Education Endowment Foundation

ABRA: Online Reading Support

Evaluation report and executive summary

October 2016

Addendum added March 2018

Independent evaluators:
Sandra McNally, Jenifer Ruiz-Valenzuela and Heather Rolfe

LSE

National Institute of Economic and Social Research

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.

This project was co-funded by the EEF and Nominet Trust as part of a funding round focused on digital technology.

Making change happen with nominet trust

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

Danielle Mason
Head of Research and Publications

p: 020 7802 1679
e: danielle.mason@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the evaluator

The evaluation team is formed by Professor Sandra McNally and Dr Jenifer Ruiz-Valenzuela at the Centre for Economic Performance, London School of Economics, and Dr Heather Rolfe at the National Institute of Economic and Social Research (NIESR) who led the work on the process evaluation.

e. J.Ruiz-Valenzuela@lse.ac.uk

Contents

Executive Summary

The project

Abracadabra (ABRA) is a 20-week online literacy programme composed of phonic fluency and comprehension activities based around a series of age-appropriate texts. Four 15-minute sessions per week are delivered by a teaching assistant (TA) to groups of three to five pupils. This report summarises the findings of a randomised controlled trial assessing the impact of ABRA on literacy outcomes for Year 1 pupils. The trial also assesses the impact of an offline, paper and pencil version of the same intervention (referred to here at 'the non-ICT intervention'). There were 51 participating schools and 2,241 pupils at randomisation, and a total of 48 schools and 1,884 pupils were included in the final analysis (84% of the initial pupils at randomisation).

The trial took place between October 2014 and May 2015. Fifty-one schools were randomly assigned to either receive some version of the intervention or to act as a 'control' school delivering business as usual. In the schools receiving the intervention, pupils were randomised to receive one of the following options: (1) ABRA, (2) the non-ICT intervention, and (3) standard literacy provision. The process evaluation involved observing sessions to understand a variety of factors in the intervention. These included an evaluation of which elements contributed to successful implementation, the perceptions and experiences of TAs and project leads, levels of pupil engagement, and the mechanisms behind the estimated impacts. This was an efficacy study, due to the involvement of the developer in the delivery of the programme. The study was funded by the Education Endowment Foundation and Nominet Trust as part of a funding round focusing on the use of digital technology to improve outcomes for disadvantaged children.

Key conclusions
1. The children who received ABRA, or its offline alternative were found to make two and three months' progress in literacy respectively compared to the children who received standard provision. This positive result would be unlikely to occur by chance.
2. For both ABRA and the offline alternative, the impact for children eligible for free school meals and children with below average pre-test outcomes was larger than for all pupils.
3. Successful implementation contributed to a well-designed and delivered training programme which emphasised fidelity and consistency, enforced by ongoing support from the project team.
4. The process evaluation found that both the ICT and non-ICT interventions may be best delivered in groups of similar rather than mixed ability. The process evaluation also suggested minor changes to the intervention to make it more culturally relevant to British pupils, and to remove some repetition in the non-ICT programme.
5. Future research will examine whether ABRA or the non-ICT intervention can be successfully delivered at scale, and will look at longer-term impacts through assessing Key Stage 1 data from this trial.

How secure are the findings?

Security rating awarded as part of the EEF peer review

These findings have moderate to high security. The trial was a three-armed randomised controlled trial that featured randomisation between schools that received ABRA and the non-ICT intervention and schools that continued normal literacy provision. It also involved randomisation within the schools that received interventions, between ABRA, the non-ICT intervention, and pupils that received standard literacy provision. The trial was large and the pupils who received the intervention were similar to the pupils in the comparison group; 16% of the pupils were not included in the analysis because they did not complete all the tests at the end of the trial.

What are the findings?

Both the ICT and non-ICT treatments were found to have positive results for literacy that were unlikely to have occurred by chance, although the size of the non-ICT effect is considerably higher. The impact was higher for children eligible for free school meals (FSM) for both ABRA and the non-ICT intervention, with both groups making the equivalent of five months' progress. Pupils with below median pre-test outcomes seemed to benefit from ABRA, whereas the non-ICT intervention seemed to benefit both below- and above-median pupils. Pupils that received normal literacy provision in the schools where the interventions took place did better than students in schools that only received normal literacy provision. This is consistent with the existence of spillover (or peer) effects. This preliminary evidence will be further investigated in future research when further data on teaching assistant surveys and log data becomes available.

The process evaluation indicated that the implementation of both interventions was successful and benefitted from a well-designed and delivered training programme which emphasised fidelity and consistency facilitated by ongoing support from the project team.

These positive findings for the ICT programme are in line with previous studies about ABRA which found that ABRA leads to improvements in literacy. This evaluation adds to this body of evidence through (1) looking at a larger number of students than previous studies, (2) comparing ABRA with a non-ICT treatment, and (3) looking at the results of students that did not receive ABRA but went to school with students that did allowing the study to see if there are benefits to students that do not receive the intervention directly. The comparison between ABRA and the non-ICT intervention, which used the same materials in a non-digital format, is particularly useful as the positive non-ICT results seem to support the idea that it is the literacy programme itself making the difference rather than the digital delivery format. There are several studies that find no evidence of ICT programs having a positive impact on pupil outcomes using strong research designs at a relatively large scale. In this broader literature, the findings of this study are more unusual.

How much does it cost?

The average cost per pupil per year over three years is £8.52 for the ICT intervention and £8.49 for the non-ICT intervention. This cost includes training the teaching assistants, cover during training, and travel costs. All of the costs are frontloaded into the first year of the programmes (which costs £25.56 and £25.47 per pupil respectively). The programme is free to deliver in the following years.

Table 1: Executive Summary Table

Intervention Group	Number of Schools	Effect Size (95% confidence)	Estimated months' progress	Security rating	Cost
ICT	48	0.138 (0.004, 0.273)	2	🔒🔒🔒🔒🔓	£
ICT FSM	48	0.368 (0.089, 0.646)	5	N/A	£
Non-ICT	48	0.231 (0.102, 0.360)	3	🔒🔒🔒🔒🔓	£
Non-ICT FSM	48	0.396 (0.195, 0.596)	5	N/A	£

1. Introduction

Intervention

This report summarises the findings of a randomised controlled trial assessing the impact of two literacy-related interventions: we analyse the effects on literacy of two methods of small-group tuition for Year 1 pupils in English schools. One of these methods used an ICT programme (Abracadabra) which has some support from small RCT efficacy studies in other countries (see background evidence section below). The other method is identical (using materials that replicate the ICT intervention) but delivered without using the computer programme (non-ICT intervention). Both methods were peer-reviewed in advance of implementation.

The content of the ICT intervention was entirely encapsulated within the Abracadabra (ABRA) web-based technology platform, a modular 'game'-based balanced literacy intervention that is fixed in content (new activities cannot and will not be added). The games are related—and are linked in terms of content—to a series of electronic texts (mainly 'stories', some non–fiction) suitable to beginner readers. Phonic fluency and comprehension activities are connected to, and draw most content from, a series of electronic texts. There are extension activities for some of the tasks within ABRA and these can be found in the Teacher area of the site. There are also activities for writing, but the implementation team chose not to include these in the 20-week programme. However, additional extension activities were provided for the comprehension tasks in the form of a printed sheet of ideas that the implementation team went through with the TAs on one of the visits. A summary of the 20-week ICT intervention can be found in Appendix A.

The non-ICT intervention is also defined by a 20-week programme of lesson plans based on pencil and paper activities with additional resources such as magnetic letters and cards. It was especially developed for this study as a paper and pencil version of the ICT intervention: each of the ABRA activities was matched by a paper/pencil version (or using magnetic letters, for example). It also used all the same stories, vocabulary items, questions, words, and letter sounds in all the activities. It was therefore almost identical in content to the ICT version and only differed in terms of delivery medium. Appendix B summarises the non-ICT intervention together with the 20-week programme. The rationale for including the non-ICT intervention in parallel to the ICT intervention was to discern how important the mode of delivery was to its success, as opposed to the pedagogy underlying both approaches.

Robert E. Slavin, director of the Center for Research and Reform in Education at Johns Hopkins University, conducted an independent review of the two interventions on behalf of the Education Endowment Foundation which concluded that they appeared to be comparable in content. Following his suggestions, training for both the ICT and non-ICT conditions were set to be equal in term of time. Training was also closely matched in terms of content but tailored for each specific mode of treatment delivery. Each TA was trained for 1.5 days prior to involvement, in groups of 12–13. Moreover, on average each TA received approximately 0.6 days of further support from the project team (a mix of in-person, phone, and email support).

Training occurred after baseline testing and randomisation and prior to the start of the intervention. For each school, a TA was assigned randomly to the ICT or non-ICT condition before the training event.[1] A small number of big schools had two TAs per condition. Both the ICT and non-ICT TAs received detailed training packs after the training sessions with a description of the activities and why they were useful. The package included the 20-week plan (available on request) outlining the activities to be performed four days per week during the 15-minute sessions. Moreover, the implementation team at Coventry provided just-in-time support to both groups of TAs on request, and they visited the TAs during the first weeks of treatment to observe how the intervention was being delivered and to provide support for the

[1] The school, pupil and TA randomisations were performed using the statistical programme STATA.

TAs. During training TAs received a list of pupils assigned to them. Prior to the start of the intervention TAs had some flexibility in arranging the small groups of pupils (around three to four pupils per group). The purpose of doing so was to give them the flexibility to divide pupils into appropriate groups, as they normally would do for any other activity. In practice, TAs grouped pupils into groups of three to four according to whether they were likely to be able to work well together. This was guided by ability, behaviour, SEN status, and personality. Finally, the process evaluation found no issues of concern over implementation or fidelity in delivery. The intervention was found to be well understood by TAs and implemented as intended. This included aspects such as timing, use of materials and organisation, and practical matters. As far as was practically possible, the two interventions were delivered separately without contamination in the active schools.

The design included two control groups. One group consisted of pupils in purely control schools, and the second group included pupils in the control group of treatment schools. This design allows us to investigate the existence of spillover effects from treatment to control pupils in those schools randomly assigned to treatment. Spillover effects may exist, for example, because teaching assistants interact with pupils at other times of the school day (they were not employed only to implement the intervention). It is possible that the training might help them in their role generally, which would then impact on the control group.

Background evidence

Many children leave primary school without the expected level of literacy.[2] This matters because it has an impact on subsequent educational performance and is linked, for example, to poor labour market outcomes (Hansen and Vignoles, 2005). As discussed by Slavin *et al.* (2011), while much is known about the characteristics of struggling readers, much less is known about practical, replicable programmes capable of helping educators prevent or remediate early reading deficits. Of those programmes covered in their systematic review, the vast majority pertain to the U.S.

As explained in the previous section, we intend to analyse the effects of two methods of small-group tuition for Year 1 pupils in English schools. The ICT programme (based on the Abracadabra Web-Based Literacy Program) has had some support from small RCT efficacy studies in other countries,[3] but the non-ICT intervention was especially developed for this study and designed to be identical in content to the computer-based programme without using ICT. Both interventions are balanced and structured reading programmes that contain a systematic phonics aspect consistent with current UK government recommendations.

The ABRA programme is a freely available computer-based programme, widely used in Canada (Abrami et al., 2010). The programme is not used extensively in Britain, to our knowledge. However, the identity of the computer programme was concealed from all schools to prevent it being used in control schools.

The ABRA programme has had some support from smaller efficacy RCTs (see, for instance, Comaskey, Savage and Abrami, 2009; Savage *et al.*, 2009; and Wolgemuth *et al.*, 2011) and a bigger effectiveness trial (Savage *et al.*, 2013). Savage *et al.* (2009) randomly allocated 174 pupils into three groups: a synthetic phonics intervention group, an analytic phonics intervention group, and a classroom control group. The intervention groups both used the computer-based ABRA programme. The authors state

[2] Currently 11% do not reach level 4 in reading by age 11—the expected standard according to the National Curriculum.

[3] There was no formal piloting in the UK. However, in Coventry University a student tested out ABRA in the context of a small-scale intervention involving 25 Year 1 children from one school, with the control group doing a maths intervention (although total literacy teaching time was equated across the ABRA and control group). The interventions lasted for six weeks (12 sessions) and the ABRA group significantly increased their single word reading scores (using the British Ability Scales II single word reading sub-test) both at post-test, and after a four-week delay (delayed post-test). This work is not published and was produced in 2012.

that effect size analyses confirmed that both interventions had a significant impact on literacy at post-tests. The later study (Savage *et al.*, 2013), describing a classroom-level RCT with just over 1,000 pupils where the intervention was delivered by teachers, also found improvements in literacy for treatment pupils at post-test.[4] The effectiveness trial in this report differs from this later study in several respects: (1) the size of the trial in terms of pupils is doubled, (2) this intervention compares ICT and non-ICT delivery of a broadly identical programme—this allows the evaluation of different delivery methods with respect to improving results for early literacy, and (3) we have a clean control group—pupils in schools that do not receive and do not know about the existence of the web-based programme while the intervention is in place. If we have reasons to believe that there might be spillover effects from treatment to control individuals within the same school, then having a control group in schools that are not treated gives additional information not found in the previous effectiveness trial (which only had a control group within the same school as the treatment group).

Evaluation objectives

The evaluation is designed to answer the following questions, as described in the protocol:

1. Do the interventions have an impact on the treatment groups relative to the control group on outcomes within one academic year?
2. How do the interventions compare to each other in terms of effectiveness?
3. Is there evidence of heterogeneity in the effects of the interventions on subgroups relating to free school meal status, gender, whether pupils speak English as a first language, and above-median attainment on pre-test?[5]
4. Is there any evidence of spillover effects on the control group?
5. Do any effects persist one year later?

Except question five for which data is not yet available (Key Stage 1 data for the cohort under study will be added as an appendix in Spring 2017) all the other questions are addressed in the Results section. The purpose of the process evaluation was to assess the fidelity of the intervention and to help explain impact of small-group tuition as measured by the RCT. We aimed to achieve this largely through an assessment of the practices and experiences of teaching assistants in participating schools and observation of the two interventions—ICT and non-ICT—focusing on teaching and learning processes and particularly pupil response and engagement. We carried out our own process evaluation using qualitative research methods and also used quantitative data collected by the project team, including their assessments of fidelity and quality of delivery.

Project team

The project team who developed and delivered the intervention are based at Coventry University and led by Professor Clare Wood and Dr Janet Vousden. The other team members are Rob Savage, Helen Johnson, Sabrina Ammi, Claire Pillinger and Sam Waldron. The ICT 20-week programme was developed by a team led by Dr Robert Savage, Associate Professor at McGill University.

[4] Savage *et al.* (2013) does not contain all the information to appropriately calculate effect sizes. However, if you calculate the approximate effect size for the outcome CTOPP blending words using the SD of the pre-test for the control group (which is similar to the SD of the pre-test for the treatment group), you get an effect size of 1.08/3.04 = 0.36, which is in line with previous work. It is reasonable to report, in general, a ballpark effect size of 0.3–0.4. However, the effect sizes also vary according to each outcome, for example, larger effect sizes are often observed for measures involving phonological processing (such as letter-sound knowledge, blending, segmenting, and so on, as opposed to, say, single word reading or comprehension).

[5] Heterogeneity according to the first language spoken by the pupil cannot be assessed due to a lack of data. See the results section for more information on this issue.

The evaluation team is formed by Professor Sandra McNally and Dr Jenifer Ruiz-Valenzuela at the Centre for Economic Performance, London School of Economics, and Dr Heather Rolfe at the National Institute of Economic and Social Research (NIESR) who led the work on the process evaluation.

Ethical review

The Coventry University Ethics Committee reviewed and approved this study. Headteachers and relevant teachers were given a written outline of the project and asked to sign consent forms for their schools to be part of the project. A participant information sheet was sent out to all parents giving information about the project and the evaluation. Parents were asked to complete a reply slip saying whether or not they consented for their child to be part of the project. They were told that a non-response would be taken to mean no objection (therefore opt-out consent). A copy of the letters sent to both headteachers and parents can be found in Appendix C.

Trial registration

The trial was registered at:

http://www.controlled-trials.com/isrctn/submission/

The registration number is ISRCTN18254678, and the link to the ISRCTN registry website for this trial is:

http://www.isrctn.com/ISRCTN18254678?q=&filters=conditionCategory:Not%20Applicable&sort=&offset=130&totalResults=707&page=13&pageSize=10&searchType=basic-search

2. Methods

Trial design

The results in this report are based on a randomised controlled trial (RCT) with randomisation taking place at the school (first stage) and at the pupil (second stage) level. Teacher assistants were also randomised into different treatments. In the first stage, recruited schools were paired according to the following characteristics: size of the Year 1 cohort, Key Stage 1 average point score for the 2013 cohort, and percentage of FSM pupils. Additionally, infant schools were paired together. Each school within each pair was randomised to cohort 1 (those receiving the treatment(s) in Year 1) or cohort 2 (those receiving the treatment(s) in Year 2).[6] In the second stage, within treatment schools, children in Year 1 were randomly allocated into three groups:

(A) support from teaching assistants using the computer programme (ICT intervention);
(B) support from teaching assistants using other methods (non-ICT intervention); and
(C) control group within treatment schools.

The design, therefore, included two control groups: one group consisted of pupils in purely control schools, and the second group included pupils in group (C) described above. This design allowed us to investigate the existence of spillover effects from treatment to control pupils in those schools randomly assigned to treatment. Finally, teaching assistants in treatment schools were randomised to deliver the ICT or the non-ICT intervention.

This design was chosen in order to overcome the problem of selection bias.[7] As Duflo *et al.* (2007) point out, selection bias can be entirely removed when individuals or groups of individuals are randomly assigned to the treatment and comparison groups.

There were no major changes to the original trial design, but some minor changes took place post randomisation and before the treatment started. These are detailed in Section 3 (under the subsection on Participants).

Outcome measures

The primary outcome was measured (pre- and post-treatment) by the Progress in Reading Assessment (PIRA) test. This is a test that evaluates the general reading ability of pupils.[8] In particular, it assesses reading ability in the following areas: phonics, literal comprehension, and reading for meaning. It is a favoured test for the Education Endowment Foundation due to having a marking key that the organisation feels is unambiguous and not sensitive to interpretation. This is a well-known test, often used in EEF evaluations. The advantages mentioned above and the fact that overall reading attainment is the cognitive outcome of most importance to interventions of this type were the reasons why the PIRA reading test was chosen as the primary outcome measure. The secondary outcomes assess more specific components of reading.

