15% of the total sample who delivered in both morning and afternoon. Where another response had been given, staff were being opportunistic in carving out time:

'We often squeezed in a session at break or lunch time' (paper-based Year 1 teacher).

'During lunchtime to enable the use of an interactive whiteboard' (ICT Year 1 teacher).

'Initially in the afternoon but then when I had time to slot them in' (ICT Year 1 teacher).

Respondents were asked which lessons were missed to deliver the programme sessions (Table 31): there was a range of responses, with many staff naming more than one subject. The most common answers included literacy-related sessions (23% of total answers, particularly guided reading and handwriting), topic or foundation subjects (with at least 10% running sessions during PE), and the core subjects of science and (to a lesser extent) maths. We did not collect any data which could explore the unintended consequences of missed lesson time or missed play or social time.

Table 31: Lesson time missed to deliver programme sessions

Response	**Total answers (n = 216)**
Literacy-related	50 (23%)
Topic	35 (16%)
Foundation subject	26 (12%)
Science	21 (10%)
P. E.	21 (10%)
R. E.	17 (8%)
Maths	11 (5%)
Assembly	10 (5%)
Deliberately varied/range	9 (4%)
Breaktime	3 (1%)
Other	13 (6%)

Teachers were asked if they had needed to stop doing anything this year so that the RUKS programme could be delivered. Three out of five said they had. This was more often the case in the paper-based arm (66%) than the ICT

arm (53%). Elaborating, most of the 48 teachers who provided this information said they had reduced or ceased other elements of literacy teaching (including guided reading, phonics, reading, and handwriting; 16 teachers) and interventions (seven literacy-focused, three numeracy-focused, and the other 12 unspecified, from a total of 20 teachers who reported intervention delivery being affected). There was also mention of having to make lessons more focused on whole-class teaching or independent learning to compensate for pupils being removed.

Subgroups

RQ4 Are there perceived or actual benefits for specific groups of pupils (for example, SEN, EAL, GRT)?

Summary

- It was perceived that the programme may be more appropriate as a targeted intervention rather than a class-level programme, particularly for low- and medium-ability pupils.

Although there was a widespread sense that the programme would be better as a targeted intervention for a subgroup of Year 1 rather than the entire cohort, this tended to be for logistical reasons (time and staffing). However, deliverers perceived benefits to be greater for lower- and medium-ability pupils whereas higher ability children could be held back. There was no agreement about whether either programme was particularly beneficial for EAL pupils: on the positive side, deliverers reported it could help support vocabulary and rectify misconceptions, on the other hand, some deliverers felt those unable to join in with discussions or keep up with the fast pace had been excluded.

Groupings

RQ5 How did schools group pupils for small group work (programme and control schools)?

a. Does the way pupils are 'grouped' have any impact on the effectiveness of the programmes (ICT and paper-based delivery models)?

Summary

- Pupil characteristics were important for group cohesion.
- Some schools faced challenges when delivering to mixed ability groups.

Irrespective of trial arm, around half the schools reported delivering literacy in small groups at least half the time and about four in five Year 1 teachers set their children for phonics. Most schools delivered sessions, as intended, to small groups of children and both deliverers and pupils tended to like this way of working.

From the observations in the programme schools, the manner in which the group operated seemed to be related more to the individual participants' characteristics (for example, how quiet or disruptive they were) and the skill of the deliverer in managing their contributions rather than the school's overarching grouping strategy. However, some of those who delivered to mixed ability groupings reported issues around successfully managing a range of ability levels without constraining higher ability pupils or leaving lower ability children behind.

> *'So, you're sort of holding half of the group back, because of one child getting the three wrong every day … if it's the same child every day, then obviously we have to do a further intervention from that' (paper-based Year 1 teacher).*

This finding implies that it may be more beneficial to group pupils by ability to increase the likelihood of the children being able to move through the programme at the same pace.

Stakeholder views

RQ6 What are the views of specified stakeholders (school staff, pupils, developers) about the implementation and effectiveness of the programmes during the trial period?

Summary

- Most deliverers viewed the programme as equal or better quality to their usual literacy interventions (no difference between delivery models) and many would use it again.
- Deliverers generally felt the programme was beneficial for pupils, particularly with regard to developing their comprehension skills.
- Staff and pupils generally liked the stories, though there were some issues raised about the quality of the printed books in the non-ICT arm and the time spent on each story.
- The Canadian origins of the programme were particularly problematic in terms of the cultural context as well as anomalies in pronunciation, phonics, and punctuation.

Comparison with usual literacy practice

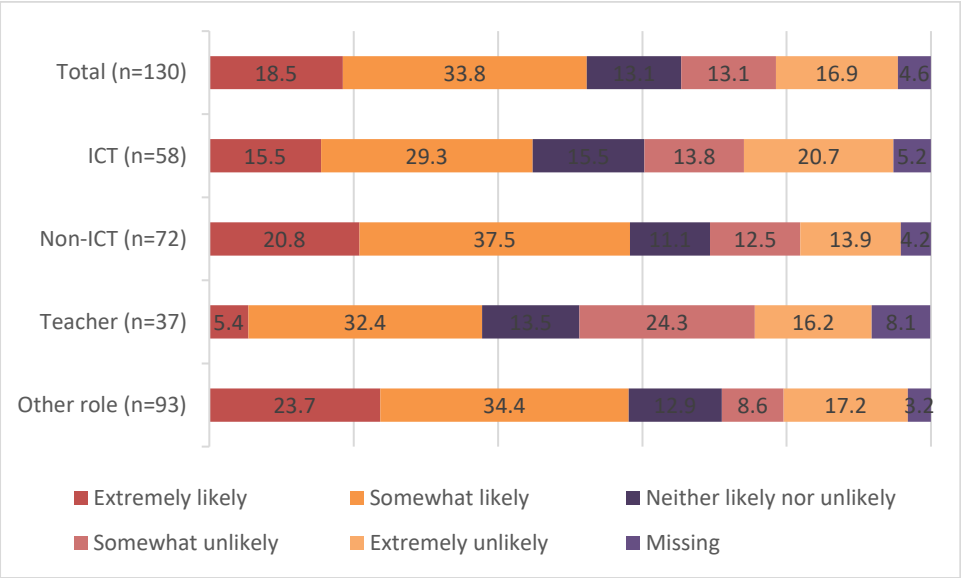
In the post-intervention survey, deliverers were asked to rate the programme in comparison with their usual literacy interventions. Responses were spread across a five-point scale (from 'much better' to 'much worse'). Overall, 37% thought it was better, 27% thought it was worse and 31% judged it to be about the same. Teachers (30% judging it better) were less positive than those in other roles (40% judging it better), but there was no meaningful difference between those delivering ICT (34% better) and paper-based (38% better).

During the observations of the staff training and during school visits, we noted that there were some comments about the programme not being tailored to the way phonics is usually taught in England. Examples cited included the pronunciation of certain individual vowels and the different way that some words are divided into sounds (phonemes) such as 'time' and 'take'. In British English these form split digraphs (that is, the two vowels are treated as one sound, the digraph, split by the m and k respectively) whereas in Canadian English the words are split into three consecutive sounds (t//i/me and t/a/ke).

One school, where the programme had replaced guided reading, had developed a system that enabled it to identify individuals' performance in their group to incorporate in the school system as evidence of individual progress for Ofsted.

All survey respondents were asked how likely they would be to use the programme again if it was their decision. Responses were spread across the five-point scale from 'extremely likely' to 'extremely unlikely' (Figure 6). Looking across the whole sample of respondents, including both programme deliverers and non-deliverers, more were likely (52%) than unlikely (30%) to use it again. Comparing subgroups, those in other roles expressed more enthusiasm than teachers (58% versus 38% likely to use again) and paper-based were more positive than ICT deliverers (58% versus 45% likely to use again).

Figure 6: Likelihood of using the RUKS programme again—deliverers and non-deliverers



The reasons given for being likely to use the programme again included learning and enjoyment on the part of both pupils and deliverers. However, even some of these more favourable survey respondents had experienced difficulties fitting sessions in and suggested that it might be better as a focused intervention rather than applied to all children (see also RQ4). Where more detail was provided using the free text box, respondents either explicitly or implicitly implied that the programme may be best suited as a targeted intervention to those with lower levels of literacy. Likewise, time was the key drawback for respondents not wanting to use the programme again: in general, time taken from the timetable and—within the programme—time devoted to individual books. Some felt there was unnecessary repetition inherent in the programme, which could adversely affect pupil's motivation.

All the Year 1 teachers, regardless of whether they delivered the programme or not, were asked how well they thought the programme integrated with their school's literacy policy (including the phonics programme). Only 31% thought the programme integrated extremely or very well; 16% said not well at all. The sample sizes were small, but there was an indication that some teachers in schools following the ICT programme judged it less integrated than those using the paper-based version.