[6] Note that the equivalent cohort in control schools will not get the treatment at any stage. Cohort 2 is a year younger. This enables us to consider the effects of the treatment on longer-term outcomes (as indicated by Key Stage 1 teacher assessments).
[7] Selection bias arises when individuals or groups are selected for treatment based on characteristics that may also affect their outcomes and makes it difficult to disentangle the impact of the treatment from the factors that drove selection (Duflo *et al.*, 2007).
[8] More information on the PIRA test can be found at **https://www.hoddereducation.co.uk/pira**

The secondary outcomes that have been measured are as follows:[9]

- DTWRP (the Diagnostic Test of Word Reading Processes). This assesses the reading of regular words, exception words, and non-words, enabling the nature of a child's reading difficulty to be easily identified.
- BPVS (the British Picture Vocabulary Scales, Third edition). This is a one-to-one test that assesses a child's receptive vocabulary. For each question, the teacher says a word and the pupil responds by selecting a picture from four options that best illustrates the word's meaning.
- PSS (the Phoneme Segmentation Subtest) and LK (the letter sound subtest)—subtests from the Primary Inventory of Phonological Awareness (PIPA) test. PIPA assesses the nature and extent of a child's phonological awareness development.
- LEST (the Letter Sound Test). This tests a person's ability to sound out single letters and letter combinations.

All of these secondary outcomes were included as secondary measures in the protocol except the LEST test which replaced the PIPA test originally specified in the evaluation protocol. The implementation team at Coventry examined the tests and realised that the letter sound subtest (LK) on the PIPA test was very basic (consisting of mainly single letters and no vowel digraphs such as 'ea' or 'ou'). Given that in the UK most children receive phonics instruction from the Reception Year, they learn the single sounds fast and are often near or at the ceiling by the end of Year 1. Moreover, children often score very high marks at the beginning of Year 1. The single letter sounds are also much less predictive of word decoding. The LEST is much more comprehensive and includes the most frequently used English digraphs ('th', 'sh', 'ea', 'ou', 'ow', 'gh' and so on). Therefore, the implementation team at Coventry decided to include the additional LEST test.

These tests took place at the beginning (September 2014, all before randomisation) and end (July 2015) of the academic year 2014/2015.

We will also analyse secondary outcomes such as Key Stage 1 outcomes on reading and writing one academic year later (as data becomes available in September 2016).

Assessments were administered by a team of Research Assistants (RAs) employed by Coventry University who did not know to what condition the children had been allocated. Furthermore, the RAs were blind to the nature of the study, in other words, they were not given any details about the project other than it was a reading project. Assessments were administered individually or in small groups with spot checks by the project team to ensure correct administration. The baseline PIRA assessment was scored by Hodder Education. All other tests were scored (and entered) by a group of RAs hired specifically for this purpose (not those who carried out the assessments), with no knowledge of how schools or pupils had been allocated to the treatment and control groups, and no knowledge of the nature of the project other than it was a reading project. Data-checking was carried out by the RAs who marked the assessments, and again by the project team. In particular, the implementation team conducted further checks on the data like range checks, age, date of birth, and date of test checks, checking zero scores against missing data, and so on. This was done in order to catch any mistakes in data entry.

[9] More information on these secondary outcomes can be found here:
DTWRP: **http://www.gl-assessment.co.uk/products/diagnostic-test-word-reading-processes**
BPVS: **http://www.gl-assessment.co.uk/products/british-picture-vocabulary-scale-third-edition**
PIPA:
http://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildLanguage/PreschoolandPrimaryInventoryofPhonologicalAwareness(PIPA)/PreschoolandPrimaryInventoryofPhonologicalAwareness(PIPA).aspx
LEST: **http://www.motif.org.au/home/test/7**

Participant selection

The implementation team at Coventry University first selected all schools with primary-aged children in its geographical area—the West Midlands, where the project team is based. There was an attempt to recruit 60 schools in Coventry, Warwickshire, Solihull, Leicestershire, Northamptonshire, Walsall, Sandwell, Worcestershire, Birmingham, and the West Midlands. These are more areas than the ones originally stated in the protocol (Birmingham, Coventry, Warwickshire, Worcester and Solihull) in order to widen the net to recruit more schools. There were a total of 1,742 schools. Coventry University excluded schools that did not contain a Year 1 group, leaving them with 1,682 eligible schools all of which were invited to take part. In particular, an effort was made to encourage schools with disadvantaged intakes to participate during the recruitment stage. Coventry University ran a recruitment event where the project was explained to leaders from interested schools. Schools needed to sign a Memorandum of Understanding in order to participate. They were required to send an opt-out consent form to all parents ahead of testing in order to make a request for administrative data (the National Pupil Database) to the Department for Education. The consent was obtained before randomisation. The participation information sheet sent to parents is found in Appendix C. All of the approached schools that agreed to participate were included in the study.

Sample size

The calculations to decide on the sample size included in the protocol were performed using the Optimal Design (OD) Software (Spybrook *et al.*, 2011).[10] Our design paired schools with the aim of grouping them into clusters of similar schools. As Spybrook *et al.* (2011) point out, the aim of pairing is to reduce the heterogeneity within pairs (also called blocks or sites). This increases the precision of the treatment effect estimate and hence increases the power of the test. In the protocol (first row of Table 3) we had assumed that 60 schools would be recruited and paired according to characteristics.[11] Each school within each pair would be randomised to cohort 1 (to receive the treatment in Year 1) or cohort 2 (to receive the treatment in Year 2). On average, schools were assumed to have 60 pupils in their Year 1 cohort. The predictive power of baseline characteristics was set at 0.7, the intraclass correlation at 0.2, the power of the test at 0.80, and the statistical significance of the effect at 0.05. Further assuming that the proportion of the explained variance by the blocking variable is 0.10 and the effect size variability is 0.01, this sample size was powered to detect an effect of at least 0.2 standard deviations. We did not suggest increasing the size of the sample to take account of potential attrition as 60 schools was already an ambitious number to recruit for the experiment. See Table 3 in the next section for more information related to the minimum detectable effect size at randomisation and at the analysis stage.

As stated in the protocol, these are conservative estimates as they are based on a multisite cluster randomisation at the school level (where the sites are the pairs of schools). In practice, we also randomised pupils within schools. This was not incorporated in the calculations because the programme used (OD) does not have this option, but this further level of randomisation at the student level should enable us to detect smaller size effects. This is because compared to the multisite cluster randomisation, in our case we had a sequential randomisation process that randomised first schools and then students in treatment schools. The variance of the treatment effect in this case should be potentially smaller. Therefore, this should enable us to estimate more precise estimates and thus detect effect sizes smaller than those predicted with the OD programme for the multisite cluster randomised trial option.

[10] The programme is free and available to download here:
https://sites.google.com/site/optimaldesignsoftware/home

[11] Recruited schools were paired according to the following characteristics, with the following order of importance: size of the Year 1 cohort, Key Stage 1 average point score for the 2013 cohort, and percent of pupils that received free school meals at the time. Additionally, infant schools were paired together.

Randomisation

Paired randomisation was performed by the evaluation team at the Centre for Economic Performance (LSE) using the statistical programme STATA, whereas recruitment was carried out by the team at Coventry University, therefore ensuring independent allocation.

The list of variables used to pair schools included the size of the Year 1 cohort as supplied by the implementation team at Coventry University, Key Stage 1 average points score for the 2013 cohort, and a measure of the percentage of FSM pupils provided by the implementation team.[12] Additionally, infant schools were paired together

The random allocation sequence to assign schools to treatment or control groups was generated:

1) by first drawing pairs of schools (we drew numbers from a uniformly distributed random variable) and then
2) from each pair we assigned the school with the highest drawn number to the treatment group, and the school with the lowest drawn number to the control group.

The random allocation of pupils to the three different conditions in treatment schools was achieved by:

1) first, in each school, randomly allocating a number to each pupil in the school drawn from a uniformly distributed random variable, and then
2) in each school we assigned the first third of pupils with the lowest random numbers to the control condition, the second third of pupils with the lowest random numbers to the non-ICT condition, and the remaining third of pupils with the highest drawn numbers to the ICT condition.

Finally, the random allocation sequence to assign TAs in treatment schools to the ICT or non-ICT conditions was performed in a way analogous to the school randomisation. Codes are available upon request from the authors. A small number of deviations from this randomisation procedure are explained in the Participants subsection of the Impact Evaluation Section (see below).

Analysis

Our analysis follows the protocol and estimates Ordinary Least Squares (OLS) regressions for each outcome variable where the outcome is regressed against dummy variables for whether the pupils are in the ICT treatment or the non-ICT treatment. In this first analysis we control for the pairing dummies, as well as for prior attainment (as given by the baseline test for each outcome), gender, FSM eligibility, and a dummy variable indicating whether the pupil had a good level of development in the Foundation Stage Profile (FSP GLD).[13] Standard errors in the regression analysis are always clustered at the school level. This takes into account the fact that students are clustered in schools, and therefore outcomes might be correlated.

[12] The FSM measure provided by Coventry for the pairing was downloaded from the Edubase dataset, **http://www.education.gov.uk/edubase**. This FSM measure is defined as the percentage of pupils receiving FSM in that school year. This is calculated as the number of pupils known to be eligible for and claiming free school meals expressed as a percentage of all pupils. Includes pupils who have full time attendance and are aged 15 or under, or who have part time attendance and are aged between five and one.

[13] FSP GLD is a dummy variable that indicates whether the pupil has achieved a good level of development in the Foundation Stage Profile. This is the case if the pupil achieved a level of 2 or 3 in each of COM (Communication), PHY (Physical development), PSE (Personal, Social and Emotional Development), LIT (Language and Literacy) and MAT (Mathematical development) results.

This specification is close to that stated in the protocol. Specifically, the protocol stated that we would control for the following baseline characteristics: baseline test score, FSM eligibility, gender, ethnicity, and whether English is spoken as a first language. Unfortunately, the conditions for the supply of ethnicity and language data at the pupil level were not fulfilled and so could not be provided by the Department for Education.[14] The decision to include the FSP GLD variable is done in order to gain further precision.

All the outcome variables and baseline tests have been standardised to have mean 0 and standard deviation (SD) 1 using the mean and SD of the outcomes (we have used the mean and SD for the full sample for each of the outcomes, both at post-test and at baseline respectively). This allows us to interpret the coefficients of the explanatory variables in terms of standard deviations of the outcome variable. All regressions cluster standard error at the school level.

Our analysis has been on an intention-to-treat basis—as explained in the Participants subsection of the Impact Evaluation, five schools dropped out after randomisation—however, for four out of these five schools we were able to obtain post-test data and they are included in our analysis. We check whether baseline characteristics are balanced for the final sample to discern whether the small amount of attrition (losing one school) creates imbalance in the different intervention and control groups. We also check whether those pupils with missing data at follow-up (post-treatment stage) had different baseline characteristics than the pupils for which we have data at post-test. Since we do not find significant differences (see the Analysis section) we do not impute missing data and work with the available final sample.

As stated in the protocol, we explore whether there are heterogeneous effects of treatment on the main outcome variable according to FSM status, gender, and above-median attainment on pre-test. For example, for FSM, we do this by interacting each treatment variable with a dummy variable that indicates whether or not the pupil is eligible to receive FSM.[15]

Finally, the protocol specified that in order to address the possibility of spillover effects from treatment pupils to control pupils in treatment schools, we would run a first regression where the control group would be formed by pupils in the same schools, and a second regression where the control group would be formed by pupils in control schools only. We could then check whether the treatment effects were larger in the second regression than in the first one. If so, this is likely to be indicative of spillover effects. However, running regression one implies working only with treatment schools (the 24 clusters). The reduction in the number of clusters is rather problematic given that we need to cluster standard errors at the school level. Although we did this, we first run a regression where an additional dummy variable is added to the full sample for pupils that are in the control group within treatment schools (CT variable). We compare the coefficient on CT to those of the ICT and non-ICT treatment dummies.

Implementation and process evaluation

The principal aims of the process evaluation of ABRA were to understand the perceptions of the programme of teaching assistants and project leads and to identify factors affecting its successful implementation through qualitative research. The evaluation aimed to identify features that appeared to contribute to successful implementation, including practicalities and the influence of factors such as staff confidence in the approach. Qualitative research within the process evaluation also aimed at

[14] This is because the consent forms sent to parents ahead of testing were opt-out consent forms. In order to be able to receive sensitive NPD data, parents should have given explicit consent (opt-in consent forms).

[15] This variable comes from the National Pupil Database, in particular from the pupil Census. A pupil is recorded as eligible for Free School Meals on Census day. From 2012/13, this variable set to 'true' (‘1’) if a pupil has an FSM period with a start date and end date blank or end date on or after census date which means they are FSM eligible on Census day.

bringing greater clarity to the quantitative research findings and understanding the reasons for, and mechanisms behind, the estimated impacts.

The process evaluation aimed at understanding participants' experiences of the programme at the training stage and during implementation and application of ABRA within schools. It was designed to include observations of ABRA sessions to gain an insight into pupil engagement with the ICT and non-ICT interventions: the original plan was to carry out visits in two stages—initial and follow-up visits. However, given the demands on TAs it was agreed with the project team that email follow-up would be preferable. This was particularly necessary in view of the scale of process evaluation work carried out by the project team, which included surveys and the collection of recording and tracking data.

The process evaluation consisted of the following components:

- Evaluation of teacher preparation and training through:
    - attendance at the introductory conference at Coventry University in May 2014;
    - observation of teaching assistant training in both ICT and non-ICT interventions at Coventry University in October 2014;
    - analysis of project training materials issued to TAs; and
    - analysis of training evaluation data collected by the implementation team at Coventry University.

- Evaluation of implementation, fidelity, and impact through:
    - visits to six intervention schools in January 2015 to carry out
        - interviews with 16 TAs,
        - observe 13 ICT-based sessions and five non-ICT sessions, and to
        - analyse the project resources in the context of their use;
    - follow-up surveys of the six intervention schools in June 2015 resulting in responses from three;
    - analysis of survey on TA's background and experience carried out by Coventry University and completed by 46 of 49 TAs;
    - analysis of data on TAs' recording and tracking carried out by Coventry University;
    - analysis of data collected on TAs' experiences of the intervention carried out by Coventry University, completed by 37 of 49 TAs; and
    - analysis of data collected by Coventry University on the practices of treatment and control schools in relation to how they teach literacy completed by 13 of 21 active schools and 14 of 24 in the control group.

ABRA was implemented in the Midlands with many schools situated in Birmingham, Coventry and Leicester. NIESR selected six schools to include a range by size, location and performance as measured by Ofsted. Information about these schools and the size of the Year 1 ABRA cohort is presented in Appendix D. The visits and follow-up survey enabled us to gain a more in-depth understanding of the implementation of ABRA while our own assessment of the project team data—including tracking, evaluation of training, and usual practice in relation to teaching literacy—gave us a broader understanding of teachers' experiences across all schools involved in the project.

Data, collected through the range of approaches described above, was analysed using thematic analysis within a framework approach.[16] This enables the analysis of qualitative data in a written form and is therefore appropriate for the analysis of transcripts of interviews with teaching assistants and project leads, as well as research notes taken during observation of training. Qualitative responses to training evaluation surveys were also analysed in this way. The method entails coding the data into

[16] Ritchie J. and Lewis J. (2003) *Qualitative research practice: a guide for social science students and researchers*, London: Sage.

themes and issues. In this case, codes were a mixture of predetermined ones, developed during the design of the process evaluation and taking account of the aims of the intervention, as well as those that emerged from the text of transcripts and observations. Codes identified different types of information, for example, more tangible ones—such as previous use of phonics programmes, experiences of the training, and of putting the approach into practice—as well as others such as values and feelings. Throughout the analysis process we looked for similarities and differences in the data. The framework approach allows for tracts of text to be classified under more than one code, and codes were, in some cases, amalgamated to form wider groups, particularly where substantial issues were concerned.

The codes and groups developed in the analysis of data formed the analytical framework and were used to structure the findings into a preliminary report. The framework was formed from the data from case-study visits which was then used to structure findings from the follow-up survey, project team data, and training evaluation. In the report, where relevant, we identify the source of a finding.

In the final stage of analysis we re-structured our document to follow the format required by the EEF. This involved structuring the findings using the main EEF process evaluation criteria: implementation, outcomes, and fidelity.

Timeline

Table 2: Timeline

Date	Activity
January–April 2014	Coventry team invites schools ahead of a recruitment event that will take place the week after Easter half-term.
January–March 2014	Coventry team develops treatment curricula for both treatments and sends to CEP for review. CEP develops an assessment form for these curricula to send back to Coventry.
March–April 2014	The EEF organises an anonymous process of peer review of the two curricula.
Autumn 2014 (and beginning of 2015 to correct mistakes in pupil identifiers and dates of birth)	CEP applies for the National Pupil Database for relevant cohorts of pupils in all participating schools.
September 2014	Coventry organises baseline tests in all participating schools and arranges for a list of pupils in Year 1 to be sent to CEP.
October 2014	CEP randomises schools into cohort 1 (treatment in Year 1) and cohort 2 (control). CEP uses year group lists to randomise pupils in treatment schools to groups A, B and C. Coventry sends CEP the CVs of all teaching assistants taking part in the experiment (November 2014). CEP randomises teaching assistants to groups A and B within treatment schools. Coventry organises training of teaching assistants. Attended by NIESR and a team member of CEP.
October 2014–May 2015	Coventry takes responsibility for implementation of treatment in groups A and B in treatment schools.
January 2015	NIESR conducts first round of process evaluation visits to six selected schools.
June–July 2015	Coventry organises second round of testing in all participating schools.
July 2015	NIESR collects follow-up data from six selected schools through a survey.
September 2015–July 2016	Coventry organises implementation of programme in schools participating in cohort 2.
From July 2015 onwards	CEP works on the EEF evaluation report (to be extended with an addendum report when KS1 outcomes become available in 2016). The final set of data was received by mid-December 2015. CEP and Coventry to work on quantitative evaluation with a view to publication in academic outlets. NIESR analyses data from interviews and observations, from training evaluation and from teaching assistant logs for the process evaluation section of the EEF report.

Costs

The cost information has been collected by the implementation team at Coventry (that is, data has been collected from the developer). The interventions cost the schools nothing in the trial. In the Costs section

(after the Results section of this report), we provide the costs associated with implementing each intervention had the EEF not funded it. This gives detailed information on costs per pupil, and how the current costs of this intervention could be further reduced.