Perceived outcomes for deliverers and pupils

In the post-intervention survey, 31% of all deliverers reported that they were 'extremely confident' in facilitating the programme sessions and 49% were 'very confident'. Confidence was higher among the paper-based deliverers (86% extremely or very confident) than the ICT deliverers (72%) suggesting that more support may be needed. There was limited evidence that the programme had improved their confidence in delivering small group work, with 21% saying it was somewhat or much better but most (69%) saying it was 'about the same'. This was true regardless of delivery arm or staff role. The overall picture was almost identical for deliverers' self-report of changes in their understanding of children's language development overall, with 67% reporting it remained 'about the same', and most of the remainder saying it was better to some degree. However, there was a suggestion of a difference between teachers (16% saying it was better and 73% about the same) and those in other roles (31% saying better and 65% about the same). Although based on small sample sizes, this may reflect TAs (who constitute the majority of those in other roles) having less initial knowledge and training in this area.

Several pupils considered that they had learnt something from the RUKS programme sessions, but there was no consensus what this was. It ranged from learning to read and spell to specific learning unrelated to literacy (for example, 'do not water a beansprout when it is raining'—a message gleaned from the book about growing beansprouts).

The Year 1 teachers who completed the survey (regardless of whether they delivered the programme or not) were asked how much they thought their pupils had improved across the year in various facets of literacy (speaking, listening,

writing, understanding, decoding, comprehension, attitudes to reading, and motivation to read). In interpreting these figures, it should be borne in mind that, although it was intended that the programme would be delivered to the entire Year 1 class, in some cases (as already discussed) this had not happened. Consequently, some teachers in the programme arms will be answering for a mix of children who did and did not attend the programme sessions, making the responses equivalent to 'intention to treat'.

For each dimension, teachers could answer that pupils had improved a great deal, a lot, a moderate amount, a little, or not at all. The most common response was that pupils had improved 'a lot' except for decoding (including phonics) where 55% overall thought their pupils had improved 'a great deal'. For most dimensions, there was no consistent difference across the three trial arms. Comparisons of the combined percentage saying pupils had improved a lot/a great deal show that the ICT pupils were rated higher for improvements in motivation to read (73% ICT versus 55% paper-based and 62% control). This may be related to the perceived levels of pupil engagement being higher among those receiving the ICT programme rather than the paper-based programme (see RQ2c). Paper-based pupils were judged to have improved less than their counterparts in the other trial arms for understanding (65% paper-based versus 82% ICT and 77% control).

When interviewed, deliverers mentioned several aspects of literacy where they felt the programme had impacted on pupils. Improvements in comprehension and the ability to answer and discuss questions were identified most frequently. Several felt that the impact had varied according to pupils' ability. The consensus was that there was not enough challenge for the high ability pupils, and the main gains were seen for those of low or average ability, hence the suggestion of a targeted intervention. There was no agreement about whether or not it was beneficial for EAL pupils: some felt it helped them with vocabulary and to address misconceptions about phrases while others felt they lacked sufficient language skills to enter the discussion or keep up with the programme given the strict timings.

Reading material

In the survey, programme deliverers (both ICT and paper-based) most commonly liked the stories because pupils enjoyed them, and they stimulated their interest and imagination (Table 32). Paper-based deliverers also talked about the visual appeal of the books, usually referring to the illustrations and limited amount of text per page. ICT deliverers were more likely than paper-based to refer to the range of stories, mainly in terms of subject matter and familiarity, and to the activities incorporated around them (for example, prediction and interactive vocabulary elements). There were also positive comments from both trial arms about most of the stories being easy to understand. The same themes were reflected in the school visits with ICT teachers in particular liking the stories (range of fiction and non-fiction and the inclusion of some traditional tales). One commented that it made a change from paper books. Another element that received favourable mention was the moral message of the stories:

> *'I've been quoting half the story today at break time because a child did something that they shouldn't have done. And I said, "What happened in* The Frog and the Well*?" So, it ends up as a really good discussion point' (ICT TA).*

In the focus groups, children mentioned specific books that they had read and had mixed views about how interesting or boring they were, and what they thought about the story content. Positives included the range, learning new information, and the humour:

> *'[The stories] help you read, and like if you're stuck, like stuck on your word, with the pen you can tap the word and you can spell it out' (ICT pupil FG).*
>
> *I: 'Which has been your favourite story that you've heard on the programme?'*
>
> *C: 'Little Red Hen.'*
>
> *I: 'Little Red Hen? And why's that?'*
>
> *C: 'Because I haven't learned about bread that they made in the olden days' (ICT Pupil FG).*

Table 32: Likes about stories and storybooks (deliverer survey)

Theme	Total (n = 95)	ICT (n = 41)	Paper-based (n = 54)	Example quote
Interest, imagination, enjoyment	31	17	15	*Engaged with children's imagination (paper-based).*
Visually pleasing	17	2	15	*The amount of print on the pages was not too overwhelming for the children. The illustrations were bright and colourful (paper-based).*
Understanding (including ease of)	15	6	9	*The stories were repetitive, which meant the children were able to remember the story when they came back to it in the next session' (paper-based).*
Range, variety, familiarity	15	12	3	*The stories were brilliant for the kids as there were some they knew and others they didn't (ICT).*
Games, activities, interactivity	11	8	3	*The words in the stories incorporated within the games. The stories allowed for children to learn new words, gave opportunities to focus on difficult words (ICT).*
Moral, point to story	9	7	2	*The children enjoyed the 'moral' aspect of the stories, liked that it was open for discussion (ICT).*
Vocabulary	6	4	2	*Good vocabulary to discuss (paper-based).*
Challenge, level of difficulty	2	1	1	*Some challenging language (ICT).*

n = number of respondents making a free text response.

Overall, about the same number of respondents furnished details of what they disliked as what they liked about the stories. However, paper-based deliverers gave many more, and more multi-faceted, responses about dislikes than ICT deliverers did (Table 33). Two or three different aspects were often combined in one answer:

> *'The quality of the books was not good. They were like pamphlets and the children did not like that. The illustrations were a bit amateurish' (paper-based teacher, survey).*
>
> *'The size of the font, the punctuation was not correct or recognisable to the children and really disliked words being broken up over two lines. Totally threw the children when trying to read them. Had to do a lot of explaining to the children when this happened, which was quite frequently' (paper-based teacher,survey).*

In terms of specific stories, although one deliverer praised the *Waterfall* book as culturally informative with a meaningful happy ending, four others identified it as an inappropriately difficult story for pupils to understand: 'The text *Waterfall*, extremely hard for the children to pronounce names which knocked confidence' (ICT TA survey).

In the case study interviews, paper-based deliverers had similar criticisms about the layout of the books, particularly the small font and compact words, which made choral reading while tracking with a finger very difficult. One deliverer had solved this by enlarging the storybook pages to A3 and putting them up on the wall. The typeface, long lines of text, and splitting words across lines were other features that made it difficult for children to access the books. One deliverer felt the text did not live up to the more exciting illustrations although, in contrast, another described the pictures as 'bland'.

One respondent in the regional trainer survey suggested that the quality of reading books, both in terms of print and story content, could be improved. This opinion was also voiced in a deliverer interview:

> *'And I just don't think the language is very good. There's something dull, I don't know what it is. I don't like the language. Whereas, if it was using a proper storybook, with lovely, rich story language … I think it's because it's a bit of a precis of a full book' (paper-based teacher).*

There was a mix of attitudes to reading among the pupils, which sometimes spilt over into their comments about the stories.

> *'I don't enjoy, like … I don't enjoy reading the books and I don't even read them every time. I don't like reading the books … reading is not fun' (ICT pupil FG).*

Table 33: Dislikes about stories and storybooks (deliverer survey)

Area of dislike	**Total (n = 87)**	**ICT (n = 35)**	**Paper-based (n = 52)**	**Example quote**
Presentation (not satisfactory, e.g., font too small)	26	0	26	*'The writing was very small and close together and some words started on one line and went into the next which our children are not used to' (paper-based).*
Children already familiar and/or losing interest	19	7	12	*'The children already knew the traditional stories and answering the comprehension and predicting was too easy' (ICT).*
Accent/pronunciation	18	18	na	*'Wish the voices could have an English/UK option as this did confuse the lower ability children' (ICT).*
Difficulty in progressing or pitched too high	11	1	10	*'Some very complex language. Lots of my groups were unable to read themselves' (ICT).*
Repetition	11	7	4	*'Very repetitive therefore some children became bored' (ICT).*
Errors including but not limited to grammar	10	0	10	*'Because they were written in Canada some of the language was different and the punctuation marks were confusing' (paper-based).*
Time/length	8	5	3	*'How long we focused on each book. The pupils got bored' (paper-based).*

n = number of respondents making a free text response.

Canadian origins

Pooled data collected from the surveys, interviews, and through observations indicated that the Canadian origins of the online software were problematic for many schools. In the post-implementation survey, 33% of respondents from the ICT arm (31 of 94) mentioned the Canadian accent or material as being alienating and confusing or something they would like to see changed in the future. In the case of the paper-based group, issues related mainly to the Canadian context of the stories, for example, Canadian animals. The Canadian context presented a much greater issue for schools in the ICT arm which, in addition to the differences in cultural context within the stories, also had to contend with difference in the accent and the phonics demonstrations. The issue of re-voicing with a British accent had been explored before the trial began but this was considered prohibitively expensive.

The implementation team found the cross-cultural differences to be three-fold. They felt that, while most deliverers were happy to adapt to the Canadian accent and talk over the computer at appropriate times, a minority resisted because they had been promised an intervention with minimal input from themselves:

R2: 'If [deliverers have] been on prior training where "thou shalt" has been built into them, and they have been taught to be empty vessels, I think those are the ones that struggled the most with needing to think about how to deliver ABRA in a way that's sensitive to the needs of their kids, and also sensitive to whatever the literacy model is that's being delivered at the schools' (implementation team).

Secondly, some anomalies in the phonics were also raised by ICT schools and the implementation team:

R1: 'For example, for a word like "cake", in the U.K., we'd split that up as in the c, and then the a would go with the e, and then you'd have the k … split digraph or magic e, however you want to describe it, but on ABRA the e goes with the final k. So you get c, a and then ke together, and all the teachers were going, "What? … What's this about?"' (implementation team).

'Some of the decoding was different—split digraphs not taught the same way, for example. Phonemes taught in different order and therefore children didn't know all of them before being asked to decode words' (ICT, learning mentor survey).

'We were not happy with the Canadian pronunciation on the programme when we are teaching different pronunciation for phonics screening' (ICT, Year 1 teacher survey).

'Unfortunately too many of the phonics were incorrect, which confused the children. They need to hear the correct sound consistently' (ICT, Year 1 teacher survey).

'And there's a lot of things that's maybe not tailored to our phonics programme in this country. Like the way they do the split diagraphs, the way the say "oh" for "o" [oh as in boat versus o as in cot]' (ICT TA).

Thirdly, the implementation team itself perceived a difference in school approach to early literacy between the two countries, judging that schools in England are more constrained and accountable, making them more reluctant to change their approach:

R1: 'In Canada, they're not so regimented. They've got much more flexibility how they teach reading … whereas in the U.K. we've got phonic screening, we've got the Key Stage 1 SATS, the league tables: everyone has to provide a systemised structured phonics approach to reading' (implementation team).

In the regional trainer survey, three respondents raised the Canadian pronunciation as a challenge for schools, either in terms of teacher resistance or children struggling with it. Asked if they would make any improvements to the ICT programme, half the regional trainers (four) suggested an English version of the programme.

The post-implementation survey responses from school staff matched the challenges identified by the implementation team and regional trainers with dissatisfaction focusing on cultural context as well as anomalies in pronunciation, phonics, and punctuation:

'If a U.K. English version was available, it would be much better as it used Canadian animals and places in Animated Alphabet *as well as Canadian pronunciations' (ICT TA survey).*

'The Canadian pronunciation of the sounds does not match our phonics' (ICT, TA survey).

'The American/Canadian accents were really off putting and for LA [low ability] children, it really confused their understanding of certain sounds' (ICT, Year 1 teacher survey).

'My only issue with ABRA was the Canadian accent as some of our children found it very hard to understand and would choose the wrong answer because of the pronunciation, not because they didn't know' (ICT, Year 1 teacher survey).

Cost

For the purposes of the trial, schools allocated to either of the programme arms only paid a one-off £200 training fee. Here we report costs associated with implementing each delivery model of the programme in a 'real-world' context.

We estimated the average cost per pupil per year for schools implementing either delivery model of the programme. Although completion of the report and subsequent publication has been delayed due to Covid-19, this report was

commissioned in 2017 and the evaluation was designed with the current EEF guidance in mind. Consequently, we have used the 2016 cost guidance (EEF, 2016) while also following the updated guidance issued in 2019 (EEF, 2019) where possible. The main aspect of cost missing relates to staff cover to facilitate implementation of the intervention (for example, if cover is needed for staff release to attend training) as we did not collect this. This means that the costs may represent a slight underestimate of the total, though this is likely to be small. Given the overall low cost it is highly unlikely that including staff cover would change its EEF cost rating based on current thresholds. In this trial, schools were initially asked to deliver the programmes to whole Year 1 classes. The cost evaluation has been prepared under the assumption that the programme would continue to be delivered at the class level assuming an average of 26 pupils per class (as per the average class size in this trial).

Table 34 details the resources needed to implement the programme. The primary sources of cost for implementing both versions of the programme are:

- training;
- materials; and
- staff time.

Table 34: List of resources

<table>
<tr><th rowspan="2">Category</th><th colspan="2">Item</th></tr>
<tr><th>ICT model</th><th>Paper-based model</th></tr>
<tr><td rowspan="3">Personnel for training</td><td colspan="2">Academics from the implementation team required for five full days for the training of regional trainers</td></tr>
<tr><td colspan="2">Ten regional trainers</td></tr>
<tr><td>School staff (school’s preference)</td><td>School staff (school’s preference)</td></tr>
<tr><td>Personnel for preparation and delivery</td><td>School staff (school’s preference)</td><td>School staff (school’s preference)</td></tr>
<tr><td>Programme costs</td><td>Programme manual</td><td>Programme manual, books</td></tr>
<tr><td>Facilities, equipment, and materials</td><td>ICT facilities: internet connection plus laptop or interactive whiteboard</td><td>Optional resources, for example, magnetic letters, letter cards</td></tr>
</table>

Note: there are no licence fees associated with the online version of the programme.

To deliver the ICT version of the programme, there are also some additional prerequisites in terms of ICT facilities. These could constitute start-up costs if schools decided to purchase these specifically to deliver the programme, however, the survey indicated that this was not the case and schools used existing facilities. There is no licence fee associated with the software hence the only programme cost for the ICT group relates to the cost of having a printed manual.

Time

Prior to implementation, staff time was required for training. At least one nominated member of staff from each school was required to attend training for a total duration of 1.5 days. Schools were asked to deliver sessions four times per week, at 15 minutes per session for 20 weeks, amounting to 20 hours. In the survey, schools were asked to estimate how much additional staff time per week—other than delivery time—was needed for each group receiving the programme. Paper-based deliverers reported an average of 29 minutes per week of additional work (ranging from no extra time to three hours and 20 minutes); for ICT delivery, it was 21 minutes (ranging from zero to three hours)—an average of 25 minutes overall. Sessions ran during a typical school day and no cover was needed for staff. No other time costs were involved. Preparation time included reading the manual and preparing resources.

Table 35. Total time devoted by school staff to deliver the programme over 20 weeks per year

	Year 1		Year 2		Year 3	
	Mean number of hours		Mean number of hours		Mean number of hours	
Activity	ICT	Paper-based	ICT	Paper-based	ICT	Paper-based
Training	9	9	0^	0^	0^	0^
Session delivery	20	20	20	20	20	20
Additional time to support delivery (e.g. preparation, set-up time)	7	9.7	3.5*	5*	3.5*	5*

^ Assumes the member of staff trained in Year 1 continues to deliver the programme over subsequent years.
* Assumes that the time needed depreciates in subsequent years with experience, a conservative estimate of a 50% reduction in time is assumed.

Financial costs—training

For the purposes of the trial a 'train the trainers' model was adopted whereby the implementation team trained regional trainers who in turn trained teachers. A similar model could be used to roll out the programme nationally.

Training of the regional trainers was delivered by two professorial-level academics and another member of the implementation team. The senior academics were funded by the EEF at a rate of £900 per day while the other member of the team was paid via their usual place of work at their day rate of £183.30. It is currently unclear who would be responsible for leading the training if the programme were to be rolled out. Regional trainers were funded by the EEF at £350 per day while receiving training and £500 per day, or £350 per half-day, for delivering training to schools.

The implementation team trained regional trainers for five days who in turn were responsible for training school staff.

School staff were trained in groups. A total of 13 full-day training sessions took place in addition to 15 half-day sessions. During implementation of the programme, schools received approximately 1.5 days of further support from regional trainers consisting of two half-day visits and approximately half a day of remote support (a mix of phone and email support) at a cost of £350 per half-day; 103 schools received training in total, 47 in the ICT model and 56 in the paper-based model. We did not collect data relating to the cost of staff cover during training or implementation so this is not included in our costing model.

Venues were hired for both the training of the regional trainers and the training of schools; these are assumed to cost £500 per day and £250 per half day.