3. Impact evaluation

Participants

As described in the Participant Selection subsection, the implementation team at Coventry University first selected all schools with primary-aged children in the geographical areas near to them. There were a total of 1,742 schools in the selected geographical areas. Coventry University excluded schools that did not contain a Year 1 group, leaving them with 1,682 eligible schools. Of the schools that attended the recruitment session, 58 signed contracts to participate in the intervention. Six schools dropped out prior to baseline testing (due to changes in circumstances) and one school dropped out after baseline testing as this school found the testing too disruptive.

A total of 51 schools were part of the randomisation (randomisation took place after baseline testing had concluded). One of the schools randomised is a private school. In the paired randomisation, this private school was paired to a blank (non-existent) school in order to decide whether to assign that private school to treatment or control. (This school is not included in the analysis.)

Among the 26 treatment schools, the initial number of pupils to be randomised into the three conditions was 1,093. Randomisation resulted in the three following groups: ICT treatment (373 pupils), non-ICT treatment (365 pupils), and control pupils in treatment schools (355 pupils). However, post randomisation and before the treatment started, some changes were requested by the implementation team at Coventry in order to reduce drop-out of entire treatment schools. As evaluators we were aware of the problems of re-randomising for some particular schools however we granted these changes so that power would not be further compromised. These few changes concerned the following:

a. Two schools had classes in Year 1 in different school locations. We were asked to re-randomise pupils to be able to form groups that would come from the same locations (otherwise, it would have been impossible for them to form the treatment groups).
b. We were asked to reduce the number of pupils in the treatment conditions or the school would drop out. This happened in two cases (two schools). In one case, ten pupils had been assigned to each treatment and we had to reduce the number to four in each treatment instead. In the second case, there were 14 pupils initially allocated to the ICT and non-ICT conditions respectively. We were asked to reduce the number of pupils in each condition to 12.
c. A small number of pupils had left the school or had withdrawn consent so they should not have been included in the randomisation. We randomly assigned pupils from the original control group to treatment only when schools pressed to have treatment groups of equal size. This happened for two schools (one pupil in each school was randomly chosen from the original control condition to join the corresponding treatment group).

The number of pupils in the sample of treatment schools after these changes was 1,083 (we lost ten pupils: four because they had withdrawn consent and six school leavers). The final distribution of pupils in treatment schools before the start of treatment was: ICT treatment (360 pupils), non-ICT treatment (350 pupils), and control pupils in treatment schools (373 pupils). We do not believe that these changes represent a threat to validity for two main reasons. First, the number of students involved in the changes is small. Second, all the changes described in points a, b, and c above were done by allocating pupils at random to meet the requirements.

There were initially 1,158 pupils in control schools. Together with the 1,083 pupils randomised to the three different conditions in treatment schools, this made an initial sample of 2,241 pupils.

A small number of pupils—26 pupils from treatment schools and 35 pupils from control schools—did not sit the baseline PIRA test (pupils sat several tests in different dates, both for baseline and post-intervention tests—we focus here on PIRA, the primary outcome).[17]

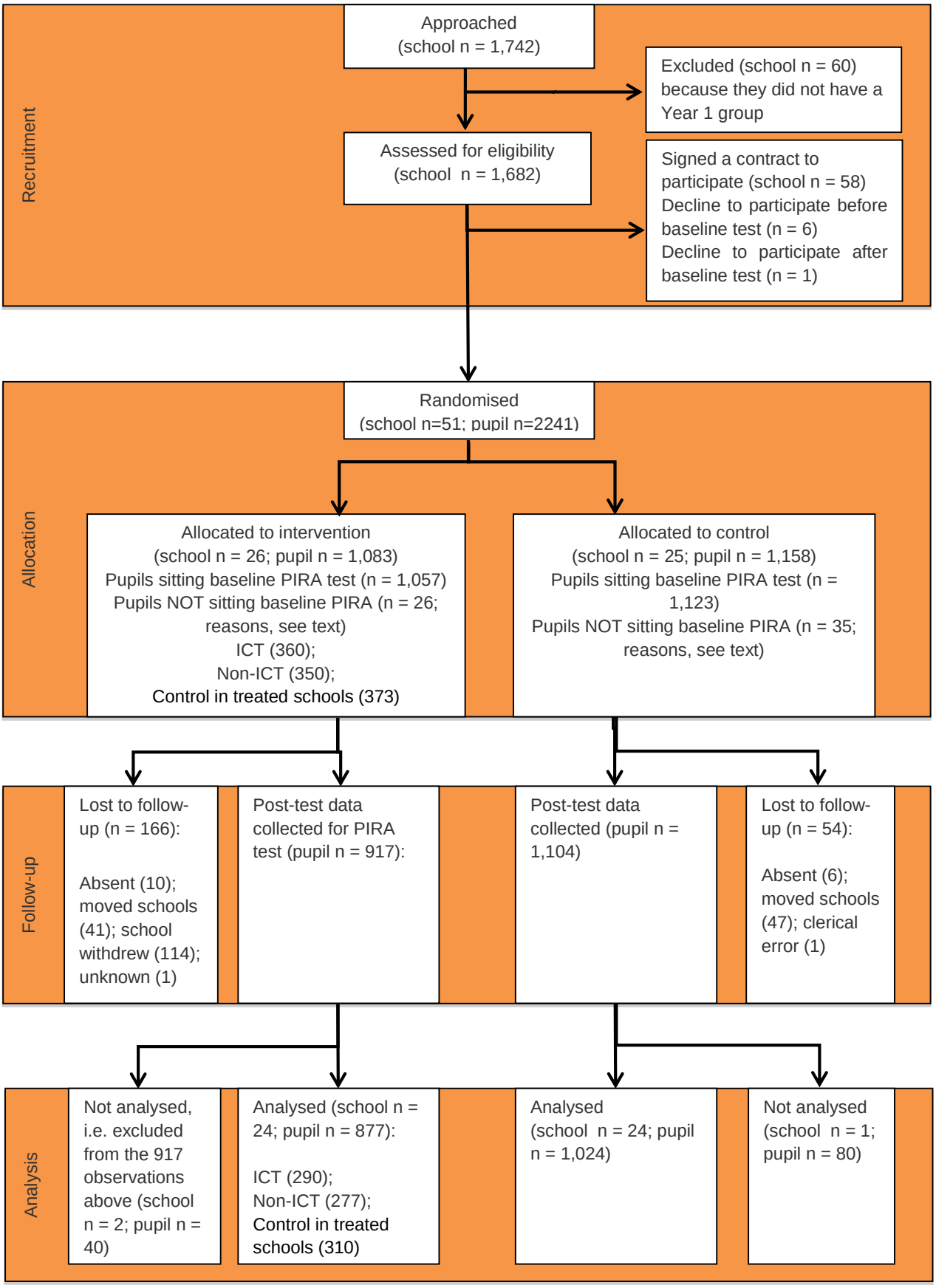
Likewise, not all pupils in the randomised sample sat the post-intervention PIRA test (the follow up tests after the intervention had finished). From the original 1,083 (1,158) pupils in the treatment (control) sample, 977 (1,104) pupils sat the post-intervention PIRA test. The reasons why some pupils did not sit the test can be found in Figure 1.

A total of five schools dropped out from the intervention post-randomisation, all of them in the treatment group. Three schools dropped out right after randomisation took place (two because they could not see how to integrate the intervention with their current literacy provision and worried that the children might get confused, and one for unknown reasons), and two dropped out during the intervention (one in November 2014 due to staffing, the other one in February 2015 due to a change in the head teacher). We advised the implementation team to collect data for all initially randomised schools. They were able to collect post-intervention data for four out of the five schools that dropped out after randomisation. This allows us to perform an intention-to-treat analysis using most of the original randomised schools. However, given that we used paired randomisation, we remove from the main analysis both the school for which we did not get any post-test data and its pair, as well as the independent school that was paired to a blank school. We are aware that this decision could introduce selection bias and reduce statistical power. However, when we include the full sample of schools (with the exception of the one school for which we don't have any post-intervention test data), and exclude the pairing dummies as controls, the results remain virtually unchanged (results available upon request).

As a result of excluding these three schools from the main analysis, and also from having to exclude pupils without PIRA tests both at both baseline and follow-up, we are left with a total of 1,901 pupils (877 in treatment schools and 1,024 in control schools), almost 85% of the randomised sample. If we were to include all pupils for whom we have pre- and post-intervention test data, we would work with 1,969 observations, almost an 88% of the randomised sample. Additionally, the FSM indicator (measuring whether pupils were FSM eligible on Census day) provided by the National Pupil database is only available for 1,887 pupils, almost 84% of the randomised sample. Thus, attrition in this study is lower than 20% as 84% of the randomised pupils were available in the final analysis.

[17] The reasons why pupils did not take the baseline PIRA test in treatment or control schools included absence at the time of the test (14/28), being unable to access test due to behaviour, EAL status, SEN status, and so on (3/4), blank test returned (might be because pupil refused to take the test or was absent) (6/0), children leaving school (1/0), unknown reasons (2/2), and clerical error (0/1).

Figure 1: Number of schools and pupils at different stages of the RCT



With these numbers, and making some assumptions, we can calculate the minimum detectable effect size at different stages of the trial. Table 3 summarises these calculations, performed using the Optimal Design Software (Spybrook *et al.*, 2011).[18] Our design paired schools with the aim of grouping them into clusters of similar schools. The list of variables used to pair schools includes the size of the Year 1 cohort as supplied by the implementation team at Coventry University, Key Stage 1 average points score for the 2013 cohort, and a measure of the percentage of pupils classified as FSM-eligible provided by the implementation team. Additionally, infant schools were paired together.

As Spybrook *et al.* (2011) point out, the aim of pairing is to reduce the heterogeneity within pairs (also called blocks or sites). This increases the precision of the treatment effect estimate and hence increases the power of the test. In the protocol (first row of Table 3) we had assumed that 60 schools would be recruited and paired according to characteristics, and that each school within each pair would be randomised to cohort 1 (treatment in Year 1) and cohort 2 (treatment in Year 2). On average, schools were assumed to have 60 pupils in their Year 1 cohort. The predictive power of baseline characteristics was set at 0.7, the intraclass correlation at 0.2, the power of the test at 0.80, and the statistical significance of the effect at 0.05. Further assuming that the proportion of the explained variance by the blocking variable is 0.10 and the effect size variability is 0.01, this sample size was reasonable to detect an effect of at least 0.2 standard deviations. Under the same assumptions, but considering the number of schools at randomisation, this design allows detecting an effect of at least 0.22 standard deviations. Finally, with the data analysed, we can calculate some of the assumed parameters (see information in the table). With these new values, we can detect effects of at least 0.27 standard deviations with a power of 0.80.

These are conservative estimates as they are based on a multi-site cluster randomisation at the school level (where the sites are the pairs of schools). In practice, we have also randomised pupils within schools. As stated in the protocol, within the schools assigned to the treatment in Year 1, pupils have been randomly allocated into three groups: (A) support from teaching assistants using the computer programme (ICT intervention), (B) support from Teaching Assistants using other methods (non-ICT intervention), or (C) a 'business as usual' control group. This should enable us to detect smaller size effects, as explained in more detail in the section above.

[18] The programme is free and available to download here: https://sites.google.com/site/optimaldesignsoftware/home

Table 3: Minimum detectable effect size at different stages

Stage	N [schools/pupils] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	Schools: 60 (30; 30) Pupils in Year 1: 60 on average (by school)	0.7	0.20	School blocking (30 blocks with 2 schools by pair)	80%	0.05	0.20
Randomisation	Schools: 50 (25; 25) Pupils in Year 1: 60 on average (by school)	0.7	0.20	School blocking (25 blocks with 2 schools by pair)	80%	0.05	0.22
Analysis (available pre- & post-test)	Schools: 48 (24; 24) Pupils in Year 1: 40 on average (by school)	0.43	0.15	School blocking (24 blocks with 2 schools by pair)	80%	0.05	0.27

Pupil characteristics

Table 4 shows whether pupil characteristics were balanced at baseline for the analysis sample using the maximum number of observations available for each variable. Outcome variables are standardised to have a mean of 0 and a standard deviation of 1 (enabling us to interpret estimates in terms of standard deviation impacts). Each column shows the results of regressing each characteristic on a dummy variable that takes on the value of 1 if the pupil is enrolled in a treatment school. The constant, therefore, shows the average value of that characteristic in control schools, and the dummy variable coefficient picks up the difference for treatment schools. Columns 1 and 2 show that there are no significant differences between pupils in treatment and control schools in terms of gender and FSM eligibility, and column 4 shows that the pupils who did not sit the PIRA test at baseline (before randomisation) are not missing differentially in treatment and control schools. Column 3 shows that pupils in treatment schools were less likely to have achieved a good level of development (GLD) in the Foundation Stage Profile (FSP). Controlling for the FSP development level between treatment and control pupils has only a minimal effect on point estimates for treatment effects. However, it does improve the precision of the estimates. Thus we include this as an additional control above that specified in the protocol.

Table 5 repeats the same exercise using all the outcome variables measured at baseline. In none of these cases do we find significant differences in the baseline tests between pupils in treatment and control schools.

Table 4: Balance at baseline (final sample)—pupil characteristics

	(1) Female	(2) FSM eligibility	(3) FSP GLD	(4) Missing PIRA at baseline
=1 if school is a treatment school	-0.005 (0.017)	-0.000 (0.023)	-0.087*** (0.022)	-0.005 (0.008)
Constant	0.519*** (0.019)	0.458*** (0.133)	0.455*** (0.044)	0.021* (0.011)
Observations	1,901	1,887	1,893	2,241

Note: Standard errors clustered at the school level in parentheses (* p < 0.10; ** p < 0.5; *** p < 0.01). FSM eligibility: pupil recorded as eligible for free school meals on Census day. FSP GLD: pupil has achieved a good level of development—achieved level of 2 or 3 in each of COM, PHY, PSE, LIT and MAT results (see main text). All regressions include pairing dummies (schools were paired for randomisation).

Table 5: Balance at baseline (final sample) – Outcome variables measured pre-randomisation

	(1) PIRA	(2) DTWRP	(3) BPVS	(4) PS	(5) LK	(6) LEST
=1 if school is a treatment school	-0.061 (0.068)	0.001 (0.056)	-0.101 (0.066)	0.046 (0.048)	0.030 (0.063)	-0.061 (0.061)
Constant	-0.383*** (0.085)	-0.026 (0.176)	-0.139 (0.134)	-0.139*** (0.050)	0.039 (0.085)	0.111 (0.251)
Observations	1,901	1,897	1,872	1,929	1,929	1,916

Note: Standard errors clustered at the school level in parentheses (* p < 0.10; ** p < 0.5; *** p < 0.01). All regressions include pairing dummies (schools were paired for randomisation). All tests (outcome variables) are standardised at the baseline level to have mean 0 and SD 1, using all the observations available at baseline.

The last row in Table 5 shows the number of observations available in the final sample for each of the outcome variables. As documented above, there are 1,901 pupils in 48 schools (clusters) available for the main outcome (the PIRA test), and this represents almost 85% of the randomised sample. For the other secondary outcomes, the number of pupils available varies slightly with respect to the main outcome sample as can be seen in the last row of Table 5. The number of observations available represents 85% (DTWRP), 84% (BPVS), 86% (PS), 86% (LK) and 85% (LEST), respectively, of the randomised sample. Table 6 explores whether pupils who were unable to be followed up from baseline tests are different from those for whom we do have post-treatment data. Columns 1 to 4 show that there are no statistically significant differences in terms of gender, FSM eligibility, FSP development level, and PIRA results at baseline. Moreover, column 5 shows that pupils that we could not follow up do not differ in terms of baseline PIRA results in treatment and control schools (see the coefficient for the interaction between the variable 'Missing pira at post-test' and the dummy variable for treatment schools). Finally, column 6 repeats the analysis in column 5, distinguishing between the two types of treatment (ICT and non-ICT). This shows that pupils who do not have a post-intervention test score are no different from other pupils, conditional on what group to which they are assigned. Given these findings, the tables shown in the next section use the final available sample without imputing missing values.

Table 6: Analysis of missing data at follow-up

	(1) Female	(2) FSM eligibility	(3) FSP GLD	(4) PIRA at baseline	(5) PIRA at baseline	(6) PIRA at baseline
Missing PIRA at post-test	-0.009 (0.036)	-0.087 (0.076)	0.018 (0.047)	0.064 (0.094)	0.144 (0.116)	0.068 (0.091)
=1 if school is a treatment school					-0.043 (0.074)	
Missing PIRA at post-test* (= 1 if school is a treatment school)					-0.096 (0.189)	
ICT						-0.053 (0.082)
NONICT						-0.043 (0.074)
ICT* (= 1 if school is a treatment school)						-0.011 (0.151)
NONICT* (= 1 if school is a treatment school)						0.031 (0.139)
Constant	0.482*** (0.020)		0.482*** (0.037)	-0.027 (0.155)	0.003 (0.167)	-0.013 (0.157)
Observations	2180	2143	2149	2180	2180	2180

Note: Standard errors clustered at the school level in parentheses (* p < 0.10; ** p < 0.5; *** p < 0.01). Missing PIRA at post-test is a dummy variable that takes on the value of 1 if we don't observe PIRA at post-test. FSM eligibility: pupil recorded as eligible for free school meals on Census day. FSP_GLD: has achieved a good level of development—achieved level of 2 or 3 in each of COM, PHY, PSE, LIT and MAT results. All regressions include pairing dummies (schools were paired for randomisation). PIRA at baseline is standardised to have mean 0 and SD 1. ICT and NON-ICT are the treatment dummies.

Finally, in Table 7 we provide the baseline demographic, school-level, attainment, and other relevant characteristics of the final sample for the intervention and control schools. The upper panel of Table 7 presents means of several school characteristics for treatment and control schools, all measured with data for the academic year 2013/2014 and for the final (analysis) sample. In particular, information is given on the type of school, whether the school is rural or urban, Ofsted overall effectiveness according to the last inspection until August 2014, the IDACI rank, and the Index of Multiple Deprivation. The bottom panel presents information on other continuous variables at the school level (percentage FSM, SEN, and so on). Although there are some differences between the treatment and control groups (such as in the distribution of schools according to the IDACI rank or Index of Multiple Deprivation), there is a reasonable and broadly comparable spread of schools within the categories considered here.