Financial costs—materials

With regard to materials, to facilitate implementation of the ICT delivery model, schools needed to have access to a computer and the internet. It was established upfront whether schools had these facilities and, to our knowledge, no schools specifically bought in IT equipment for the specific purposes of delivering the programme. Schools in both arms required copies of the programme manual, these are costed at £29.45 for two copies per school for the ICT arm and £83.93 for the paper-based arm. Schools in the paper-based arm required physical copies of books, each deliverer received their own set and, given an average of two deliverers per school, the cost is estimated at £50.40 per school. Costs for both the manuals and the books are based on the actual printing costs incurred in the trial. These may vary somewhat in the future depending on how the manuals and books are produced going forward. For example, the manuals produced for the purposes of the trial were printed in bulk with a 'bulk-buy' discount. If manuals were produced individually by a publisher in the future there may be greater printing costs. Equally, printing costs may reduce if schools printed manuals in house. Costs could be substantially reduced if manuals were provided digitally.

As part of the survey, schools in both programme arms were asked about any other additional, or hidden, costs needed to implement the programme. Only one school in the ICT-arm reported additional costs of £250 but they did not specify what this was for. Considerably more schools in the paper-based arm reported additional costs. These ranged from £10 to £300 and were largely related to preparing paper resources. These have been built into the costing model.

Table 36: Costs of the implementation of the programme, per resource

				Year 1			**Year 2**	**Year 3**
			Start-up or recurring	Quantity required	Price per unit	Total cost	Total cost	Total cost
Total costs of training 103 schools								
Personnel (training the region trainers)		Academic trainer (professorial level)	Start-up	3 days in total	£900 per day	£2,700	£0	£0
		Academic trainer (non-professorial level)	Start-up	2 days in total	£183.30 per day	£366.60	£0	£0
		Regional trainer (x10)	Start-up	7 days in total	£350 per day	£24,500	£0	£0
Personnel (training the school staff)		Regional trainer (x10)	Start-up	13 x full days in total; 15 x half days in total	£500 per full day; £350 per half day	£11,750	£0	£0
Other training costs		Venue hire	Start-up	18 x full days in total; 15 x half days in total	£500 per day	£12,750	£0	£0
Total cost of training all programme schools						£52,066.60	£0	£0
Total cost of training per school						£505.50	£0	£0
Other costs per school								
Personnel (ongoing support)		Regional trainer	Start-up	3 x half days per school	£350 per half day	£1,050	£0	£0
Programme costs	ICT	Manuals	Start-up	2 per school	£14.50	£29	£0	£0
	Paper-based	Manuals	Start-up	2 per school	£42	£84	£0	£0
		Books	Start-up	2 sets per school	£25.20	£50.40	£0	£0
Equipment and materials	Paper-based	Materials	Optional recurring	Variable	Variable	Approx. £75 (£0–£300)	Approx. £75	Approx. £75

Note: As school staff time is not costed, it is not included in this table. In the trial, one manual was provided to each deliverer; on average there were two deliverers per school hence we assume two copies. The manual in the paper-based arm is much larger hence more costly. An equal assumption is made for the books used in the paper-based programme.

Overall costs

In total, £160,216.60 was spent on training and supporting the relevant personnel within the trial (regional trainers and school staff), which equates to £1,555.50 per school (total spent divided by the number of schools trained). This constitutes the largest portion of the overall cost. Additional costs included: the price of the manuals, £29 per school for the ICT arm and £84 per school for the paper-based arm, the price of the books, £50.40 per school in the paper-based arm, and additional materials, £75 in the paper-based arm.

Table 37 and Table 38 present the total cost per pupil per year over three years. This calculation assumes an average class size of 26 (as per our trial numbers) and that no additional training of school staff is required. Additional costs could be incurred if schools needed to replace manuals or books during the three-year period.

Table 37: Cost of delivering the RUKS programme for the ICT (ABRA) delivery model

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
One-off training/in-school support	Start-up cost per school	£1,555.50	£1,555.50	£19.94
Materials	Start-up cost per school	Manuals: £29	£29	£0.37
Total				£20.31

Table 38: Cost of delivering the RUKS programme for the paper-based delivery model

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
One-off training	Start-up cost per school	£1,555.99	£1,555.50	£19.94
Materials	Start-up cost per school	Manuals: £84 Books: £50.40 Total = £134.40	£134.40	£1.72
Additional materials	Optional recurring cost per school	£75	£225	£2.88
Total				£24.54

Conclusion

Table 39: Key conclusions

Key conclusions	
1.	Children in schools receiving the paper-based programme made the equivalent of two additional months' progress in reading, on average, compared to children in control schools. This result has a high security rating.
2.	Children in schools receiving the ICT-based programme made no additional progress in reading, on average, compared to children in control schools. This result has a high security rating.
3.	Children in both the paper-based and ICT-based programme schools made additional progress in decoding and letter-sound knowledge compared to children in control schools. The paper-based programme also had a small positive impact on pupil attitudes towards reading.
4.	The implementation and process evaluation found that the training model worked well. Schools felt well supported to implement the programmes and reported that they were able to implement them as intended, although a lower proportion of deliverers in the ICT arm reported delivering the intended number of sessions. Around half of deliverers felt the programmes were of equal or better standard than their usual provision and around half said they would use it again (58% in the paper-based arm and 45% in the ICT arm).
5.	79% of the deliverers of the ICT programme reported experiencing technology issues. These largely centred on unstable internet connections. This may help to explain the lack of impact on reading attainment observed for the ICT programme. Some teachers also suggested that the Canadian pronunciation and cultural context in the ICT programme may have been challenging for pupils.

Impact evaluation and IPE integration

Evidence to support the logic model

This evaluation was driven by a clear logic model that detailed the integral components necessary for implementation and the hypothesised outcomes. The model starts with the recruitment of schools and the cascade, 'train-the-trainers' model of training with ongoing just-in-time support provided to schools throughout implementation. Ultimately, 157 schools accounting for 4,007 pupils were included in the evaluation. This was lower than the target sample size of 201 schools and 5,427 pupils. There was some imbalance between the ICT arm and the other two arms in terms of the type of school. In the ICT arm a high proportion of schools were classed as academy-converters (40%) while in the other two arms a high proportion of schools were classed as community schools (paper-based, 47%; control, 47%). Additionally, there was some imbalance in terms of school setting with more schools in the paper-based arm being rural (paper-based, 23%; ICT, 19%; control, 16%). There was also some imbalance in the Ofsted ratings with fewer schools in the ICT arm being classified as 'outstanding' (6% compared to 15% paper-based and 16% control) and more schools in the ICT arm having no Ofsted status (28% compared to 15% in the paper-based and 14% control). It is unlikely that these differences would have impacted the results in any meaningful way. Rates of attrition were slightly higher in the paper-based arm (19%) compared to the ICT arm (14%) and control arm (11%), though a sensitivity analysis suggests that this has little impact on the trial results.

The training and support component of the logic model was well supported. Training and support appeared to be strong and well received at each level. The 'train the trainers' or 'cascade' model worked well. Regional trainers felt sufficiently prepared to deliver training to school staff who, in turn, felt well-prepared to deliver the programme to pupils. Schools generally felt supported by regional trainers. Evidence from the IPE suggests that the training model was generally appropriate. However, there was also evidence that some aspects of the training could be improved. For example, there was some misunderstanding of the progression rules, both at the school level and at the trainer level. Some schools felt the 'just in time' support could have been better and would have preferred more frequent check-ins from their regional trainer. There was also a tendency not to proactively contact the trainers and to keep questions for trainer visits to the school, which meant that problems or queries were not always resolved promptly. The workload of each regional trainer appeared appropriate, though this had to be increased part way through the trial due to a redistribution of workload from a regional trainer who withdrew from the trial.

The evidence to support the implementation component of the logic model was somewhat mixed. School staff involved in the delivery of sessions in both arms seemed to understand the overall principles of the programme and attempted to deliver it with fidelity, although there were challenges. In the lesson observations, sessions scored well on exposure to programme activities and instructional guidance. The only element that caused widespread confusion was how to apply the progression rules and complete the records. This confusion (partly stemming from an initial misunderstanding by some of the trainers) was usually clarified in the first visits trainers made to the schools. Paper-based deliverers were

more likely than those delivering the ICT version to think the programme was flexible. There were several instances where they adapted the materials to be accessible to all members of the group, irrespective of ability, whereas ICT deliverers seldom made modifications.

Several of the delivery challenges applied across paper-based and ICT delivery. Timing proved problematic on several fronts. Fitting a session into 15 minutes was difficult for many deliverers. The observed sessions did poorly on pacing and efficient use of time. Running four sessions a week for the whole class was considered by school staff to be very burdensome and there was a strong sense from many schools that the programme would be more practicable as a targeted intervention for specific pupils. Indeed, this fits with the issues at recruitment where several schools asked to deliver the programme to a sub-sample of pupils as opposed to the whole class. This provides evidence from the IPE that a more targeted approach could be appreciated by schools to aid efficient delivery and reduce the burden of the programme while ensuring those pupils for whom the programme is perceived by the school to be most appropriate would be targeted. This could be explored further through further research.

Deliverers and children tended to be positive about the small group working inherent to the programme. However, there were some issues in schools that had chosen to use mixed ability groupings, with reports that lower ability pupils were left behind while those at the top of the range did not receive sufficient challenge. This could have impacted the motivation and attitudes component of the logic model. Other schools had chosen to group by ability because they judged that, as a result, children worked better together. Additionally, grouping pupils by ability more closely mirrored usual practices in school with regard to literacy teaching.

Schools faced numerous logistical difficulties. Although it was intended that the programme would be delivered in addition to other literacy provision, one in four deliverers reported that children were being removed from literacy-related classes to participate in sessions. Many schools lacked an ideal venue for holding the sessions and while the majority took place in a separate room, significant minorities were held in an open area or (especially if run by teachers) in the main classroom. Additionally, according to the responses provided in the end of trial survey, four in five of those delivering the ICT programme reported IT problems, often related to internet issues, and the regional trainers identified IT as the main issue they had come across. Survey data indicated that 64% of deliverers in the ICT arm delivered the full programme compared to 79% in the paper-based arm.

Evidence to support the outcomes component of the logic model was well supported for the paper-based version of the programme but less so for the ICT version, both through the findings of the impact evaluation and the process evaluation. Teacher perceptions as to whether children had demonstrated improved comprehension was relatively consistent across trial arms. There is some evidence that children in both programme arms (ICT and paper-based) made some additional progress compared to children receiving teaching as usual on two of the secondary outcomes of the Letter Sound Test and the Diagnostic Test of Word Reading Processes. Children in the ICT arm made one month of extra progress on the DTRWP ($p = 0.27$) and two months' extra progress on the LeST ($p = 0.06$) compared to children receiving teaching as usual. By comparison, children in the paper-based arm made two months' extra progress on both the DTRWP ($p = 0.01$) and the LeST ($p = 0.09$) compared to children receiving teaching as usual. This implies the programme improved skills in decoding and phonics more than business-as-usual approaches. However, for the ICT programme arm of the trial, this did not translate into additional progress on the reading attainment primary outcome measure, the Progress in Reading Assessment: with children receiving the ICT version of the programme making no additional months of progress compared to children receiving teaching as usual ($p = 0.99$). However, there was strong evidence of an effect for the paper-based version of the programme with these children making an additional two months' progress compared to controls ($p = 0.04$).

In addition to the primary analyses, we also conducted three planned sensitivity analyses and one post-hoc sensitivity analysis. In the first of the planned sensitivity analyses we explored the impact on pupils whose pre-test was collected post-randomisation (for the majority, pre-test was collected prior to randomisation). By excluding pupils whose pre-test was collected post-randomisation, the effect size found in the comparison between paper-based and control reduced slightly from 0.15 to 0.11 but remained similar for the comparison between ICT and paper-based (-0.14 to -0.13) and for the ICT to control comparison (0.00 to 0.02). Given that the direction of effect remains the same, the very slight difference between these findings and those in the primary analysis is likely due to a loss of power resulting from the reduced sample size included in the analysis.

In the second of the planned sensitivity analyses we explored the impact of replacing the pre-test PiRA score with the EYFSP result for literacy as the measure of prior attainment covariate in the analysis models. This was to examine whether routinely collected data could be used as a pre-test to negate the need for additional pre-test assessments in future research in the area, Also, if the EYFSP result was more strongly correlated to the post-test PiRA than the pre-test PiRA score, this would give us additional power. This again produced very similar results to the primary analysis, though the 95% CIs were slightly wider, likely indicating that using the EYFSP does not give as much power as using the PiRA as the pre-test.

After randomisation, several schools informed us that they could no longer support the programme at the full class level due to staffing limitations. In these cases, we randomly selected a number of pupils from each school to continue receiving the programme. In the primary analysis all pupils were analysed as randomised under the rules of intention to treat; however, we wanted to explore the impact that these pupils, who we know did not receive the programme to which they were assigned, had on the results. Hence, in the third of the planned sensitivity analyses, we explored the impact of pupil deselection. Removing these pupils increased the effect sizes found in the comparison between paper-based and control very slightly (from 0.15 to 0.17, p = 0.03) and between ICT and paper-based (from -0.14 to -0.17, p = 0.02) in favour of paper-based. The effect size for the ICT to control comparison changed negligibly (from 0.00 to -0.01, p = 0.87). These differences do not change the interpretation of the primary analysis; hence, we can be confident that pupil deselection did not have a significant impact.

An additional post-hoc sensitivity analysis was conducted to include 27 pupils whose post-test data had been collected by their teachers rather than by blinded assessors as per all other included pupils. Including these pupils in the analysis reduced the effect size for the paper-based to control comparison slightly (from 0.15 to 0.12, p = 0.12) but had a limited effect on the effect size for the ICT to paper-based comparison (from -0.14 to -0.13, p = 0.07) and the ICT to control comparison (from 0.00 to -0.02, p = 0.81). It was noted that the 27 pupils performed well above the average for the rest of the trial sample. This could indicate that teachers either administered the assessment differently to the blind assessors or were considerably more generous in their marking.

Taken together, the results of all sensitivity analyses broadly align with the results of the primary analysis thus we can be confident of their accuracy and this component of the logic model.

The paper-based version of the programme appeared to have a small positive impact on children's motivation or attitudes towards reading. No meaningful difference was observed for those receiving the ICT version of the programme. Data collected through the IPE found that children generally liked the stories but sometimes did not enjoy the time spent on each book and found the repetition demotivating. In the paper-based arm, school staff remarked on the poor presentation of the books. In the proposed revised logic model, we have omitted improved attitudes towards reading and improved motivation as contributory factors to overall improved reading attainment (Appendix S). Overall, the results of this evaluation show some support for the logic model, particularly when considering the paper-based version of the programme.

Interpretation

This trial evaluated the RUKS programme, which was developed by the implementation team based on ABRACADABRA, a free, interactive, web-based literacy platform designed for use with primary level children by academics at Concordia McGill Universities, Canada. The RUKS programme is available in two delivery models, either using ICT to access the web-based programme or via a non-digital paper-based format. The effectiveness of the online ABRA platform has been supported through smaller efficacy RCTs (Comaskey, Savage and Abrami, 2009; Savage et al., 2009) and a larger effectiveness trial conducted in Canada (Savage et al., 2013). Other, non-RCT evidence has also been generated worldwide and summarised via systematic review (Abrami et al., 2020).

The results of the present evaluation are somewhat in line with the findings of the efficacy study (McNally et al., 2016), the only other U.K.-based study of the programme, in that there is some evidence of a benefit to children for both delivery models of the programme, both ICT and paper-based, but these are greater for the latter. As per the efficacy trial, children receiving either version of the programme made some additional progress on one of the secondary outcomes —the Letter and Sound Test, a measure of a child's ability to sound out single letter and letter combinations—with children in both arms making an additional two months' progress compared to children receiving teaching as usual. This level of progress is equivalent to that observed in the efficacy trial for children in the paper-based arm but is lower for

children in the ICT arm who made three months' improvement rather than two months' improvement in the earlier trial. With regard to the other secondary outcome—the Diagnostic Test of Word Reading Processes, an assessment of children's ability to read regular, exception, and non-words—children in the paper-based arm made an additional two months' progress compared to controls, in contrast to the children in the ICT arm who made an additional month of progress. This differs somewhat to the results of the efficacy trial where an additional two months of progress was observed in both programme arms.

Although some level of progress was made on both secondary outcomes for children in both versions of the programme, evidence of additional progress on the primary outcome, the Progress in Reading Assessment, was only observed for children receiving the paper-based version of the programme who made an additional two months' progress. There was no evidence that children in the ICT arm made additional progress in reading compared to controls. This is inconsistent with the efficacy trial where children in both programme arms made progress: the equivalent of two months for the ICT delivery model and three months for the paper-based model.

There are several differences in how the programme (both versions) was delivered in this effectiveness trial, which could have contributed to this lower level of success. First, although the train-the-trainers model worked well, the academic implementation team had much closer contact with schools during the efficacy trial and were on hand to resolve problems and provide additional guidance in a timely manner. In the present trial, where regional trainers were responsible for schools, the onus was on the school to contact the regional trainer with any issues between scheduled visits. Data collected as part of the IPE suggested that schools were not particularly proactive in doing this, instead saving any problems arising after the first visit until the Week 10 visit, by which time half of the specified 20-week programme had been completed. Additionally, the progression rules were misunderstood by several trainers and this misinterpretation was subsequently passed on to school staff. Second, schools in the ICT arm of the efficacy trial were provided with a laptop with the ABRA software installed which was used offline meaning schools were not reliant on internet facilities coping with the demands of the programme. Given that four in five of the schools assigned to the ICT arm in the present trial reported technology issues, and 12% were using tablets to deliver sessions despite receiving guidance to the contrary, this could be a significant contributory factor to its lower success relative to the earlier trial. Finally, the Canadian pronunciation posed considerable issues in the ICT arm with phonics practices and cultural context being repeatedly identified as problematic due to the nature of how the software operates with the Canadian voice. Most deliverers did 'correct' to British English pronunciation or phonics where necessary, but some were reluctant to do this which could have further impeded children's progress.