Table 7: Baseline comparison (school-level variables for the final sample, 2013/2014 data)

Variable	Intervention group		Control group	
School-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Type of school				
Academy (converter/sponsor led)	3/24 (0)	12.5%	6/24 (0)	25%
Community school	11/24 (0)	45.8%	9/24 (0)	37.5%
Voluntary (aided/controlled)	10/24 (0)	41.7%	9/24 (0)	37.5%
Urban vs rural				
Urban > 10k (less sparse)	21/24 (0)	87.5%	20/24 (0)	83.3%
Town and fringe/village (less sparse)	3/24 (0)	12.5%	4/24 (0)	16.7%
Ofsted overall effectiveness				
Outstanding	4/24 (0)	16.7%	3/24 (0)	12.5%
Good	14/24 (0)	58.3%	15/24	62.5%
Requires improvement	6/24 (0)	25%	6/24	25%
IDACI rank (1 most deprived)				
1 (0 to quartile 1)	8/24 (0)	33.3%	8/24 (0)	33.3%
2 (quartile 1 to quartile 2)	7/24 (0)	29.2%	3/24 (0)	12.5%
3 (quartile 2 to quartile 3)	4/24 (0)	16.7	6/24 (0)	25%
4 (from quartile 3 above)	5/24 (0)	20.8%	7/24 (0)	29.2%
Index of multiple deprivation (1 most deprived)				
1 (until quartile 1)	10/24 (0)	41.7%	8/24 (0)	33.3%
2 (quartile 1 to quartile 2)	5/24 (0)	20.8%	4/24 (0)	16.7%
3 (quartile 2 to quartile 3)	5/24 (0)	20.8%	7/24 (0)	29.2%
4 (from quartile 3 above)	4/24 (0)	16.7%	5/24 (0)	20.8%
School-level (continuous)	**n (missing)**	**Mean**	**n (missing)**	**Mean**
Number of Year 1 pupils	24 (0)	47	24 (0)	49
Number of Key Stage 1 pupils	24 (0)	91	24 (0)	95
Percentage of pupils known to be eligible for free school meals	24 (0)	20	24 (0)	21
Percentage of registered pupils with SEN (without a statement)	24 (0)	15	24 (0)	15
Percentage of pupils whose first language is known or believed to be other than English	24 (0)	25	24 (0)	15
Percentage of pupils classified as white British ethnic origin	24 (0)	63	24 (0)	73
Measures of school attainment				
Cohort level Key Stage 1 average points score	21 (3)	15.17	20 (4)	15.09
% pupils low KS1 attainment	21 (3)	19.2%	20 (4)	19.2%
% pupils med KS1 attainment	21 (3)	54.8%	20 (4)	58.9%
% pupils high KS1 attainment	21 (3)	26.0%	20 (4)	21.8%

Outcomes and analysis

The first table in this section (Table 8) shows the raw means for the two treatment groups (the ICT and the non-ICT condition) and the two controls groups (control pupils in intervention schools, and control pupils in control schools) together with 95% confidence intervals for these means. The first column shows the number of observations available at follow-up, that is, the number of pupils for which we do have a result in the post-intervention test for each of the outcome variables. The second column shows the number of pupils that were part of the randomisation and for which test data is missing at follow up. In general, the raw means for groups in treatment schools in the follow-up test are higher than the means for the group of pupils in control schools (except for the BPVS test).

Table 8: Raw means for all outcomes in the different intervention and control groups

	n*	(missing at follow up)**	Raw means	(95% CI)
Outcome variable: PIRA				
ICT	306	54	98.60	(97.13 , 100.06)
Non-ICT	290	60	99.72	(98.16 , 101.27)
Control pupils in treatment schools	321	52	98.52	(97.06 , 99.96)
Control pupils in control schools	1,104	54	97.29	(96.50 , 98.07)
Outcome variable: DTWRP				
ICT	307	53	109.86	(108.23 , 111.48)
Non-ICT	288	62	109.82	(108.00 , 111.63)
Control pupils in treatment schools	319	54	109.85	(108.13 , 111.55)
Control pupils in control schools	1,102	56	108.59	(107.65 , 109.53)
Outcome variable: BPVS				
ICT	304	56	89.77	(88.47 , 91.05)
Non-ICT	284	66	91.33	(89.98 , 92.67)
Control pupils in treatment schools	315	58	90.62	(89.39 , 91.83)
Control pupils in control schools	1,094	64	92.94	(92.26 , 93.61)
Outcome variable: PS				
ICT	307	53	13.42	(13.01 , 13.83)
Non-ICT	289	61	13.67	(13.23 , 14.11)
Control pupils in treatment schools	319	54	13.26	(12.85 , 13.66)
Control pupils in control schools	1,103	55	13.43	(13.22 , 13.64)
Outcome variable: LK				
ICT	307	53	11.71	(11.56 , 11.85)
Non-ICT	289	61	11.64	(11.46 , 11.81)
Control pupils in treatment schools	319	54	11.75	(11.58 , 11.90)
Control pupils in control schools	1,103	55	11.48	(11.38 , 11.57)
Outcome variable: LEST				
ICT	307	53	0.83	(0.72 , 0.94)
Non-ICT	288	62	0.81	(0.69 , 0.92)
Control pupils in treatment schools	318	55	0.79	(0.67 , 0.90)
Control pupils in control schools	1,103	55	0.66	(0.59 , 0.72)

*(n) Number of pupils for which post-treatment test results are available.

** Missing values calculated from subtracting (n) to the number of pupils in the samples available at randomisation (for more information, see subsection Participants in Section 3).

The impact of the different interventions is summarised in Table 9 for the main outcome variable, the Progress in Reading Assessment (PIRA) test. As stated above, this is a test that evaluates the general reading ability of pupils. The specifications in the three different columns make use of the maximum

number of observations available. All the columns regress the standardised PIRA measure for the follow-up test on the ICT and non-ICT treatment dummies, and include dummies for the randomisation pairs. As explained in the Participants subsection, this is why the number of schools shown in the second to last row of the table falls to 48 schools. Standard errors are clustered at the school level in all regressions. For each of the variables, the first row shows the value of the coefficients (that should be interpreted in terms of standard deviations (SD) of the outcome variable), the second row shows the p-values in parentheses, and the third row shows 95% confidence intervals.[19] The control group in all regressions consists of all control children (that is, control children in treated schools and control children in control schools). Results using control children in control schools only will be shown later in the section.

The first column in Table 9 shows the results of a specification that only controls for the treatment and pairing dummies. The second column includes a control for the standardised baseline PIRA test. The third column further controls for gender, FSM, and whether the pupil achieved a good level of development in the Foundation Stage Profile (FSP GLD). As we progress from columns 1 to 3 there is a small increase in the point estimates of the treatment effects and a marked increase in precision. Both these changes are more important for the ICT treatment. The results in the most detailed specification show the ICT and the non-ICT treatments to have a positive effect on the general reading ability of pupils when compared to all control pupils. The ICT treatment increases the general reading ability at post-test by almost 14% of a SD whereas the non-ICT treatment increases it by 23% of a SD. The last row shows the p-value from testing the null hypothesis that the ICT and non-ICT coefficients are equal. The p-value suggests that we can marginally reject the null hypothesis that the effects of the two different treatments are equal (at the 10% level).

This most detailed specification in column 3 is our preferred specification and will be the one used throughout the analysis in this section. The section on Methods (in particular, the subsection Analysis) compares this final specification with the one stated in the protocol.

[19] Given that the information provided by the p-value and confidence interval is very similar, this third row is only added in this first table.

Table 9: Analysis for the main outcome variable

Outcome: Standardised PIRA	(1) Pairs	(2) + Baseline test	(3) +FEM, FSM, and FSP GLD
ICT	0.074 (0.321) [-0.074, 0.222]	0.116 (0.112) [-0.028, 0.260]	0.138** (0.044) [0.004, 0.273]
Non-ICT	0.171* (0.052) [-0.001, 0.344]	0.218*** (0.003) [0.080, 0.356]	0.231*** (0.001) [0.102, 0.360]
PIRA baseline (standardised)		0.620*** (0.000) [0.568, 0.671]	0.470*** (0.000) [0.422, 0.519]
Female			-0.014 (0.704) [-0.085, 0.058]
FSM eligible (SPR14)			-0.175*** (0.001) [-0.276, -0.074]
FSP GLD			0.521*** (0.000) [0.426, 0.617]
Constant	-0.491*** (0.000) [-0.561, -0.420]	-0.193** (0.016) [-0.349, -0.037]	-0.403*** (0.000) [-0.497, -0.309]
Pupils	1,952	1,901	1,884
Missing pupils from randomisation	289	340	357
Schools	48	48	48
Ho: ICT-NONICT=0 (p-value shown)	0.335	0.104	0.103

Note: Standard errors clustered at the school level. P-values in parentheses (* p < 0.10; ** p < 0.5; *** p < 0.01), 95% confidence intervals in brackets. ICT and non-ICT are the treatment dummies. All regressions include pairing dummies (schools were paired for randomisation). Both the PIRA outcome variable and PIRA at baseline are standardised to have mean 0 and SD 1, so the coefficients are interpreted in terms of SD of the outcome variable.

As stated in the protocol, Table 10 explores whether there are heterogeneous effects of treatment on the main outcome variable according to FSM status, gender, and whether the pupil has a score above median attainment on the pre-test. For example, when we consider whether effects are heterogeneous by FSM status, we interact each treatment variable with a dummy variable that indicates whether the pupil is eligible to receive FSM. All regressions control for the set of controls included in the preferred specification (not shown in Table 10). To ease comparability, the first column reproduces the main results (i.e. column 3 of Table 9, for the coefficients of interest). The presence of heterogeneous effects has also been explored by running separate regressions for each condition (FSM and non-FSM pupils; males and females; and above and below median baseline attainment). The results remain virtually unchanged.

The results in the second column of Table 10 show that the impact of the intervention is bigger for disadvantaged pupils (as measured by the FSM eligibility indicator). The magnitude of both the ICT and non-ICT dummies increase considerably for FSM pupils: ICT and non-ICT treatments increase general

reading ability by almost 37% and 40% of a SD for FSM pupils respectively. Non-FSM pupils in the non-ICT condition still benefit from the intervention (their results on the PIRA test post-treatment are 18% SD higher compared to the pupils in the control groups). However, non-FSM pupils in the ICT treatment condition do not seem to benefit from the intervention. The last two rows of column 2 show statistically significant differences between the FSM and non-FSM coefficients for both the ICT and non-ICT interactions.

Column 3 repeats the exercise with respect to gender. The results suggest that both males and females benefit equally from the ICT intervention. In fact, both coefficients barely differ from the ICT coefficient in the main results table (shown here in column 1). Females seem to benefit more from the non-ICT intervention than males, although both groups benefit substantially (27% and almost 19% of a SD for females and males respectively). However, the results in the last row of Column 3 suggest that these coefficients are not statistically different from each other.

Finally, column 4 examines the interaction between the intervention dummies and the attainment of pupils in the PIRA test before treatment. Baseline attainment in the PIRA test is measured with a dummy variable that indicates whether the pupil's baseline test results are above or below the median. The results indicate that the non-ICT intervention similarly benefited below- and above-median pupils (the p-value shown in the last row of column 4 suggests that these coefficients are not statistically different from each other). However, those pupils with below-median prior attainment benefited from the ICT intervention whereas this intervention did not affect above-median pupils. In this case, the null hypothesis that the coefficients for the two ICT subgroups are equal can be rejected at the 10% level.

Table 10: Analysis of heterogeneous effects of the interventions on the main outcome variable

Outcome: Standardised PIRA	(1) Final specif	(2) FSM interaction	(3) Outcome: Standardised PIRA	Gender interaction	(4) Outcome: Standardised PIRA	Pre-test interaction
ICT	0.138** (0.044)					
NONICT	0.231*** (0.001)					
ICT*FSM		0.368** (0.011)	ICT*Female	0.139* (0.085)	ICT* (> median)	0.055 (0.383)
ICT*NOFSM		0.081 (0.217)	ICT*Male	0.139* (0.087)	ICT* (< median)	0.215** (0.022)
NONICT*FSM		0.396*** (0.000)	NONICT*Female	0.273*** (0.002)	NONICT* (> median)	0.235*** (0.001)
NONICT*NOFSM		0.181** (0.014)	NONICT*Male	0.187** (0.033)	NONICT* (< median)	0.230** (0.027)
Pupils Schools	1884 48	1884 48	Pupils Schools	1884 48	Pupils Schools	1884 48
Ho: ICT FSM - ICT NOFSM=0		0.035	Ho: ICT Female - ICT Male=0	0.999	Ho: ICT Above - ICT Below=0	0.068
Ho: NONICT FSM - NONICT NOFSM=0		0.000	Ho: NONICT Female - NONICT Male=0	0.421	Ho: NONICT Above - NONICT Below=0	0.964

Note: Standard errors clustered at the school level. P-values in parentheses (* p<0.10; ** p<0.5; *** p<0.01). ICT and NONICT are the treatment dummies. All regressions include pairing dummies (schools were paired for randomisation) and the controls of the main specification: PIRA baseline (std), Female, FSM eligible, and FSP GLD. The PIRA outcome variable is standardised to have mean 0 and SD 1. '> median' and '< median' stand for above and below median pre-treatment PIRA attainment.

In Table 11 we address whether there have been spillover effects from treatment pupils to control pupils in treatment schools. The protocol specified that we would run a regression where the control group would be formed by pupils in the same schools, and a second regression where the control group would be formed by pupils in control schools only. We could then check whether the treatment effects were larger in the second regression than in the first one. This would be indicative of spillover effects. Although we do this, we also run a regression where we simply include an additional dummy variable for pupils that are in the control group of treatment schools (denoted by CT). The advantage of this approach is that it produces similar insights without having to reduce the number of clusters (which is necessary when running regressions for treatment schools only: 24 schools). In Table 11, column 1 replicates our preferred specification for the main outcome (as shown in column 3 of Table 9, showing coefficients on the treatments). In column 2 we include a dummy variable (CT) indicating whether pupils are in the control condition within treatment schools. Column 3 uses only the observations related to pupils in treatment schools and column 4 shows results where we exclude pupils assigned to the control condition within treatment schools. The coefficient CT in column 2 shows that control pupils in treatment schools did significantly better at post-test than control pupils in control schools. In fact, they did just as well as those exposed to the ICT treatment. However, both groups are still out-performed by those exposed to

the non-ICT treatment. A very similar picture emerges in column 3. Column 4 presents results where we exclude pupils assigned to the control condition within treatment schools (in other words, the control group is only comprised of pupils in control schools). The magnitudes of both the coefficients on the ICT and non-ICT treatments increase considerably in this case. Along with the results in columns 2 and 3, these findings are consistent with the presence of spillover effects from treatment pupils to control pupils in treatment schools, although the mechanism for these spillovers is not clear. A potential explanation for spillover effects is that teaching assistants interacted with pupils on other occasions. They were not only employed to implement this experiment. Potentially, the training enhanced their ability to help pupils at other times. This would have affected the control group as well as the treatment groups.

Table 11: Spillover effects

Outcome: Std PIRA	(1) Final specif	(2) +Control pupils in treated schools	(3) Only pupils in treated schools	(4) Only controls in control schools
ICT	0.138** (0.044)	0.179** (0.027)	0.013 (0.792)	0.186** (0.025)
NONICT	0.231*** (0.001)	0.272*** (0.001)	0.080* (0.092)	0.284*** (0.001)
CT		0.167** (0.028)		
PIRA baseline (std)	0.470*** (0.000)	0.469*** (0.000)	0.523*** (0.000)	0.478*** (0.000)
female	-0.014 (0.704)	-0.011 (0.769)	0.078 (0.123)	-0.044 (0.272)
FSMeligible_SPR14	-0.175*** (0.001)	-0.178*** (0.001)	-0.168*** (0.001)	-0.161*** (0.006)
FSP_GLD	0.521*** (0.000)	0.531*** (0.000)	0.428*** (0.000)	0.517*** (0.000)
Constant	-0.403*** (0.000)	-0.437*** (0.000)	-0.243*** (0.000)	-0.407*** (0.000)
Pupils Schools	1,884 48	1,884 48	870 24	1,576 48
P-values Ho: ICT-NONICT=0 Ho: ICT-CT= 0 Ho: NONICT-CT=0	0.103	 0.102 0.820 0.039	0.244	0.092

Note: Standard errors clustered at the school level. P-values in parentheses (* p < 0.10; ** p < 0.5; *** p < 0.01). ICT and NONICT are the treatment dummies. CT is a dummy that equals 1 for pupils in the control group of treatment schools. All regressions include pairing dummies (schools were paired for randomisation). Both the PIRA outcome variable and PIRA at baseline are standardised to have mean 0 and SD 1.

Finally, Table 12 shows the results using the preferred specification (that is, controlling for pairing dummies, gender, FSM eligibility and FSP GLD) for all outcome variables. The number of observations differs between columns because we have included all available observations for each outcome variable. However, if we restrict the observations in these regressions to include those pupils for which we observe all six outcome variables, the results are almost the same. As in Table 9, the control group in all regressions in this table consists of all control children (those in treatment schools and those in control schools).

Of the five secondary measures analysed, the largest impact of the intervention is shown for the Letter Sound Test (LEST in column 6)—a test that assesses a person's ability to sound out single letters and letter combinations. This is directly related to one of the four modules of ABRA—the letters and sounds module teaches a subset of the letter combination sounds tested in LEST. Both the ICT and non-ICT interventions have a positive and sizable impact on the LEST outcome (almost 22% and 17%

respectively, although the two coefficients do not significantly differ from each other). The ICT and non-ICT treatments are also found to have a positive effect on the DTWRP test (although only the ICT treatment has a significant effect at conventional significance levels). Pupils in the non-ICT condition also seem to benefit over controls in terms of their ability to perform phoneme segmentation as measured by the subtest of PIPA (column 4). Finally, the interventions did not seem to affect attainment in the Letter sound subtest of PIPA (column 5), and the results suggest that the ICT intervention had a negative effect on a child's receptive vocabulary as measured by the BPVS test (column 2). This will be explored further in future research.