Several other potential barriers to success were also identified, for example, schools in both programme arms found that too much time was spent on each story leading to pupils losing interest. Schools in the paper-based arm expressed some dissatisfaction with the storybooks themselves, especially the appearance of the text (small font, splitting words over two lines). Despite these issues, a small positive improvement was observed on the reading attitudes questionnaire for children in the paper-based arm indicating that the programme may be beneficial in improving children's motivation to read. No impact was observed for children in the ICT arm.

Another potential reason as to why the paper-based delivery model may have been more successful than its digital counterpart may have been related to the perceived flexibility and the number of adaptations made. While most deliverers in both arms said they usually followed the guidance in the manual, deviations were more common among those following the paper-based programme and often they were making adaptations for specific pupils or amending timings to give more time to those struggling. By contrast, those delivering the ICT version were twice as likely as their paper-based counterparts to consider the RUKS programme to be inflexible as well as reporting spending less time preparing and planning sessions. Additionally, data from the pre-implementation survey revealed that over a quarter (28%) of deliverers had some level of discomfort with regard to using ICT to support teaching. The familiarity of the paper-based approach may have aided in delivery.

Despite these issues, observations of programme sessions, as well as feedback from trainers and delivers, showed that pupils were engaged with the programme, and this was particularly true for the ICT version. The use of technology was felt to provide novelty and a degree of interactivity that the children enjoyed. However, high levels of engagement did not translate into greater progress in terms of reading. This may be due to issues surrounding the consistency of the technology or the issues around the Canadian context. The majority of deliverers reported that the programme was of equal or higher quality to their usual literacy interventions and many wished to use the programme again in the future.

Overall, delivering four sessions a week to small groups across the entire Year 1 cohort was often considered by schools to be too time-consuming and resource hungry. Several schools informed us at the outset of the trial that they needed to reduce pupil numbers in order to be able to deliver the programme suggesting that for some schools it would not be feasible to deliver the programme to the whole class. Collectively, the evaluation and implementation teams got a sense from the schools delivering both versions of the programme that it would be more positively received as a targeted intervention for lower ability children and, particularly in the ICT arm, if there was a British English version available.

Targeting specific pupils to receive the programme as a small group teaching intervention would be more consistent with the EEF Teaching and Learning Toolkit, which recommends that small group teaching is most likely to be effective if it is targeted at pupils' specific needs and that lower attaining pupils gain the most benefit. This could be supported by feedback collected through the IPE with schools reporting that higher ability pupils were not being challenged sufficiently and were sometimes being held back, though this could also be due to misinterpretation of the progression rules. Further research could be used to established whether particular subgroups of pupils are more likely to benefit than others.

The findings of this evaluation support the rollout of the paper-based version of the programme. Should this version of the programme be rolled out at a national level, we would recommend that the initial three-day training of regional trainers be condensed into two days and that the training of school staff (which often finished early) made better use of the time available to incorporate more opportunities to practise and study lesson plans. We would also recommend further development of the printed materials used in the paper-based arm to tie in with feedback received from schools around presentation. Given the feedback from schools, we would recommend that schools are given more guidance around grouping. Finally, we would recommend that contact with trainers is increased, especially during the early weeks of programme delivery, and specifically that trainers take a more proactive approach to 'checking in' on schools.

Given the positive results of the efficacy trial, which was not impacted by the ICT issues noted in the present work, we would recommend further exploration and development of the ICT version of the programme prior to any rollout. If possible, an offline version of software could be made available to make the digital version of the programme more accessible to those schools with limited ICT facilities and less reliable internet connections. This could be particularly relevant for remote rural schools. We would also recommend, if possible, that a British English version of the digital programme be made available and its effectiveness explored. This would be particularly beneficial if the programme were to move forward as a targeted intervention as it was reported that the lower ability pupils struggled the most with the Canadian voice related issues.

Limitations

There were some limitations to the evaluation. Two schools, accounting for 27 pupils, were excluded from the primary analysis as the schools conducted the PiRA with the children for their own purposes just before the post-test was due. Although schools sent us the pupil's scores, it was deemed inappropriate to include these in the primary analysis as this could introduce bias thus reducing the sample size slightly. The sensitivity analysis conducted to explore this appeared to confirm that teacher markings may introduce more bias as these pupils had considerably higher scores than the sample average but this only had a small impact on the effect sizes found relative to the primary analysis.

Only around half of the schools returned session registers so we cannot be fully confident that dosage was completely adhered to. Due to the insufficient number of registers, we were unable to undertake the planned CACE analysis to explore the effect of compliance and cannot thus fully understand what impact this might have had on the results. It is unclear as to why compliance with this aspect of the evaluation was so low. It may be that deliverers simply forgot to record attendance as this may have been an unusual action for them. Fidelity to the programme as measured in both groups was high when assessed through observations, however fidelity in terms of dosage is unclear. The paper-based version of the programme is more accessible and implementable by schools as this does not rely on internet connection. Where appropriate IT infrastructure is in place, the observational work conducted as part of the IPE would suggest that the ICT version of the programme is also implementable. The results can be generalised to other primary schools with similar demographics to the schools included in the evaluation and the cascade model of training could be implemented to roll out the programme nationally.

As described in the IPE section, we were unable to evaluate the use of progression monitoring, which limits our understanding of this component of the programme. We were also unable to observe integration with other classroom practices thus we are unable to understand this aspect also.

To deliver the programme, schools had to displace either lesson time or recreational time (for example, if the programme was delivered during break or lunch time). We did not collect data to evaluate the impact of this, particularly in terms of the impact on attainment in other subjects or on pupil wellbeing.

Unfortunately, we were unable to conduct any longer-term follow-up by exploring the impact of the programme on Key Stage 1 outcomes because these assessments did not take place due to the Covid-19 pandemic. This means that although we can be confident that there is an immediate short-term benefit of the programme (particularly for the paper-based version), this may not be sustained overtime and may not translate into additional longer-term attainment outcomes.

Additionally, we were unable to explore the independent effects attributable to children receiving small group teaching, which has been shown to be beneficial through past research and is recommended in the EEF Toolkit. Although all schools were asked prior to randomisation to identify pupils with whom they would deliver some small group teaching if allocated to the teaching-as-usual arm, very few schools did this. Several advised us that small group teaching does not form part of their usual practice for Year 1 literacy and that if they were not delivering a specific programme, such as RUKS, then they would not undertake small group teaching. However, given the evidence of additional progress being made in the paper-based arm compared to the ICT arm, both of which are implementing small group teaching, we can be confident that at least some of the benefit observed in the paper-based arm is directly due to the programme content and structure.

Future research and publications

In terms of future research, the evaluation has provided clear support for the paper-based version of the RUKS programme. There does, however, remain some uncertainty around the effectiveness of the ICT version within the English context. Although there was no evidence of additional progress in terms of overall reading attainment, children receiving the ICT version of the programme did make some additional progress on the Letter and Sound Test and the Diagnostic Test of Word Reading Processes suggesting the programme could be beneficial. If the challenges faced by schools in implementing the ICT version of the programme—specifically internet issues, training processes, and possibly the Canadian context—could be resolved then the benefit of the programme may be greater. This could potentially be explored in the future.

One of the barriers identified by deliverers was the issue of delivering the sessions within the allocated timeframe. Further research could explore the benefit of making the programme more flexible, extending the recommended session duration or giving schools more guidance around managing time.

As stated in the evaluation protocol (Ainsworth et al., 2018), we had planned to conduct further research by way of an addendum to this report exploring the impact of the programme on KS1 outcomes. Unfortunately, due to the Covid-19 pandemic this is no longer possible as all Key Stage 1 assessments were cancelled for the cohort of children involved in this trial. Should any future research be centred on this programme, it would be wise to include an assessment of the impact on KS1 outcomes, if possible, given that the programme was associated with improved outcomes in the efficacy trial (McNally et al., 2016).

Children were withdrawn from a range of lessons, including other core subjects such as maths and science; we did not collect any data that could explore unintended consequences of this missed lesson time. This, too, could be factored into any future research.

Following the publication of this report, we expect to publish academic papers in collaboration with the implementation team at Nottingham Trent University.

References

Abrami, P. C., Savage, R. S., Deleveaux, G., Wade, A., Meyer, E. and Lebel, C. (2010) 'The Learning Toolkit: The Design, Development, Testing and Dissemination of Evidence-Based Educational Software', in P. Zemliansky and D. M. Wilcox (eds), *Design and Implementation of Educational Games: Theoretical and Practical Perspectives*, Hershey, PA: IGI Global (pp. 168–187).

Abrami, P. C., Lysenko, L. and Borokhovski, E. (2020) 'The Effects of ABRACADABRA on Reading Outcomes: An Updated Meta-Analysis and Landscape Review of Applied Field Research', *Computer-Assisted Learning*, 36, pp. 260–279. https://doi.org/10.1111/jcal.12417

Ainsworth, H., Gascoine, L., Fairhurst, C., Torgerson, D., Torgerson, C., Elliott, L., Hewitt, C. and Bell, K. (2018) 'ABRA: Online Reading Support Trial Protocol', London. Education Endowment Foundation.