Table 12: All primary and secondary outcomes

	(1) PIRA	(2) DTWRP	(3) BPVS	(4) PSS	(5) LK	(6) LEST
ICT	0.138**	0.107*	-0.132**	0.024	0.090	0.219***
	(0.044)	(0.062)	(0.015)	(0.678)	(0.172)	(0.001)
Non-ICT	0.231***	0.096	-0.058	0.096*	0.068	0.168**
	(0.001)	(0.140)	(0.200)	(0.094)	(0.345)	(0.035)
Standardized values of each pre-test*	0.470***	0.690***	0.674***	0.004	0.423***	0.537***
	(0.000)	(0.000)	(0.000)	(0.883)	(0.000)	(0.000)
Female	-0.014	-0.067**	-0.063**	-0.020	0.026	-0.010
	(0.704)	(0.046)	(0.022)	(0.670)	(0.633)	(0.819)
FSM eligible (SPR14)	-0.175***	-0.099**	-0.028	-0.107**	-0.163**	-0.109*
	(0.001)	(0.016)	(0.481)	(0.040)	(0.024)	(0.066)
FSP GLD	0.521***	0.284***	0.219***	0.323***	0.081*	0.351***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.087)	(0.000)
Constant	-0.403***	-0.312***	-0.300***	-0.276**	0.069	-0.563***
	(0.000)	(0.000)	(0.000)	(0.036)	(0.620)	(0.000)
Pupils	1,884	1,881	1,857	1,912	1,912	1,899
Schools	48	48	48	48	48	48
Ho: ICT-NONICT=0 (p-value shown)	0.103	0.879	0.224	0.167	0.739	0.508

Note: Standard errors clustered at the school level. P-values in parentheses (* p < 0.10; ** p < 0.5; *** p < 0.01). ICT and non-ICT are the treatment dummies. All regressions include pairing dummies (schools were paired for randomisation). All the outcome variables and the outcome variables at baseline are standardised to have mean 0 and SD 1. In each column, the variable standardized values of each pre-test correspond to the standardised value of the pre-test for that particular outcome variable. For instance, in column 1 this is the standardised value of PIRA (DTWRP) at baseline.

Cost

The interventions cost the schools nothing in this trial. Here we report the costs associated with implementing each intervention had the EEF had not funded it.

The software for the ICT intervention is available free on the internet. The activities for the non-ICT intervention mirrored those in the ICT intervention. The 20-week lesson plan for both the ICT and non-ICT intervention were supplied free of charge to the schools.

We used professorial level academics to run the training for the TAs who participated. This was funded by the EEF at a rate of £750 per day for the non-ICT intervention and £500 per day for the ICT

intervention. However, the training would be delivered in future by academics on a lower pay grade (£300 per day).[20]

Although the interventions themselves were delivered without charge, there were some costs associated with printing the 20-week programme for each intervention: £3.50 per manual for the ICT intervention and £22 per manual for the non-ICT version.

The implementation team trained each TA for 1.5 days prior to involvement in groups of 12 to 13. On average, each TA received approximately 0.6 days of further support from the project implementation team (a mix of in-person, phone, and email support) at a cost of £300 per day. Forty-nine TAs were trained to deliver the intervention to 710 children.

In addition, to use the ICT intervention, schools need to have access to a computer and the internet. This was already in place in all the schools that participated.

Total cost of ICT intervention in first year (following year's cost = £0):

Cost of training 25 TAs = 1 group of 12 TAs at £500 for 1.5 days = £750; 1 group of 13 TAs at £500 for 1.5 days = £750;
Total training cost = £1500
Cost of support for 25 TAs = £300 x 0.6 x 25 = £4500
Cost of 25 manuals = £3.50 x 25 = £87.50
Total cost = £1500 + £4500 + £87.50 = £6087.50 for 360 children = £16.91 per child

Total cost of non-ICT intervention in first year (following year's cost = £0):

Cost of training 24 TAs = 1 group of 12 TAs at £750 for 1.5 days = £1125; 1 group of 12 TAs at £750 for 1.5 days = £1125;
Total training cost = £2250
Cost of support for 24 TAs = £300 x 0.6 (of a day) x 24 = £4320
Cost of 24 manuals = £22 x 24 = £528
Total cost = £2250 + £4320 + £528 = £7098 for 350 children = £20.28 per child

The cost of delivery based on a post-doctoral academic delivering the training (a research fellow) would reduce the cost of the ICT intervention to £15.24 per child, and £16.42 per child for the non-ICT condition.

Total cost of ICT intervention in first year (following year's cost = £0) based on a post-doc delivery:

Cost of training 25 TAs = 1 group of 12 TAs at £300 for 1.5 days = £450; 1 group of 13 TAs at £300 for 1.5 days = £450;
Total training cost = £900
Cost of support for 25 TAs = £300 x 0.6 x 25 = £4500
Cost of 25 manuals = £3.50 x 25 = £87.50
Total cost = £900 + £4500 + £87.50 = £5487.50 for 360 children = £15.24 per child

Total cost of non-ICT intervention in first year (following year's cost = £0) based on a post-doc delivery:

[20] These price differences are due to the delivery organisation and not reflective of different levels of qualification

Cost of training 24 TAs = 1 group of 12 TAs at £300 for 1.5 days = £450; 1 group of 14 TAs at £300 for 1.5 days = £450;
Total training cost = £900
Cost of support for 24 TAs = £300 x 0.6 (of a day) x 24 = £4320
Cost of 24 manuals = £22 x 24 = £528
Total cost = £900 + £4320 + £528 = £5748 for 350 children = £16.42 per child

Costs could be reduced further by supplying the 20-week programme to schools in electronic format. This would bring it down to £15 per child for the ICT and £14.91 for the non-ICT intervention.

Addendum 1:

In this trial, schools were paid for TAs to be able to attend training (they were paid £128 each). A total of £128 x 49 TAs = £6272. For 710 children, adds £8.83 per child. Their travel costs were also paid, at an estimate to be £25 per TA: £25 x 49 TAs = £1225. This adds £1.73 per child. Thus, the total cost for expenses to cover training are equal to £8.83 + £1.73 = £10.56 per child.
If we consider the previous estimates with the 20-week programmes in electronic format, the cost per child would be £25.56 (ICT) and £25.47 (non-ICT).
NOTE: Following the EEF guidance, the impact on staff time should be reported separately in units of time. Given that most schools did not appoint new TAs or extended hours of existing TAs, we have not included a salary cost for the TAs in the above estimate. In terms of time, TAs are calculated to have employed 4160 hours for 710 children. This amounts to 5.85 TA hours per child per year. If we further consider that TAs needed about 5 minutes per day to prepare the sessions, then the estimate of TA hours per child per year increases to 6.32 hours.

4. Process evaluation

The purpose of the process evaluation was to assess the fidelity of the interventions and to help explain the impact of the ABRA programme as measured by the RCT. We aimed to achieve this largely through assessing the practices and experiences of teaching assistants in participating schools, and through observation of the two interventions—ICT and non-ICT—focusing on teaching and learning processes, particularly pupil response and engagement. We interviewed staff and observed sessions in six selected intervention schools, following up these visits close to the end of the intervention. We also observed introductory and training sessions. In addition we used quantitative data collected by the project implementation team, including their assessments of fidelity and quality of delivery.

We have structured the findings from the process evaluation according to the main criteria used by the EEF: implementation, outcomes, and fidelity. We have also assessed control group activity based on data collected by ourselves and by the project implementation team.

Implementation

Background information on TAs in the active schools was collected through a survey carried out by the project implementation team at Coventry University. This found that TAs had a range of qualifications for the job, commonly NVQ levels 2 or 3, Higher Level Teaching Assistant (HLTA) qualifications and, in two cases, an early years' degree. Most had experience of working across year groups with years of service ranging from a few years to more than 30. TAs main duties varied widely. They included standard TA duties such as supporting class teachers, playground duties, and first aid. Teaching phonics, guided reading, and maths interventions were also part of the role for many.

The TA survey found that almost all had been involved in delivering literacy programmes as part of standard literacy provision. Specific programmes included 'Letters and Sounds', 'Jolly Phonics', 'Phonics Bug' and 'Read, Write Inc'. All of the TAs interviewed had assisted children with literacy before being involved in the ABRA programme. Many, but not all, had done group work and in some cases this was for teaching phonics. For TAs, the main difference between their previous experience and ABRA was its status as an 'intervention' which required delivery in specified ways. But this in itself was not new to most TAs: many had previous experience of interventions and felt this equipped them well for the project. Despite their previous experience of literacy work, the programme itself in either the ICT or non-ICT versions was different to anything that most had used before. The main difference was in the complete and packaged nature of the intervention and the requirement to follow it closely, including through time allocation of components within delivery. As one TA explained:

> *'I've never done it quite so, two minute, three minute, five minute. I've never done it quite so rigid as that before' (non-ICT TA).*

The pace of the sessions was also new to many TAs who had been accustomed to more leisurely activities with individuals or groups. The group work aspect (where support also distinguished it from programmes such as Better Reading Partnership, used in one of the schools) constituted another novel feature for many of them: TAs were more accustomed to delivering support on a one-to-one or one-to-two basis than to groups of pupils.

The ICT intervention was seen as more novel than the non-ICT package. The TA survey found that most said they had used ICT for literacy and numeracy. However, those who we interviewed said that while they had used programmes with pupils before it had been in quite a limited way and rarely for phonics teaching.

Training and preparation

TAs had first known about the intervention shortly before the training session. Most said they had been interested in ABRA when told they would be taking part. However, some had been concerned about how they were going to fit an additional activity into their working day.

Following an introductory conference in May 2014 attended largely by senior leaders, the project team organised 1½ days of training for teaching assistants before the start of the intervention. These were delivered during autumn 2014 (for cohort 1) and autumn 2015 (for cohort 2). During the full day of training TAs were introduced to the interventions and trained how to use them. The half-day event took place two weeks later after they had had time to try out the activities and feedback on their own experience of delivering the intervention activities.

At the full day of training in October 2014, TAs were introduced to teaching reading as well as trained to use the 20-week intervention, with hands-on practice and feedback from instructors. The training was interactive, and TAs were encouraged to ask questions throughout the day. The ICT and non-ICT interventions were designed to be as similar as possible, focusing on effective small group literacy strategies and the specific activities used throughout the programme. Training in the ICT intervention was delivered by Robert Savage, Associate Professor at Canada's McGill University, part of the team responsible for the development of ABRA. Training in the non-ICT intervention was delivered by Morag Stuart, Emeritus Professor in the Psychology of Reading at UCL's Institute of Education. Topics covered by the training are shown in Figure 2.

Figure 2: ABRA training

Introduction to teaching reading:
• how to use the interventions as a tool to teach children skills to maximise their reading outcomes in the broadest sense
• basic reading skills—decoding, fluency, and comprehension
• why the basic reading skills are important to reading outcomes
• teaching multi-ability groups
• managing behaviour in groups and setting group rules

The training on the 20-week intervention:
• the length and number of sessions to deliver
• the aims of each of the activities and how to deliver them
• how to keep records of pupils' progress and attendance
• how to set (and track) the level of each activity to match that of the pupils
• how to access help on each of the activities (in print for non-ICT, on the laptop for ICT)
• how to access (just in time) support during delivery of the intervention

Hands-on practice:
• free time to explore the activities and resources
• group time to deliver/role play individual activities
• group time to deliver/role play a whole session (i.e. 3 or 4 activities)
• structured sessions to feed back experience of delivering sessions and activities
• structured sessions to troubleshoot and share good practice

At the end of the full training day, TAs were asked to go away and practice some of the sessions with children not in the study, and be ready to feed back on their experiences at the half day training two weeks later. They were given contact details to access help with the intervention delivery (email and phone contacts). At the follow-up half day of training, TAs fed back on their experiences of delivering sessions, shared good practice, and discussed any problems they had encountered—and solved these with each other and the instructor. Any questions they had about delivering the interventions were answered during the session. Any extra time was used

to practice activities they had not yet practiced. Once delivery of the interventions had started, Coventry University provided support to the TAs, through a combination of visits, email, or phone depending on the kind of support needed.

As we explain later, the initial training emphasised the importance of fidelity and in particular the need to follow lesson plans and keep to timings. The training also included guidance on managing group work, for example by establishing ground rules.

We asked teaching assistants about their experiences of the training during our visits to schools in January 2015 and also analysed training evaluation data collected by the project team. We also attended training days 1, 2 and 4 where TAs were trained in separate groups in the ICT and non-ICT interventions.

Data collected through the interviews and evaluation forms shows TAs consistently rated the training highly with almost all in both the ICT and non-ICT intervention training strongly agreeing that the goals of the training had been met and that they were satisfied with their increased understanding of the topic. Those in the ICT group were more likely to express this than those in the non-ICT training. The evaluation forms asked TAs their level of agreement with the following statements:

- The goals of the training were clearly defined.
- The topics covered were relevant.
- The introduction to each session stated the objectives clearly.
- There was sufficient opportunity for interactive participation.
- The training was too technical and difficult to understand.
- The training experience will be useful in my work.
- Most of my questions were answered during the training.
- The materials were pitched at the right level.
- The materials for the training were helpful.
- The schedule for the training provided sufficient time to cover all of the proposed activities.
- The handouts provided were helpful.

Almost all TAs in both ICT and non-ICT groups agreed or strongly agreed with these statements, with the exception of ‘the training was too technical and difficult to understand’ where the TAs felt this was not the case. TAs in the ICT group were more likely than those in the non-ICT group to agree strongly with the statements, with around half in the ICT group agreeing strongly with most of them. These high levels of agreement do not indicate that the training was easy. TAs interviewed delivering the ICT based intervention felt there was a lot to learn, particularly those who saw themselves as not technically-inclined. Some said that, on reflection, they would have liked a little more time to explore the different parts of the programme, but that they had practised enough to be able to implement it in school. TAs interviewed during the school visits felt the training had prepared them well for the intervention. They were also reassured that they could contact the Coventry University team when they needed to, and some said they had done this. Examples of where this had been necessary included whether to remove a child because his behaviour was preventing other group members from learning. The Coventry team had also provided additional equipment such as laptop speakers and replacement timers.

Organisational issues

For schools, the main organisational issue involved allocating TAs and their time to the intervention. TAs on the project were largely drawn from those working with Year 1 pupils, although some from other years were included where additional numbers were required. All

equipment was provided by the project. Schools then had to allocate time in the school day to deliver the programme. Schools were asked to deliver it during literacy-based lessons but not core literacy instruction, including phonics work. This is because the intervention was designed as an additional programme, not a substitute for normal classroom delivery of literacy. Visits by the project team enforced this feature of the programme, correcting schools which were timing sessions to coincide with core literacy work. The TA survey found that no school was taking pupils from core literacy classes and most delivered the intervention during general literacy-based sessions, such as guided reading or topic work. This was confirmed by school visits by the external evaluator: sessions were reported to take place largely during non-core literacy time, although was sometimes also during break time or PE (physical education). In some schools the intervention was timed flexibly so that pupils did not regularly miss the same lesson, particularly where this was literacy.

Support from schools and the project team

While some schools faced some challenges, particularly with space to deliver the intervention (see below), TAs said that they had been fully supported by senior leaders within their schools, including covering their duties, as well as by the project team at Coventry University and had experienced few problems. The TA survey found that most felt very well supported by their school, with 31 of 37 saying they felt very or somewhat supported; 25 of 37 TAs said they felt very supported by the project team at Coventry University and the rest said they felt somewhat supported. Despite this support, a small number of TAs responding to the survey indicated that the project impacted on their workload and they had to spend additional time on preparation and writing up. At the follow-up stage there were reports that TAs felt that the demands of the project impacted on the time they could spend supporting other children and class teachers.

Technological and skills issues

The survey carried out by the project implementation team asked TAs whether their use of the intervention was ever affected negatively by technological problems and limitations. Most TAs delivering the ICT intervention did report problems but these appeared to be minor and occasional, for example a screen freezing. No TA reported frequent technological problems. Very few TAs said their use of the intervention was negatively affected by their skills. Of these few, their reasons were varied, including lack of preparation time and confidence that they were delivering it correctly.

Practicalities

There were a number of practical issues for TAs: taking pupils to and from sessions, space within the school, and the short length of the sessions. The survey found that 24 out of 37 TAs said the intervention was sometimes negatively affected by school based problems or limitations, such as time and location. Six TAs said there were frequent such limitations and problems. Only six said these never occurred. Problems included space and clashes with other school activities. TAs interviewed confirmed that space was one of the main practical issues encountered by both the ICT and non-ICT interventions. Sessions were delivered in a range of locations within the schools visited, including spare rooms, classrooms alongside lessons, and in corridors. As one TA explained, this was sometimes time-consuming:

> *'You walk through the school. Everywhere is occupied. It all takes time, especially if you have to do three groups a day.'*

Some of these settings were distracting for pupils and made it difficult for them to hear the ICT programme voiceover.

TAs interviewed also reported practical issues with the need to take pupils from class to the activity. This was reported to be sometimes time-consuming, for example where pupils were

collected from PE and needed to change their trainers for shoes (not a fast process for a five year-old). Teachers were reported to be sometimes concerned at the disruption to their lessons.

Length of sessions and group size

The length of sessions, at 15 minutes, was seen by some TAs as too short. For the ICT programme in particular, the 15 minutes was eroded by time required to settle pupils down, log on, and refresh pupils about the previous session (see later). At the same time, TAs delivering both interventions felt that much longer would challenge pupils' attention. One view was that time constraints made the ICT programme less beneficial to less able children who need more time to work through the activities before moving onwards. Some TAs delivering the non-ICT intervention found it challenging to keep to the timings within sessions, but again became accustomed to their structure. TAs in both the ICT and non-ICT interventions were instructed to move between levels only when the group had reached a correct response rate of 80%. Some TAs had been a little concerned about this requirement, feeling it might be difficult to decide. In practice, they found this easy to follow.

The size of pupil groups, at between three and five, was seen as about right, allowing each child to have an input but to also learn from others in the group. For the ICT intervention a smaller group allowed pupils to interact more directly by using the mouse themselves. In larger sessions the TA tended to control the mouse for most of the sessions observed. More than five was seen as too many for the ICT intervention but less so for the non-ICT sessions. While the project allocated pupils to the two interventions or the control, TAs were able to allocate pupils into groups within the intervention, and did this largely according to ability level. This was seen to work better than mixed ability groups, particularly for the ICT intervention.

Attractiveness of the intervention to stakeholders

TAs found the ABRA programme attractive in both its ICT and non-ICT forms. They welcomed the structured nature of the programme and the full set of resources provided for both interventions. TAs delivering the non-ICT intervention liked the paper resources, which include a series of storybooks. TAs delivering the ICT version liked the programme, its graphics and activities. A number commented that, before using the programme, they usually had to source at least some resources for themselves and liked these being brought together in a complete package. One advantage of the non-ICT intervention was its perceived greater adaptability for different ability levels.

TAs on both interventions enjoyed delivering small group teaching. They liked improving their literacy teaching skills and those in the ICT group felt they had improved their navigation and mouse skills. They also liked the programme itself although, in time, some became somewhat bored with the repetition it inevitably involved. TAs delivering the non-ICT intervention also liked the resources, in particular their completeness as a package and the absence of photocopied worksheets.

As a TA-led intervention, we were interested to know participants' views on this aspect of the design. Mixed views were expressed on whether pupils responded differently to a TA-led intervention than if it had been delivered by a teacher. Some TAs felt that pupils were more likely to take risks in their responses with a TA, particularly within a small group. One described the atmosphere of a TA-led group as more 'homely' or nurturing because of TAs' pastoral role. However, others felt that it was small group teaching which changed pupil behaviour, that pupils were not able to draw a distinction between TAs and teachers and that either could form a relationship of trust. As one TA delivering the non-ICT intervention explained:

> *'If someone else is answering the question, they've not got to answer it so will just sit there and zone out sort of thing, but when you're*

> *asking them it's more one-to-one, they're actually having to get involved. They're concentrating more because they have to be able to answer the question.'*

The positive response of pupils was an important motivating factor for TAs. This was particularly strong for the ICT intervention which some TAs said appealed strongly to pupils. Their belief that the intervention was making a difference to pupils' literacy skills was also an important factor in their commitment to, and interest in, the project.

Preparation time was an issue for some TAs interviewed. During the training, TAs were told that little planning time was required for each session. However, TAs found that they did in fact need some time to prepare, but usually no more than five or ten minutes. Both interventions required some set up time and the ICT version required additional log-in time. Some schools had allocated such time, for example a TA with three non-ICT groups had 15 minutes a day for preparation. In another school, TAs were simply re-allocated time assigned for guided reading preparation. Despite these allowances, the intervention presented some challenges to the time management skills of some TAs.

A potentially less attractive side of the intervention was the requirement on TAs to keep records of each session, indicating the level that pupils had reached. They had been concerned that record-keeping would be onerous but generally found it straightforward and even useful for subsequent sessions, reducing planning time. Some TAs took only rough notes at the end of sessions and wrote them up more fully later and did not see this as problematic or excessively time consuming. Some also made their own additional notes, for example recording high frequency words identified by pupils.

Pupil response

As stated above, TAs were motivated by the positive response of pupils. They reported that pupils enjoyed the sessions, both the ICT and non-ICT versions. Pupils were reported to look forward to going and to enjoying being taught in a small group. As one TA delivering the non-ICT version explained:

> *'They love it, not because of this particular session but because they get the group and one-to-one attention.'*

One TA delivering the non-ICT intervention described how children skipped across the playground in haste to get to the session. Pupils' enjoyment of the intervention was explained with reference to a number of factors: the small group format, the interactive nature of the sessions, the humour and the 'game' aspect, and more generally the appeal of the resources.

Some TAs felt that pupils enjoyed sessions more when groups were small, for example only three pupils, and where each pupil could participate. Higher-ability pupils were seen to particularly welcome being involved, since interventions are usually targeted at less able pupils. A TA explained:

> *'Lots of the children who are getting to do it would never be in an intervention, it's a novelty. They say that splitting the children up in abilities of their peers makes for more effective delivery.'*

These pupils were found to enjoy the interactive nature of both the ICT and non-ICT sessions and the opportunity to demonstrate their knowledge and skills. As another TA explained:

> *'One of the questions in the Frogs and the Well is "have you seen a pond like this before?". The lowers say "Yes, I've seen a pond in Cannon Hill Park". The higher-ability kids say "Yes on TV I saw this", and they have all this knowledge that they are desperate to share. I*

> *think within the class they don't get to use these voices and to talk about what they know. So this I think is really important, for the lowers and the higher. I think it's got something for both, and the middles just the same.'*

Pupils clearly enjoyed the ICT based sessions. They engaged in the programme, characters, language and stories. They enjoyed the humour of the programme, for example recognising incongruous words and phrases, such as 'went down to the river to get a hamburger'. Pupils also enjoyed shouting out words as they appeared on the screen. In a number of the groups pupils expressed an enthusiastic preference for particular activities, for example changing a letter to change the word. The pace of the ICT lessons was fast, led by the programme. This kept most pupils engaged and could give an advantage to higher achievers and faster learners. However, it was common for pupils to work as a group rather than for individual pupils to dominate response activities. TAs commented on the game format of the ICT programme which they felt helped pupils to learn while playing. As TAs from two different schools explained:

> *'It's more of a game situation so they don't have to think as much… They find it easier to do something if it's a bit more fun.'*

> *'When they're coming out of the class it feels like this is going to be a bit of fun, let's crowd around the computer, let's just have a bit of a game. Sometimes they'll say, "If we go back to the class we're learning in there. This isn't learning, this is games" and they don't even realise that what they're doing is integral learning for them.'*

This also had its downside in pupils sometimes not understanding why they couldn't choose which 'games' to play.

Pupils were also reported to find the non-ICT resources attractive. They liked the books and stories and receiving stickers at the end of the session. As one TA explained, in contrast to the usual photocopied sheets that pupils work with:

> *'The books are coloured, so for them to open coloured books is different. You don't get that in class. Normally for phonics you have paper worksheets.'*

However, the pace of the non-ICT sessions was reported to be slower with pupils expressing boredom with reading the same book for a number of weeks.

A few drawbacks to the ICT intervention were identified. First, the Canadian accent and, to a lesser extent, vocabulary, was a minor distraction for both TAs and pupils. One example of where this presented difficulty was in changing word meanings through letter change: 'ten' to 'tan' is less apparent to an English ear when spoken in a Canadian accent. Some TAs highlighted any differences to pupils and therefore made this into a learning point. For example, one TA told pupils a few times 'I would say that differently'. Some words familiar in Canada were not understood, for example 'recess', 'dike' and 'blimp'. Children in one group mistook 'Caribou' for 'car boot'. This led to a short discussion on whether a caribou is a reindeer or a moose. Some commonly used English words were also not always understood. For example, one group of pupils did not understand some common terms such as 'current' in the context of a river, or knew what a wig was. Other words or phrases were just seen as too advanced, for example 'indestructible igloo' and 'Ursula's upturned umbrella'. Some TAs commented that as well as including Canadian vocabulary, the programme had a middle class bias, again apparent in the language (although examples were not given).

A second drawback to the computer-based intervention concerned the ICT skills of some pupils. It was apparent that the mouse skills of some pupils were not well developed. In some cases

interaction with the programme was largely TA-led. This appeared to be done in some cases to ensure that the pace of the lesson was kept up so that activities were completed on schedule.

Pupil engagement was good during the sessions observed for the external evaluation. However, TAs had to work harder to engage pupils, to keep their attention, and to ensure a steady pace. One TA used a teddy bear to add interest to the session (teddy was told a word and pupils were asked what letter it begins with). Some pupils and groups appeared to find the non-ICT intervention somewhat repetitive. Pupils of higher ability in one group complained they found the decoding activities 'too easy'. For this reason, TAs in another school had changed the balance of activities with higher-ability pupils spending more time on comprehension than on decoding and blending.

Our observations found pupils engaging more fully with the ICT programme than the non-ICT intervention. While pupils liked the books and activities, these had less appeal than the ICT programme. At the same time, some pupils were very positive about the non-ICT sessions, with one group going as far as describing the sessions as 'fantastic' to the NIESR evaluator. In that particular group, of higher-ability pupils, the TA felt the activities enabled pupils to learn from each other's responses.

We referred earlier to the space problem experienced by some schools. Pupil response to both interventions depended in part on location of the sessions. Some sessions were held in classrooms where other activity was in common areas or corridors. TAs had to work harder to keep pupils' attention in such circumstances, and faced more challenges with the non-ICT sessions. TAs managed these challenges very well.

Fidelity

The importance of implementing the project faithfully was impressed on TAs during the training. Particular emphasis was placed on not sharing information about the training or project materials with TA colleagues delivering the different intervention. TAs also understood from the training that the intervention had to be delivered as specified. This meant:

- keeping to timings;
- moving on to higher levels only when pupils were ready, measured at having reached 80% accuracy in responses; and
- keeping records.

The Coventry University project team put a range of fidelity control measures in place. These included requiring TAs to record progress made in each session through completing tracking sheets. The project team also kept in regular contact with schools and TAs and visited each school to observe all TAs involved in delivering the project. The purpose of this was to check that both the ICT and non-ICT interventions were being delivered at the level of expected quality.

Our visits found that some TAs had been concerned about these requirements but in practice found few difficulties. As one explained:

> *'When I first saw [the record sheets] on the training, I was like "Oh God, paper work, how am I going to deal with it?" But they're quite easy' (non-ICT).*

TAs took the requirements of the project very seriously, with very few saying that they deviated from the guidance in any way. Those who did made very small adaptations, for example writing notes during the session and writing them up later. Responses to the survey carried out by the project implementation team indicate that none delivered it in any setting other than a small group.

TAs were also asked to keep the ICT and non-ICT interventions distinct by not sharing information about the content and delivery of the two programmes. The project team survey found that most TAs did not see the other intervention being delivered, though six did. However, the level of awareness of the actual content of the alternative intervention was low: most simply said they knew that one involved ICT and the other paper-based resources. Those who appeared to have more of an idea of differences, for example that the ICT intervention was 'games focused' were not confident that they really knew. It is therefore very unlikely that either approach was contaminated by the other in terms of content, delivery, or anything else.

Scheduling and delivery

TAs were advised at the training stage to follow the programmes to the letter and to tell the project team if they were not able to keep to the lesson plans. The project implementation team did not wait for TAs to approach them with any concerns or questions, although some TAs did make such contact and in some cases quite regularly. The Coventry team very actively engaged with TAs by making regular contact. This would seem to be an effective way of ensuring that an intervention is delivered with fidelity. It may be particularly important for project teams to take this proactive approach in TA-led interventions given that TAs are likely to have less time than teachers to prepare or follow-up on project-based activities (Blatchford *et al.*, 2012).

Visits by both the project implementation team and by external evaluators found schools implementing ABRA according to the project design. This included keeping attendance records and keeping to the lesson plans and activity timings. TAs were advised not to advance to the next level of activity until the group had scored 80%. Those interviewed during the evaluation visits said they had followed this guidance. TAs also said they recorded progress made in each session, despite initially expressing concerns about this.

TAs responding to the project implementation team survey reported spending varying periods of time delivering the sessions. Fourteen said they spent 15 minutes, as specified in the training and guidance; eighteen said they spent between 15 and 20 minutes. Only three said they spent more than 20 minutes but that this included getting children ready, which others reported was time-consuming.

The tracking sheets required TAs to show that a session had taken place and whether any pupils in the group were absent. Two members of the project implementation team scored these sheets independently for quality, compared results, and arrived at a final score. Records were collected from 49 TAs. The exercise found that 34 out of 49 TAs provided complete records of their sessions. Of those who did not, only four TAs were scored at five (out of ten) or below, indicating consistently poor record-keeping rather than some missing records. The tracking sheets also asked TAs to record evidence that they were moving pupils appropriately through the levels, moving on once they achieved 80% accuracy in response. The project team found this somewhat difficult to assess from the tracking sheets. TAs were asked to record this for the activities of blending, decoding, and word changing, but not for comprehension. TA records were found to often give an indication of appropriate movement of pupils in one of two of these areas, but not to give clear evidence of all three. The project team therefore had to make a qualitative judgement that, on balance, TAs showed evidence of appropriate movement. Their scoring found that 19 out of 49 TAs showed very strong evidence, scoring eight or nine out of ten; 21 TAs showed fairly good evidence, scoring between six and seven. Only three returned no records at all and two of these were from a school that withdrew from the project on the appointment of a new head teacher.

The tracking sheets also show whether the TAs completed the full 20 weeks of the intervention. Records of ten TAs indicate some shortfall, however two of these are missing only one day of delivery. Two of the others were from the school that withdrew from the project. In the other

cases, TAs either job-shared or passed the project over to another TA in the school who delivered it to the end point. The indications are, therefore, that the TAs delivered the full programme as intended.

The project team also carried out a full programme of school visits, observing all TAs delivering the ICT and non-ICT interventions. These observations used eight criteria to assess the sessions: exposure, planning, instructional guidance, opportunities to succeed (levels and differentiation), group cohesion, pacing, and efficient use of time and behaviour. An overall rating was made of the session. Methods to carry out the assessments are described in the footnote.[21] The project team found the quality of delivery generally high with only 10–20% falling below expected standards. These TAs were then given advice and guidance on how to improve their delivery of the intervention. This indicates that the project was very largely delivered at the expected standard. Differences were found between individual teaching assistants rather than between participating schools.

Adaptations to the programme

Few variations seem to have been made, partly because the ICT programme was not seen to be flexible or allow departures from the schedule. The non-ICT intervention was seen as more adaptable, and some TAs altered the balance of activities either towards more blending and decoding for lower-ability children, or more comprehension for pupils who were more fluent readers. These changes would not appear to affect the fidelity of the programme. They were also important in schools' considerations about whether to continue with the programme after the end of the intervention.

TAs were asked not to share information about their respective intervention with colleagues delivering the alternative design. This was difficult for some TAs to follow because the activities often took place within schools in full view of staff and pupils. As noted, TAs were not aware of the detail of the alternative intervention to their own. It was therefore equally difficult to keep other children unaware that there was an intervention which did not involve them. However, this was not seen as problematic by TAs because of the young age of the children involved. As one TA delivering the ICT intervention explained:

> *'Our children go round and say "we did Abracadabra today" and the others are like "what's that?" and I am like "shush" but it's been fine, they are children and they are going to talk about something they enjoy. The word Abracadabra is quite interesting, but I don't think they know how to explain it. When someone says "what is it?" they just say "we go on a computer". So in all fairness they don't say much.'*

Little flexibility was offered in delivery and the need for this was understood by TAs. One area of flexibility which TAs were concerned to have was in the allocation of pupils to groups (within a treatment condition). It was suggested at the training that they might wish to group pupils by ability. They were also advised to take account of gender and personalities. At the training stage, TAs' main concern was to ensure that pupils with behavioural problems were not grouped together and that groups would function well. Trainers told TAs that they could allocate

[21] The project implementation team shot four videos of mock sessions (a different 'teacher' each time) in order to calculate Inter Rater Reliability (IRR). Project implementation team members then rated each of the four videos and calculated the IRR to be 0.71. They identified two sub-scales that had been interpreted differently. They consequently changed the sub-scales, re-scored the video sessions, and re-calculated the IRR as 0.85 (95% CI: 0.508–0.988). Most of the sessions that they evaluated were observed by two members of the team and an average score was used for the final rating.

pupils within the two intervention conditions but not move pupils between the two interventions. This guidance seems to have been followed. Only six TAs responding to the survey said that they had done any rearranging of groups at all. Where they had it was because children had left the school or because of the behaviour of individual pupils.

Outcomes

TAs expressed different views on which ability groups might benefit most from ABRA. The non-ICT intervention was seen to meet the needs of pupils of all ability levels. Pupils of lower ability were seen to benefit from its emphasis on blending and coding. Children with very poor literacy skills were found to struggle with the level of the text but still to benefit from the sessions. Higher-ability children were seen to benefit less from the basic decoding and blending activities which were seen as too basic for them, however some TAs felt the programme benefited higher-ability children whose technical reading ability is good but who have relatively poor comprehension. This was explained with reference to the opportunity for discussion in a group setting. That said, a small number of TAs still felt that higher-ability pupils required more structured activities within the non-ICT intervention. Some suggested that there should be more differentiation for lowest and highest ability pupils.

The ICT intervention was seen to have something in it for all abilities and to have an advantage in this respect over the non-ICT version. Lower-ability pupils were seen to enjoy the interactive element of the programme. As one TA explained:

> *'It's fun. It's on a laptop so for them it's exciting. They get to come out and do something. They get to click buttons and all sorts and a lot of it is actual phonic sounding out which is so brilliant.'*

However, some TAs thought the pace was too fast for some lower-ability children and too slow or repetitive for higher-ability groups. The programme itself was seen as adaptable, but TAs wished to follow the guidance and keep to the timings. They also faced the problem of making sessions less suitable for one ability group by making it better for another. This was less of a problem for the non-ICT intervention since pupils were seen to gain differently from the activities and stories.

The ICT intervention had the advantage of allowing pupils with good ICT skills but weaker literacy skills to have a role. This was seen to have the effect of improving the confidence of some lower performing pupils. Another perceived advantage of the ICT programme was that some pupils viewed it as play, and therefore learned in a more natural, non-resistant way. Overall, there was no clear message from the process evaluation about which pupils benefit most from either intervention.

TA views on the impact of ABRA

TAs delivering both programmes believed they were having an impact on pupils with most (32 out of 37) surveyed saying they thought it had a positive or strong positive effect in terms of its benefit to pupils. The external evaluation visits shed some light on this finding. Some TAs explained this with reference to the frequency of the intervention, at four times a week. TAs delivering both the ICT and non-ICT interventions had noticed that pupils were using decoding and blending skills in class. However, some TAs felt that, while intense, the intervention was too short. As one TA delivering the ICT intervention stated:

> *'If it's going to make a difference I think it's going to have to be extended a bit.'*

A number of features of the intervention were seen as promoting learning, in particular the group format: TAs felt that this encouraged participation and enabled children to learn from

each other's responses. This was reported by TAs in both interventions. Individual pupils were seen to have grown in confidence and to be contributing in ways not in evidence in their class situation. This was seen to have the potential to impact on their learning. TAs also felt they had themselves benefited from being involved in the intervention. In the survey most TAs said they felt more confident in delivering small-group work after completing the intervention and that they had a better understanding of phonics and children's language development.

TAs were asked in both the survey and the visits whether they would continue with the programme in the longer term. In the survey 13 out of 32 who responded said they would, 8 said they would not, and 11 were not sure. TAs interviewed were generally inclined to continue but would prefer to do so in a less structured way, for example by using the programme less frequently or more flexibly. TAs interviewed during the external evaluation visits said they would like to see some small changes to the programme: some believed, for example, that the vocabulary presented a barrier to progression for some pupils, particularly in the ICT intervention, and felt this would need to be adjusted for pupils to benefit fully. The non-ICT intervention was seen as in less need of adaptation, but some TAs felt that timings should be adjusted, for example by reducing the length of time spent on each book.