Altman, D. G., Moher, D. and Schulz, K. F. (2012) 'Improving the Reporting of Randomised Trials: The CONSORT Statement and Beyond', *Statistics in Medicine*, 31, pp. 2985–2997.

Benton, T. and Sutch, T. (2014) 'Analysis of Use of Key Stage 2 Data in GCSE Predictions', Ofqual, University of Cambridge. **https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/429074/2014-06-16-analysis-of-use-of-key-stage-2-data-in-gcse-predictions.pdf**

Braun, V. and Clarke, V. (2006) 'Using Thematic Analysis in Psychology', *Qualitative Research in Psychology*, 3 (2), pp. 77–101.

Campbell, M. K., Piaggio, G., Elbourne, D. R. and Altman, D. G., (2012) 'Consort 2010 Statement: Extension to Cluster Randomised Trials', *BMJ*, 2012;345:e5661.

Comaskey, E. M., Savage, R. S. and Abrami, P. C. (2009) 'A Randomised Efficacy Study of Web-Based Synthetic and Analytic Programmes Among Disadvantaged Urban Kindergarten Pupils', *Journal of Research in Reading,* 32, pp. 92–108.

DfE (2016) 'Schools, Pupils and Their Characteristics: January 2016', London: Department for Education: https://www.gov.uk/government/statistics/national-curriculum-assessments-key-stage-2-2019-revised

DfE (2019) 'National Curriculum Assessments at Key Stage 2 in England, 2019 (Interim)*',* London: Department for Education: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/624576/SFR30_2017_Text.pdf

Dunn, G., Maracy, M. and Tomenson, B. (2005) 'Estimating Treatment Effects from Randomized Clinical Trials with Noncompliance and Loss to Follow-Up: The Role of Instrumental Variable Methods', *Statistical Methods in Medical Research*, 14 (4), pp. 369–395.

EEF (2021) 'Teaching and Learning Toolkit', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit

EEF (2019) 'Cost Evaluation Guidance for EEF Evaluations', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/Cost_Evaluation_Guidance_2019.12.11.pdf

Fairhurst, C. and Mitchell, A. (2019) 'ABRA: Online Reading Support Statistical Analysis Plan', London: Education Endowment Foundation: https://d2tic4wvo1iusb.cloudfront.net/documents/projects/ABRA_Protocol_1.2.pdf?v=1630926355

Hariton, E. and Locascio, J.J . (2018) 'Randomised Controlled Trials—The Gold Standard for Effectiveness Research', *BJOG: An International Journal of Obstetrics and Gynaecology*, 125 (13), p.1716.

Heary, C. M. and Hennessy, E. (2002) 'The Use of Focus Group Interviews in Pediatric Health Care Research', *Pediatric Psychology,* 27 (1), pp. 47–57. DOI: 10.1093/jpepsy/27.1.4

Higgins, S., Katsipataki, M., Coleman, R., Henderson, P., Major, L. E., Coe, R. and Mason, D. (2020) 'Improving Literacy in Key Stage 1', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/literacy-ks-1

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M. and Lamb, S. E. (2014) 'Better Reporting of Interventions: Template for Intervention Description and Replication (TIDieR) Checklist and Guide', *BMJ*, 2014;348:g1687.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Handbook.pdf

McNally, S., Ruiz-Valenzuela, J. and Rolfe, H. (2016) 'ABRA: Online Reading Support Evaluation Report and Executive Summary' (October 2016), London. Education Endowment Foundation.

Moher, D., Schulz, K. F., Simera, I. and Altman, D. G. (2010) 'Guidance for Developers of Health Research Reporting Guidelines', *PLoS Medicine*, 7 (2), p.e1000217.

Morgan, M., Gibbs, S., Maxwell, K. and Britten, N. (2002) 'Hearing Children's Voices: Methodological Issues in Conducting Focus Groups with Children Aged 7–11 Years', *Qualitative Research*, 2 (5), pp. 5–20.

Mulcahy, E., Bernardes, E. and Baars, S. (2019) 'The Relationship Between Reading Age, Education and Life Outcomes', London: The Centre for Education and Youth: https://cfey.org/wp-content/uploads/2019/03/The-relationship-between-reading-age-education-and-life-outcomes.pdf

O'Reilly, M., Ronzoni, P. and Dogra, N. (2013) *Research with Children: Theory and Practice*, London: SAGE.

Savage, R. S., Abrami, P., Hipps, G. and Deault, L. (2009) 'A Randomized Controlled Trial Study of the ABRACADABRA Reading Intervention Programme in Grade 1', *Educational Psychology,* 101 (3), p. 590.

Savage, R., Abrami, P. C., Piquette, N., Wood, E., Deleveaux, G., Sanghera-Sidhu, S. and Burgos, G. (2013) 'A (Pan-Canadian) Cluster Randomized Control Effectiveness Trial of the ABRACADABRA Web-Based Literacy Programme', *Educational Psychology,* 105 (2), p. 310.

Sharples, J., Rob Webster, R. and Blatchford, P. (2015) 'Making Best Use of Teaching Assistants: Guidance Report', London: Education Endowment Foundation.

Thomas, D. R. (2006) 'A General Inductive Approach for Analyzing Qualitative Evaluation Data', *American Journal of Evaluation*, 27 (2), pp. 237–246.

Torgerson, C., Brooks, G., Gascoine, L. and Higgins, S. (2018) 'Phonics: Reading Policy and the Evidence of Effectiveness from a Systematic "Tertiary" Review', *Research Papers in Education,* 1-31.

Wears, R. L. (2002) 'Advanced Statistics: Statistical Methods for Analyzing Cluster and Cluster-Randomized Data', *Academic Emergency Medicine*, 9 (4), pp. 330–341.

Appendix A: EEF cost rating

Appendix Figure 1: Cost Rating

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

Appendix B: Security classification of trial findings

OUTCOME: Reading measured by PIRA – ICT Arm

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

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	Adequate allocation sequence and blinding. No imbalance between groups on the pre-test reading scores.
Threat 2: Concurrent Interventions	Low	No evidence of differential practice between intervention arms and control group related to concurrent interventions during evaluation period.
Threat 3: Experimental effects	Low	No evidence of contamination.
Threat 4: Implementation fidelity	Low	Implementation well defined. Though only half of schools returned a session register so not possible to know how many sessions were completed by all schools and CACE analysis not possible IPE indicated generally high fidelity. School registers that were received showed that on average, schools completed more than 80% of sessions ("full compliance" criteria). The ICT arm was less likely to make adaptations than the paper-based arm.
Threat 5: Missing Data	Moderate	There are some differences in attrition between intervention arms and control groups. However, sensitivity analysis showed the results of the primary analysis was unchanged when missing data was taken into account.
Threat 6: Measurement of Outcomes	Low	Justified, reliable and valid measure used. Blinded administration. Unblinded post-test marking was excluded from primary analysis.
Threat 7: Selective reporting	Low	Study is registered and published protocol and statistical analysis plan adhered to. No evidence of selective reporting.

- **Initial padlock score:** 4 Padlocks – Randomised design, MDES=0.2, 13.8% attrition for ICT intervention arm 11% attrition for control
- **Reason for adjustment for threats to validity**: 0 Padlock – one moderate threat to internal validity with direction of bias unknown.

- **Final padlock score:** initial score adjusted for threats to validity = 4 Padlocks

OUTCOME: Reading measured by PIRA – Paper-based Arm

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

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	Adequate allocation sequence and blinding. No imbalance between groups on the pre-test reading scores.
Threat 2: Concurrent Interventions	Low	No evidence of differential practice between intervention arms and control group related to concurrent interventions during evaluation period.
Threat 3: Experimental effects	Low	No evidence of contamination
Threat 4: Implementation fidelity	Low	Implementation well defined. Though only half of schools returned a session register so not possible to know how many sessions were completed by all schools and CACE analysis not possible IPE indicated generally high fidelity. School registers that were received showed that on average, schools completed more than 80% of sessions ("full compliance" criteria). The ICT arm was less likely to make adaptations than the paper-based arm.
Threat 5: Missing Data	Moderate	There are some differences in attrition between intervention arms and control group, particularly in the paper-based arm. It is however noted that missing data would be similar across groups if not for 3 withdrawing schools in the paper-based arm, while the reason for withdrawal was not directly related to being allocated to paper-based (high proportion of SEN, change of staff, lack of time for post-testing.) Sensitivity analysis showed the results of the primary analysis was unchanged when missing data was taken into account.
Threat 6: Measurement of Outcomes	Low	Justified, reliable and valid measure used. Blinded administration. Unblinded post-test marking was excluded from primary analysis.
Threat 7: Selective reporting	Low	Study is registered and published protocol and statistical analysis plan adhered to. No evidence of selective reporting.