Formative findings

The intervention was well implemented and there are no obvious ways in which it could have been improved. TAs were well prepared, they clearly understood the purpose, design, and delivery plan. They then implemented it as intended. This included aspects such as timing, use of materials, and organisation and practical matters. As far as was practically possible, the two interventions were delivered separately without contamination in the active schools. The project team provided proactive support, and did not wait for problems to be reported.

Three changes might usefully be made for future implementation of the programme:

TAs felt that mixed-ability groups worked less well than ability sets in that the latter allowed pupils to move onwards at the same pace and not to become bored. Future implementation should aim to ensure delivery in ability sets where possible.

- The vocabulary could be adjusted to remove Canadian terms and made more culturally relevant to British children.
- The delivery of the non-ICT version could be adapted to allow for faster progression through the books since this was perceived to be slower than in the ICT version.

Other changes which could improve implementation are largely beyond the control of the programme and need to be addressed by schools themselves. These principally concern issues of space—delivery in corridors or in classroom corners is unsatisfactory both for pupils and TAs. Schools also need to factor in time for pupils to be taken out of class to attend sessions, and to return. To deliver the ICT programme, TAs require access to reliable PCs or lap-tops and a small amount of time for preparation and record-keeping.

Control group activity

Data on control group activity was collected both from treatment and control schools through a survey carried out by the project implementation team at Coventry University. The survey asked lead teachers about the ways in which literacy is taught in their schools. It included questions about hours a week spent on literacy instruction, programmes used, and the time allocated to activities such as blending, decoding, reading and comprehension. It also asked about the use of worksheets, exercise books and other formats, and the use of technology to support literacy teaching. Responses to the questionnaire provide an indication of whether active and control schools were in any way different in how they deliver literacy teaching.

Survey responses show few meaningful differences between active and control schools that could have affected the intervention results. Both allocated similar amounts of time to literacy instruction—around 1½ hours a day. Most schools used a combination of whole-class, small-group and individual pupil approaches, although the treatment schools said they used more whole-class teaching. This finding would seem to discount the possibility that small group teaching was more common in treatment schools and that they therefore had a head start in this respect. Both treatment and control group schools placed a moderate or great deal of emphasis on teaching phonics and had used programmes, in particular 'Letters and Sounds', 'Read and Write Inc' and 'Jolly Phonics'. There were no indications that either group had used these to a greater or lesser extent than the other.

Little difference was found between treatment and control schools in the time they allocated to activities such as blending, decoding, reading and comprehension. Control schools reported spending slightly less time on decoding, teaching fluency, and reading comprehension than treatment schools but more time on 'reading activities'. They reported spending more time on sequencing (putting parts of a story in the correct order), on vocabulary development, and on creative literacy activities such as drawing and writing plays. However, the sample sizes are small so caution should be exercised in reaching any conclusions from this data. Finally, treatment schools were less likely than those in the control group to report that events such as the weather and other activities have hindered the teaching of literacy. There were no differences in reported use of technology to support literacy teaching, with smartboards almost universal, computers used by around two-thirds, and tablets and iPads by a slightly smaller number.

Data collected from TAs in the treatment schools on the scheduling of sessions and what lessons were missed by intervention pupils provides additional information. This data suggests that the programme did not substitute for core literacy instruction (including phonics work) and was used, as intended, as an additional programme, not a substitute for normal classroom delivery of literacy. The external process evaluation did not find schools were compensating for the programme by delivering additional help to pupils in the control group.

5. Conclusion

Key conclusions
1. The children who received ABRA, or its offline alternative were found to make two and three months' progress in literacy respectively compared to the children who received standard provision. This positive result would be unlikely to occur by chance.
2. For both ABRA and the offline alternative, the impact for children eligible for free school meals and children with below average pre-test outcomes was larger than for all pupils.
3. Successful implementation contributed to a well-designed and delivered training programme which emphasised fidelity and consistency, enforced by ongoing support from the project team.
4. The process evaluation found that both the ICT and non-ICT interventions may be best delivered in groups of similar rather than mixed ability. The process evaluation also suggested minor changes to the intervention to make it more culturally relevant to British pupils, and to remove some repetition in the non-ICT programme.
5. Future research will examine whether ABRA or the non-ICT intervention can be successfully delivered at scale, and will look at longer-term impacts through assessing Key Stage 1 data from this trial.

Limitations

The power calculations performed during the design phase suggested recruiting a total of 60 schools to be able to detect an effect size of 0.20 standard deviations with 80% of power. Despite the impressive recruitment efforts of the implementation team at Coventry, these numbers could not be met. A total of 48 schools were available for the final analysis. Given the data collected from the trial, we calculate that we can detect effects of at least 0.27 standard deviations with a power of 0.80. However, and as we stated in the protocol, in practice we could possibly detect smaller effect sizes. This is because there was a second level of randomisation at the pupil level, and the calculations are based purely on a paired cluster randomised trial. However, even if we have been able to detect effect sizes smaller than 0.20, we might lack power to detect significant effect sizes in subgroup analyses, for instance.

The number of clusters (schools) available in the final sample should not pose a problem for the calculation of clustered standard errors. Dividing the sample to study subgroups where the number of clusters is reduced considerably, though, becomes rather problematic and results in less than 30 schools and clustered standard errors should be taken with caution. This is why we have performed the subgroup analysis by studying interactions instead.

An additional source of bias in this evaluation results from school attrition. A total of five schools dropped out after randomisation. Fortunately, data for all of these schools (except one) was collect for post-treatment tests. This means that we have been able to perform an intention-to-treat analysis using all, but one, of the initially randomised schools. A further problem might result from some pupils not providing consent or leaving school in the middle of treatment. Nevertheless, the missing data analysis performed (with a total of 84% of the initial sample available for analysis for the main outcome) suggests that both pupil and school attrition are not a major source of bias in this evaluation.

Moreover, for participating schools the process evaluation found that the intervention was delivered as planned. Three changes might usefully be made for future implementation of the programme: (1) pupils could be allocated to same-ability groups, (2) the vocabulary of the programme could be adjusted to remove Canadian terms and made more culturally relevant to

British children, and (3) the non-ICT version could allow for faster progression through the books. Additionally, some changes which could improve implementation are largely beyond the control of the programme and need to be addressed by schools themselves. These principally concern: (1) issues of space, with delivery in corridors or in classroom corners unsatisfactory both for pupils and TAs, (2) schools need to factor in time for pupils to be taken out of class to attend sessions, and to return, and (3) to deliver the ICT programme, TAs require access to reliable PCs or lap-tops.

In terms of the generalizability of the results, we can compare some descriptive statistics in our final sample of schools with those of national State-Funded Primary schools (selected from the School Level Census 2013/2014). The schools in our sample are bigger on average (93 pupils in Key Stage 1 in our sample compared to 74 in State-Funded Primary schools), and they have a bigger fraction of FSM-eligible pupils and pupils whose first language is known or believed to be other than English. In particular, around 20% of the pupils in our sample are FSM-eligible and a similar proportion do not have English as their first language. These statistics are almost 16% (for FSM eligibility) and 14% (English is not the first language) for pupils in English State-Funded Primary schools. This is in line with the recruiting strategy, since an effort was made to encourage schools with disadvantaged intakes to participate during the recruitment stage. Therefore we cannot assume our findings are generalizable to primary schools generally as this is clearly not a representative sample of all schools.

Interpretation

The positive findings for the ICT programme are in line with previous studies of the ABRA programme. In particular, the ABRA programme has had some support from smaller efficacy RCTs (see, for instance, Savage *et al.*, 2009) and a bigger effectiveness trial (Savage *et al.*, 2013). Both studies found that ABRA leads to improvements in literacy for treatment pupils at post-test. This trial, however, differs from the latter in several respects: (1) the size of the trial in terms of pupils is doubled, (2) this intervention compares ICT and non-ICT delivery of a broadly identical programme—this allows the evaluation of different delivery methods with respect to improving results for early literacy, and (3) we have a clean control group—pupils in schools that do not receive and do not know about the existence of the web-based programme while the intervention is in place. If we have reasons to believe that there might be spillover effects from treatment to control individuals within the same school (and we have some preliminary suggestive evidence of that) then having a control group in schools that are not involved in the intervention gives a much cleaner identification of the treatment effect.

In the broader literature, there are few rigorous studies about the use of ICT to teach literacy that show positive effects. A review by Cassen *et al.* (2015) states that although there is evidence of positive effects of ICT more broadly on English learning, the results are very inconsistent and restricted by the rate of ICT use and access in schools. Three relatively large-scale studies with a strong methodological design find no effect of teaching with ICT on pupil learning (Angrist and Lavy, 2002; Rouse *et al.*, 2004; and Berlinski and Busso, 2014). In contrast, the findings in this study will potentially contribute to this broader literature by suggesting that an ICT-based intervention has a positive effect, however at this point we cannot make strong claims about the impact of technology in this trial. Future research will explore alternative data sources to try to understand the difference in magnitude between the ICT and non-ICT estimates.

Even though we find positive effects relating to both interventions, the impact of the non-ICT intervention is significantly larger in magnitude than that of the ICT intervention. Several factors could be driving these results and further analysis will be performed in the future in order to address the potential mechanisms. For instance, the process evaluation highlighted that TAs perceived the non-ICT intervention to be more adaptable to different ability levels. Incorporating

the data from TA logs and surveys into the statistical analysis could potentially help to shed more light on this, and may also help to account for the differential impacts found by subgroup, and the preliminary evidence suggesting the existence of positive externalities (peer effects) for pupils in the control group of intervention schools.

With respect to the process evaluation, it provides a number of explanations for the impact findings. These principally concern the active engagement of both TAs and pupils with the intervention. TAs felt well prepared for the intervention, were supported by the project implementation team, and took the responsibility of delivery seriously. TAs felt they benefited through being involved in the intervention, improving their skills in small-group tuition and, for the ICT programme, their IT skills. They delivered the programmes according to plan and moved pupils appropriately through the levels. Other aspects of the delivery were also found to be of high quality, maximising the possibility of impact for the programme.

TAs liked delivering to small groups, they felt it fitted well with their role and liked the resources. All of these factors meant they delivered the intervention both diligently and with some enthusiasm. This appeared to inspire the pupils who were observed to be actively engaged with both the ICT and non-ICT programmes. Pupils enjoyed the small-group format led by a TA, the interactive nature of the sessions, the humour and the 'game' aspect, and, more generally, the appeal of the resources. They appeared to be particularly engaged by the ICT programme. However, the attractive features of both interventions are likely to have increased their engagement and assisted their learning.

The process evaluation also found the intervention was delivered as planned and that there are no other likely explanations for the higher scores of pupils in the treatment schools.

Future research and publications

Future research will examine Key Stage 1 results for treatment and control pupils with the goal of determining whether the positive impact found for both interventions persists one year after the end of the intervention. Moreover, additional data collected by the implementation team in terms of pupil group composition (number of pupils, average ability of pupils, and so on), TA surveys, TA logs, and other materials will be incorporated in the analysis with the objective of understanding the higher impact of the non-ICT intervention. Moreover, we expect to publish academic paper(s) in collaboration with the team at Coventry University.

Abbreviations:

IDM: Index of Multiple Deprivation

PIRA: Progress in Reading Assessment

DTWRP: Diagnostic Test of Word Reading Processes

BPVS: British Picture Vocabulary Scales 3

PSS: Subtest from the Primary Inventory of Phonological Awareness (PIPA): phoneme segmentation subtest

LK: Subtest from the Primary Inventory of Phonological Awareness (PIPA): letter sound subtest

LEST: Letter Sound Test

References

Abrami, P.C., Savage, R.S., Deleveaux, G., Wade, A., Meyer, E. & Lebel, C. (2010). The Learning Toolkit: The design, development, testing and dissemination of evidence-based educational software. In P. Zemliansky & D.M. Wilcox (Eds.), Design and implementation of educational games: Theoretical and practical perspectives (pp. 168-187). Hershey, PA: IGI Global. **http://dx.doi.org/10.4018/978-1-61520-781-7.ch012**

Angrist, J. and Lavy, V. (2002) 'New Evidence on Classroom Computers and Pupil Learning', *Economic Journal*, 112, 735–765.

Berlinski, S. and Busso, M. (2015) 'Challenges in Educational Reform: An Experiment on Active Learning in Mathematics**',** *IDB Publications (Working Papers)* 88335, Inter-American Development Bank.

Blatchford, P., Webster, R. and Russell, A. (2012) 'Challenging the Role and Deployment of Teaching Assistants in Mainstream Schools: The Impact on Schools', Final Report on the Effective Deployment of Teaching Assistants (EDTA) project

Cassen, R., McNally, S. and Vignoles, A. (2015) *Making a Difference in Education. What the evidence says*, London: Routledge.

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 children', *Journal of Research in Reading*, 32, pp. 92–108.

Duflo, E., Glennerster, R. and Kremer, M. (2007) 'Using randomization in development economics research: A toolkit', *Handbook of development economics*, 4, pp. 3,895–3,962.

Hansen, K. and Vignoles, A. (2005) 'The United Kingdom Education System in an International Context', in Machin, S. and A. Vignoles (eds) *What's the Good of Education? The Economics of Education in the United Kingdom*, Princeton: Princeton University Press.

Higgins S., Kokotsaki D. and Coe R. (2012) *The teaching and learning toolkit: Technical appendices*, London: Education Endowment Foundation.

Ritchie J, Lewis J: *Qualitative research practice: a guide for social science students and researchers*. London: Sage; 2003.

Rouse, C., Krueger, A. and Markman, L. (2004) 'Putting Computerized Instruction to the Test: A Randomized Evaluation of a "Scientifically-Based" Reading Program', *NBER Working Paper*, 10315.

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', *Journal of 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', *Journal of Educational Psychology*, 105(2), p. 310.

Slavin, R. E., Lake, C., Davis, S. and Madden, N. A. (2011) 'Effective Programs for Struggling Readers: A Best-evidence Synthesis', *Educational Research Review*, 6, pp. 1–26.

Spybrook, J., Raudenbush, S. W., Liu, X., Congdon, R. and Martinez, A. (2008) 'Optimal Design for Longitudinal and Multilevel Research: Documentation for the "Optimal Design" Software', **http://sitemaker.umich.edu/group-based/optimal_design_software**

Wolgemuth, J. R., Savage, R. S., Helmer, J., Lea, T., Harper, H., Chalkiti, K., Bottrell, C. and Abrami, P. (2011) 'Using computer-based instruction to improve indigenous early literacy in Northern Australia: A quasi-experimental study', *Australasian Journal of Educational Technology*, 27, pp .727–750.

Appendix A

ICT intervention content and delivery

A. Content

The content of the ICT intervention is entirely encapsulated within a web-based technology platform, a modular 'game'–based balanced literacy intervention that is fixed in content (new activities cannot and will not be added). The games are based around and are linked in terms of content to a series of electronic texts (mainly 'stories', some non–fiction) suitable to beginner readers. That is, phonic fluency and comprehension activities are connected to and draw most content from, a series of electronic texts.

It is important to say that there are no additional 'extension' activities beyond those presented in the technology (e.g. for writing). One can by viewing the 20-week curriculum, the times associated with it, and the software have exactly the same experience of the intervention as the participants themselves.

The term 'balanced literacy' is held to mean any programme or curriculum that gives, over the course of the intervention, equal attention to 'word-level' reading skills (including fluency) and broader 'text-level' discourse comprehension (and fluency) skills. The current programme nevertheless starts with a focus on word-level skills in the early weeks and then adds comprehension skills and moves, as the programme progresses toward becoming entirely focused on comprehension skills as word-level fluency is achieved.

It is also important to note that 'word-level' skills includes skills aimed at promoting the phonemic decoding of individual words (the 'blending train' activity for example) as well as their associated sub-skills (most notably letter-sound knowledge and phonemic awareness tasks), but also includes activities supporting the direct 'sight reading' of words including words that are exceptions to phonic rules. These skills and the rationale for including them are described in detail in text before Tables 1, 2, and 3 below.

B. Delivery

The ICT platform delivers 'game'-based short literacy activities based around 17 fiction and non-fiction electronic stories/texts. The intervention described in the 20-week curriculum document attachment will be delivered through laptop computers running the software. The delivery is 4 x 15 minutes x 20 weeks of intervention. Unselected samples of ability-grouped year 1 children in small

groups of 3-4 will receive the intervention delivered on the computer and facilitated by a trained TA precisely following the curriculum and activity durations specified in the curriculum document.

TA training will consist of one full day of university-based training on the electronic resource for groups of 8-10 TAs, (9.00-4.00 PM) with 2 breaks plus short lunch plus a follow-up half-day session to problem-solve and reiterate learning prior to implementation.

Training day agenda:

Morning: This session is primarily led by the university facilitator and will detail (briefly) the background to, rationale and evidence for, the software resource using PowerPoint and questions and group discussion. Then TAs will consider elements of literacy (alphabetics, fluency, comprehension), what they look like in practice and how they can be taught. This element of the training will be sensitive to the fact that not all of the TAs will have prior knowledge and understanding of teaching literacy in the UK. TAs will be introduced to the software tool directly, be shown basic navigation, and activities and will have a short hands-on orientation to the software and discussion before lunch.

Afternoon: TAs will be provided with laptops with the software on it, will work in pairs and will develop a curriculum for a lesson, and then a whole week. They will then role-play the lesson to the group who will discuss and co-facilitate the activity. Open discussion, and the setting of homework (plan the first 3 weeks) will be set prior to a second and a short half-day follow up where TAs will come with questions and issues having tried the resource out 'live' with children not involved with the project.

The trained TA will have learned by the end of the day how to establish and maintain good group work competencies through the use of the STAR principle (Sharing, Turn-taking, Active attention, Responding) with a real star (sticker) as motivator for these core behaviours, linked to class teachers reward systems. The TA will introduce and time monitor activities and will, where appropriate, facilitate in the response to very open activities such as story response, previewing and predicting, recording successes and establishing whether children should progress to the next most difficult activity.

The intervention as described here and in the 20-week curriculum document reflects a cognitive skill-based approach to both word and text-level reading competencies but also one where children are exposed to stories that share content with all activities (text-linked cognitive strategies). As such the software and thus this curriculum was designed to embody much of the most

uncontroversial evidence-based practice for early reading as summarized in multiple narrative and systematic reviews of the literature (e.g. Ehri, *et al..*, 2001; National Reading Panel, 2000).

Skills taught in the intervention

Three forms of broad skills are taught in the intervention:

A. Alphabetics is the ability to associate sounds with letters and use these sounds to create words. The sounds associated with letters are referred to as phonemes (smallest units of spoken language) while the written letters associated with these sounds are called graphemes. There are 26 graphemes in English but over 40 phonemes. Research suggests that a limited number of phonics strategies and letter knowledge is essential to acquisition.