- **Initial padlock score:** 4 Padlocks – Randomised design, MDES=0.2, 18.7% attrition for paper-based intervention arm 11% attrition for control
- **Reason for adjustment for threats to validity**: 0 Padlock – one moderate threat to internal validity with direction of bias unknown.
- **Final padlock score:** initial score adjusted for threats to validity = 4 Padlocks

Appendix C: Changes since the previous evaluation

Appendix table 1: Changes since the previous evaluation

	Feature	Efficacy to effectiveness stage	Rationale for Change
Intervention	Intervention content	An extra story book was added.	Reflected an update to the ABRA platform whereby additional material was added around an additional story.
		Formalised progression rules using record sheets were added.	To provide additional structure and guidance to schools.
		The activity logs/registers were updated to capture usage outside the core usage (four 15 minute sessions a week for twenty weeks).	To gain a greater insight as to how the programme was been used beyond the mandated timings.
		Four types of programme extension were offered: 1. Extended sessions to 20-30 minutes; 2. Extended programme duration to 24 or 30 weeks; 3. Introduction of related activities (e.g. grow a beansprout when working on the book about this); and 4. Give extra tuition to those struggling.	These were introduced to offer support to those struggling or conversely to add more challenging activities to higher ability pupils.
	Delivery model	**Changes for all schools**	
		As opposed to the efficacy trial where training was limited to teaching assistants, in the present trial training was open to, and delivered to, a range of nominated school staff.	This was implemented to accommodate for school preferences and to give access to schools who did not have TA's available for training.
		Intervention schools were required to pay a cost of £200 to cover the costs of training rather than receiving the training free of charge.	This is typical process when scaling up to effectiveness from efficacy. All other costs were subsidised by EEF.
		Changes for the ICT delivery model	
		It was stipulated that the programme be delivered at least on a laptop or desktop computer connected to the internet, but not via tablets	This was to ensure all pupils could see the screen and work as a group.
		Nominated school staff were provided with ICT facilities during training to access and practice with the software	This was viewed as a valuable addition to training to ensure staff were comfortable with the ICT requirements of the programme.
		Nominated school staff were provided with best practice guidelines sheets	These provided useful insight gained via the efficacy trial.
		ABRA has been updated from a Flash-based version to an html5 version.	Software required updating.
		School needed to access online software rather than the software being available offline	This reflects the real world processes for accessing the ABRACADABRA platform.
	Intervention duration	There were no changes made to the intervention duration for this effectiveness trial.	N/A

Evaluation	Eligibility criteria	There were no changes to the eligibility criteria for pupils. Schools were asked at the outset to confirm whether they had ICT facilities in place to facilitate the delivery of the ICT-based programme, where schools reported that they did not, they were only randomised between the paper-based programme arm and the control arm.	N/A
	Level of randomisation	The level of randomisation was between rather than within-school.	It was viewed by the evaluation team that the risk of contamination was high using within school randomisation.
	Outcomes and baseline	The Progress in Reading Assessment (PiRA) was used as the primary outcome as per the efficacy trial. Two of the four secondary outcomes used in the efficacy trial, the DTWRP and the LEST, were used in the effectiveness trial as these had shown promise in the efficacy trial. An additional secondary outcome was added for the effectiveness trial, the Reading Attitudes Questionnaire.	The developers/implementation team added the hypothesis that exposure to the programme may influence reading attitudes and motivation to the logic model hence an instrument to test this was added.
	Control condition	There were no changes to the control arm of the trial.	N/A

Appendix D: Effect size estimation

Appendix Table 2: Effect size estimation. Note that due to rounding error the adjusted difference provided in this table will not be exactly equal to the effect size multiplied by the square root of the pooled variance.

Outcome	Adjusted differences in means (95% CI)	Pooled variance	Effect size (95% CI)
	ICT vs Control		
PiRA	-0.01 (-1.52, 1.51)	123.8	0 (-0.14, 0.14)
DTWRP	1.32 (-1.04, 3.68)	227.3	0.09 (-0.07, 0.24)
LeST	0.94 (-0.06, 2.02)	39.8	0.15 (-0.01, 0.32)
RAQ	-0.05 (-1.33, 1.24)	80.1	-0.01 (-0.15, 0.14)
	Paper-based vs Control		
PiRA	1.70 (0.10, 3.30)	122.9	0.15 (0.01, 0.30)
DTWRP	2.60 (0.68, 4.52)	227.4	0.17 (0.05, 0.30)
LeST	0.82 (0.13, 1.71)	40.1	0.13 (-0.02, 0.27)
RAQ	0.97 (-0.48, 2.43)	81.5	0.11 (-0.05, 0.27)
	ICT vs Paper-based		
PiRA	-1.56 (-3.07, -0.05)	123.8	-0.14 (-0.28, 0)
DTWRP	-1.33 (-3.59. 0..93)	227.3	-0.09 (-0.24, 0.06)
LeST	0.06 (-0.88, 1.14)	39.8	0.01 (-0.14, 0.18)
RAQ	-0.87 (-2.39, 0.66)	80.1	-0.10 (-0.27, 0.07)

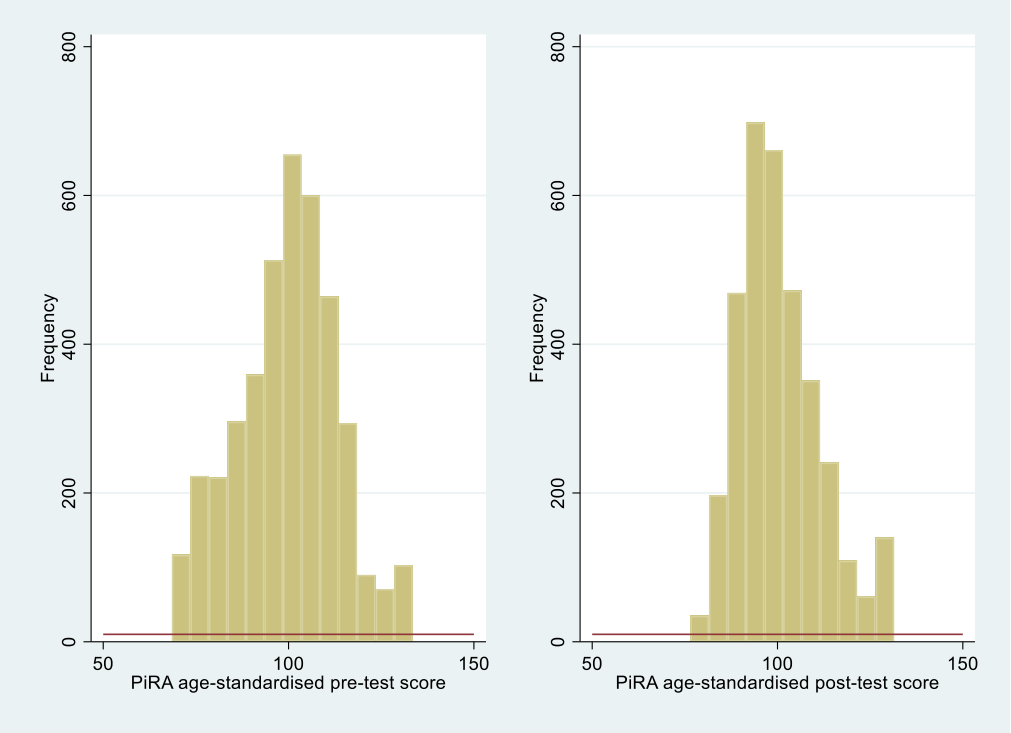
Appendix table 3: Intracluster correlations for each outcome from the unadjusted models, for the schools with ICT facilities.

Outcome	ICC (95% CI)
PiRA	0.16 (0.12, 0.20)
DTWRP	0.10 (0.06, 0.15)
LeST	0.10 (0.06, 0.15)
RAQ	0.08 (0.04, 0.15)

Appendix table 4: Intracluster correlations for each outcome from the unadjusted models, for the schools randomised to paper-based or control.

Outcome	ICC (95% CI)
PiRA	0.17 (0.13, 0.23)
DTWRP	0.10 (0.07, 0.16)
LeST	0.11 (0.06, 0.17)
RAQ	0.10 (0.05, 0.18)

Figure 2: Histograms of the pre-test age-standardised PiRA score and the post-test age-standardised PiRA score. A red line has been drawn at frequency =10 to indicate all bins exceed 10 and are therefore not disclosive.


Frequency
0
200
400
600
800
50
100
150
PiRA age-standardised pre-test score
Frequency
0
200
400
600
800
50
100
150
PiRA age-standardised post-test score

Further appendices:

Please see accompanying document *'Technical Notes'*.

This document is available for download at https://educationendowmentfoundation.org.uk

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

https://educationendowmentfoundation.org.uk
@EducEndowFoundn
Facebook.com/EducEndowFoundn