Children begin their path to reading by being able to distinguish between different sounds and develop their phonemic awareness. They can begin to associate these sounds with print and the correspondence between letters and sounds when they are given practice and modeling.

The ability to manipulate these phonemes either individually (letter by letter) or in chunks like rhyming families is a more sophisticated skill in the area of phonics. Here, the learner begins to segment words by breaking them apart into units that make it easier to read, then blend these units together. Decoding (ability to interpret symbols) then begins to take place, as the learner is able to access these strategies to read words. Research suggests children who do not have a solid foundation in these alphabetic principles are less skilled readers as they progress through school.

The software resource has 17 different activities specifically aimed at reinforcing the alphabetic principles. Each activity has multiple levels so that classrooms with a wide range of pupil needs can find an entry point. There is scaffolding built within each activity to encourage autonomous use of the tool.

Many of the alphabetics activities are associated with the leveled stories embedded in the software. This helps to build the context for practicing specific vocabulary. Also, the activities are appropriate for those pupils who are at the beginning stages of alphabetic skills development. These activities would focus predominantly on listening skills, auditory discrimination and letter naming. For those pupils who are ready for more advanced practice, the activities support word family manipulation, decoding games, and blending tasks with text support. Of course, there are fun games that provide practice like Letter Bingo and Letter-Sound Search. See Appendices for example screenshots of some activities.

Table 1: Detailed descriptions, leveling and group work for alphabetics activities

Activity Name	Description	Levels	Group activity
Letter Bingo	Given the pronunciation of the letter name, the pupil must determine if the letter is on his/her bingo card.	**Level 1** - Upper case letters **Level 2** - Lower case letters Challenge - Often problematic lower case letters (e, d, p, q, g, l, m, n, u, v)	This can be a group or individual activity. If played as a group, pupils should take turns finding one letter.
Animated Alphabet	Pupils will be able to hear the sounds made by certain single and clustered letters. These letters are then also used in fun alliterative sentences (e.g. for sound /r/ 'ruth rutherford's red robe ripped' (with an associated amusing graphic).	**Level 1**: The first page presents the 26 letters of the alphabet. **Level 2**: The second page has 11 of the most common consonant and vowel clusters for more advanced practice (sh, th, ch, or, ck, ai, ee, ea, ow, oo, oa).	As above, children can also support each other in this task.
Letter-Sound Search	Given the pronunciation of a letter sound, the pupil must identify and find its associated letter 'hidden' (entirely visible but placed in details) in a picture.	The teacher can have pupils focus on lower, upper, or mixed letters depending on the pupils' needs **Level 1** - 10 letters camouflaged in an easy background; i.e., s, p, m, t, c, d, n, a, h, r **Level 2** - 14 letters camouflaged in a medium background; i.e., g, l, o, k, u, b, e, f, l, z, d, h, p, r **Level 3** - 14 letters camouflaged in a hard background; i.e., w, j, v, y, q, x, g, i, b, l, a, t, e, h	
Same Phoneme	The pupil will be required to distinguish between phonemes (letter sounds) that are the same or different.	All letter sounds are presented in this activity.	This activity can be used by pairs (alternating after doing a pair of words) and

			as a whole class activity
Auditory Blending	Given a phonemic breakdown of a word, the pupil will identify a word; i.e. given /c/ /a/ /t/ child chooses "cat" from among 4 picture choices NOTE: if child get task wrong number of choices reduces from 4 to 3 (and as needed to 2) as support.	**Level 1**: 2-phoneme words with clusters, short, long, and r-controlled vowel sounds (e.g. a-t , ar-t). **Level 2**: 3-phoneme words with short vowel sounds. (e.g. c-a-t) **Level 3**: 3-phoneme words with clusters, short, long, and r-controlled vowel sounds (e.g. c-ar-t) **Level 4**: 4-phoneme words with blends mainly at the beginning, short, long, and r-controlled vowel sounds. (g-r-ai-n) **Level 5**: 4-phoneme words with blends mainly at the end, short, long, and r-controlled vowel sounds. (m=i-l-k) **Level 6**: 5-phoneme words with blends at the beginning and end, short, long, and r-controlled vowel sounds (e.g. p-l-a-n-t). **Level 7**: 5-phoneme challenge words (e.g. t-ar-g—e-t)	Pupils can do this activity in pairs and take turns (one word per turn) to tell the word and find its matching picture
Blending Train	Given the phonemic breakdown of a word, pupils will identify the word it makes; e.g., /r/ /o/ /d/ = "rod". NOTE: TA / peers must judge accuracy of articulation. If child struggles with task as a purely phonological one then a 'help' button provides	**Level 1**: 2-phoneme words with short and long vowel sounds. (e.g. a-t, ea-t) **Level 2**: 2-phoneme words with clusters, short, long, and r-controlled vowel sounds. (e.g. (e.g. ar.-t, t-ar) **Level 3**: 3-phoneme words with short vowel sounds (e.g. c-a-t-) **Level 4**: 3-phoneme words with clusters,	As well as group work, pupils can do this activity in pairs. They take turns (one word per turn) listening to the breakdown of a word then putting

	the letters to go with the sounds as they blend them.	short, long, and r-controlled vowel sounds (e.g. c-ar-t) **Level 5**: 4-phoneme words with blends mainly at the beginning, short, long, and r-controlled vowel sounds. (e.g. gr-ai-n) **Level 6**: 4-phoneme words with blends mainly at the end, short, long, and r-controlled vowel sounds. (e.g. m-i-l-k) **Level 7**: 5-phoneme words with blends at the beginning and end, short, long, and r-controlled vowel sounds (e.g. p-l-a-n-t). **Level 8**: 5-phoneme challenge words (e.g. t-ar-g-e-t).	the sounds together to tell the word they make. Equally all peers or pairs can judge whether word supplied by other child is correct or not.
Basic Decoding	Given the visual representation of a word, the pupil must say the sounds of the letters then blend them together to read the word. They must then find the correct picture (among three incorrect pictures) to match the word read. The help button in this case gives the sounds of the letters sequentially allowing the child just to focus on the blending and the selection of the correct picture to go with the word.	**Level 1**: 2-phoneme words with clusters, short, long, and r-controlled vowel sounds. **Level 2**: 3-phoneme words with short vowel sounds **Level 3**: 3-phoneme words with clusters, short, long, and r-controlled vowel sounds. **Level 4**: 4-phoneme words with blends mainly at the beginning, short, long, and r-controlled vowel sounds. **Level 5**: 4-phoneme words with blends mainly at the end, short, long, and r-controlled vowel sounds. **Level 6**: 5-phoneme words with blends at the beginning and end, short, long, and r-controlled vowel sounds.	This activity combines the sounding out of letters and blending them to make words. The teacher or other pupils can assist with scaffolding or picture matching when necessary.

		Level 7: 5-phoneme challenge words.	
Word Changing	The pupil must change individual letters in a word to form a new word. First a child identifies the letter that needs to be changed (e.g. when asked to change 'bow' to 'cow' Note: help is provided by computer if wrong letter is chosen (first, sounds given, or choice reduced in second tier of support, finally task is modeled if child still not successful)	**Level 1** - CVC words. Only the first letter is manipulated (e.g. 'b' 'and 'r' in bat-rat) **Level 2** - CVC words. All letters are manipulated (e.g *b*at-*r*at ba*t*-ba*g*, b*a*t-b*i*t) **Level 3** - CVC words with long vowel sounds. Only the first letter is manipulated (e.g. *d*ine-*l*ine) **Level 4** - CVC words with long vowel sounds. All letters are manipulated (e.g. *d*ine-*l*ine, di*n*e-di*v*e, d*i*ne-d*u*ne).	This activity can be done individually or in pairs. If a child makes a mistake, the teacher or a peer can encourage the pupil to pay attention to the required sound.

B. Comprehension

Comprehension is the cognitive process in which the reader interacts with a text in an attempt to understand what is taking place in the writing. Reading comprehension is a culmination of all the pupils' reading skills that have been gained. It involves good oral comprehension, vocabulary and decoding skills. (Depending on the grade level of the pupils, the importance of each of these prerequisite skills varies.) Comprehension is also affected by children's prior knowledge and personal experience that can help them relate to varied texts.

Reading for meaning is the ultimate goal of learning how to read. This is the element that takes children from being good decoders to being good readers and literate beings.

Research suggests that there are several strategies that have a positive influence on children's ability to understand what they read. The National Reading Panel (2000) suggests practicing skills like asking and generating critical questions, using story maps, and monitoring comprehension through cooperative learning opportunities.

There is also evidence that modeling appropriate strategies such as using context clues and allowing children to gain practice through providing them with multiple opportunities to interact with texts. Pupils should have time to read individually, with peers and with adults. Content here thus reflect these forms of evidence-based strategy.

This technology has focused on developing several specific skills that contribute to overall comprehension. These activities are leveled and range from simpler tasks like placing a well-known story in order (beginning, middle, and end) to summarizing an

entire text. Other activities allow pupils to respond to a question prompt encouraging pupils to think critically about the text and respond appropriately given the context of the story. In addition, the vocabulary activity helps pupils build a bank of words that they may be able to decode but do not know the meaning. See Appendices for example screenshots of some activities.

Table 2: Detailed descriptions, leveling and group work for comprehension activities

Activity Name	Description	Levels	Group activity
Prediction	Based on information from the story, the pupils will predict future events.	There are no 'levels' as such to this activity. Children predict for the story they are working with in 20-week curriculum. These stories are however in order of length in the curriculum so comprehension tasks increase in difficulty throughout for prediction/ summarizing activities etc. The story order is: 'I can move like a….' (an imitation activity), Beansprouts (non-fiction), The Dove and the Ant, The Frog and the Well, Henny Penny, the Little Red Hen (all fairy tales), Waterfalls (a longer folk story)	This activity is first done as an oral project. Pupils can also read the story as a group using the tracking function or have the page read by the computer using the audio icon. If done orally, encourage pupils to discuss their ideas with peers.
Comprehension Monitoring	After reading a story, the pupil will identify words that do not make sense. (There is one incorrect word on each page.)		Pupils can work as a group reading the story and looking for the nonsense words and take turns to be in control of the mouse and computer. If a word is not known, they can click on it to get help.
Sequencing	After reading a story, the pupil will place story events in their correct order.	**Level 1**: 3 story events (summary of story pages), displayed in random order.	Pupils can work as a group (reading the story events and discussing the appropriate

		Level 2: 5 story events (summary of story pages), displayed in random order.	order) and take turns to be in control of the computer.
Summarizing	The pupil will answer story related questions to help him/her summarize the story at the end.	Story related questions are asked after certain pages in the story.	This activity is best done in pairs or in small groups. Pupils read the story (using the tracking function) or have the pages read by the computer (audio feature). TA encourages discussion among the pupils.
Vocabulary	Given the definition of a word, the pupil must determine which of two sentences uses the word appropriately.	Some words are selected for each story.	This feature can be used to preview vocabulary words from a particular story before reading it or to work with words from a story after it was read.
Story Response	Pupils are given open-ended questions about the story that they will discuss with one another (e.g. did you ever try to help someone like the little red hen did?)	Six questions are presented in a set—three of which are general while the remaining three are story-specific.	This activity should be done in small groups. After pupils have discussed a question, they take turns scoring goals.
Story Elements	The pupils must answer questions about events that took place in a story (e.g. how did the story end?)	Pupils must select the answer to six questions about the story. Because there are audio prompts to assist, this activity can be done by all pupils	

		once they are familiar with the story.	

C. **Reading (Fluency)**

Fluency describes the level of automaticity that children have when reading a text. It is when learners are able to decode with little or no effort helping them to concentrate on the content of what they are reading as opposed to figuring out words and sounds. The sub-skills under the umbrella term fluency add to this skill correlating to comprehension. Pupils who read with correct pacing and expression with few or no mistakes are better able to focus on the meaning of the text.

Research suggests that pupils who are unable to acquire the skill of automatically reading words will struggle with attending to meaning. In addition, children who struggle with reading accurately and at an appropriate pace tend to struggle with motivational problems (Reading Rockets). The reading load increases as pupils progress through the grade levels making lack of fluency a major obstacle in school success.

The technology has a range of activities that allows for practice in reading fluency. The activities cover all of the sub-skill areas, including expression and speed. Through fun games, pupils can test their pace against cartoon characters. The technology has models of reading built into the story-related activities so pupils can monitor where they can improve. In addition, the programme models how to decode words within the context of the story so they are receiving additional and consistent practice at figuring out words. The more learners are able to practice a specific skill, the better the chance it has of becoming automatic. See Appendices for example screenshots of some activities.

Table 3: Detailed descriptions, leveling and group work for Fluency activities

Activity Name	Description	Levels	Group activity
High Frequency words	Given a list of high frequency words, pupils will have to read them out loud within a certain time period before another word pops up.	**Level 1**: A set of four words are randomly presented 4 times with gradually less time to read them **Level 2**: A set of seven words are randomly presented 4 times with gradually less time to read them	Pupils can do this activity in groups. TA reminds pupils that speed in their word recognition ability is an important element in their becoming good readers.

Tracking	While reading the story, the pupil will be able to read with one-to-one correspondence and without skipping words.	This activity is appropriate for all stories and levels.	Pupils can take turns to read the story and be in control of the computer (peers can monitor accuracy).
Accuracy	After listening to the computer read a passage, the pupil must read the same page accurately.	This activity is appropriate for all stories and levels.	This activity should be done individually (but peers can monitor and take turns reading it with least capable going last and benefiting from repeated observation of task). As the computer reads, the pupil follows. When s/he reads, s/he can click on a word if help is needed. At the end of each passage, the words clicked will be reviewed in a game-like feature to help the pupil become more familiar with them.
Expression	The computer reads a text using different expressions and the pupil must decide if the passage was read appropriately. The pupil must then read the same text practicing correct use of expression.	This activity is appropriate for all reading levels.	This activity can be carries out in turns by pupils and decision made collectively, or completed in pairs.
Reading Practice	This activity exposes pupils to a new story, which will allow for more reading	Suitable for skilled readers at end of intervention.	Repeated reading is shown to be an effective strategy in the development of

	opportunities and practice. This activity can be done in pairs or in small groups and individually.	NOTE pupil who can read new passages independently and who have completed all other activities may be eligible to 'graduate' (i. e. discontinue) the programme and return to full-time classroom teaching.	Fluency and other reading skills. We suggest that you have pupils reading passages orally multiple times while receiving guidance or feedback from peers, parents, or teachers.

Progression in activities

There is a very closely planned progression in both word-level and text-level activities from the less to the more complicated (with some elements of progression also in fluency activities). The aim in this is to allow all children to have appropriately challenging material throughout the intervention for their developmental level. These progressions are detailed for all lessons in a 20-week lesson plan (see 'Detailed 20-week Lesson Plan attachment'). As shown there, of each 3 elements of literacy (alphabetics, fluency, and comprehension) are carried out as streams that run at the same time through the 20-week intervention. There is a general overarching progression in the sense that at the star of the intervention there is a greater focus on children's word reading and associated alphabetic skills. There is a form of gradual 'hand-over' towards fluency and comprehension activities (that is, more time per 15 –minute session is devoted to these two activities later in the intervention as children's word reading progresses).

Beyond this the progression within the three streams is governed by a general rule. The **general rule** is that *children should on any given task be working at around 90%+ accuracy (this figure is drawn from quite conventional curriculum based assessment routines*). After 3 successes at this level on any given level a child (or group) moves on to the next level up in the ladder of difficulty. Typically this is monitored by the TA and it means that children can remain at a level until competency is evident. It can also mean on occasion that children (or groups) move backwards though activities if they have not shown sufficient competence at a given level (less than 80% accuracy at a given level). Progressions within streams are described below:

A) **Alphabetics/Phonics progressions**: This progression principle is easily mapped in phonic activities (that is, letter-sound knowledge and

phonological awareness activities) described in Table 1. For 2 of the 3 **letter-sound activities** (letter-sound bingo and animated alphabet) the progression is as follows:

Level 1 90% success rate 3 times in a row: move to next level
Level 2 90% success rate 3 times in a row: consider discontinue task

For letter-sound search the progression is as follows:

Level 1 90% success rate 3 times in a row: move to next level
Level 2 90% success rate 3 times in a row: move to next level
Level 3: 90% success rate 3 times in a row: consider discontinue task

For the majority of the phonological and decoding activities (namely the auditory blending, blending train, and basic decoding the progression is as follows:

Level 1: 2-phoneme words with clusters, short, long, and r-controlled vowel sounds.
Level 2: 3-phoneme words with short vowel sounds
Level 3: 3-phoneme words with clusters, short, long, and r-controlled vowel sounds.
Level 4: 4-phoneme words with blends mainly at the beginning, short, long, and r-controlled vowel sounds.
Level 5: 4-phoneme words with blends mainly at the end, short, long, and r-controlled vowel sounds.
Level 6: 5-phoneme words with blends at the beginning and end, short, long, and r-controlled vowel sounds.
Level 7: 5-phoneme challenge words. Level 1

In each case, 90% success rate 3 times in a row: move to next level, (consider) discontinuation at level 7.

For basic decoding there is a 4-level version of the same principle with the same 90% success discontinue rule:

Level 1 - CVC words. Only the first letter is manipulated (e.g. 'b' 'and 'r' in bat-rat)
Level 2 - CVC words. All letters are manipulated (e.g *b*at-*r*at ba*t*-ba*g*, b*a*t-b*i*t)
Level 3 - CVC words with long vowel sounds. Only the first letter is manipulated (e.g. *d*ine-*l*ine)

Level 4 - CVC words with long vowel sounds. All letters are manipulated (e.g. *d*ine-*l*ine, di*n*e-di*v*e, d*i*ne-d*u*ne). (Possible) discontinuation here.

Please note that as the web-based tool was designed to be entirely modular these common structural features of programmed learning are not built into the system but have to be superimposed through the curriculum and the adult supervising the sessions.

The precise details of these progressions are based on the sum of much research in the 1980s and 1990s on development of skills such as phonological awareness and word reading accuracy (full citations can be given for reviews of this work).

The progression for word-level skills is essentially built on a widely-accepted developmental model of the combined causal influence of letter knowledge and 'levels of phonological awareness' from simple phoneme blending tasks to complex phoneme segmentation tasks and their joint deployment in phonic tasks of graduated increased difficulty. Adjacent to these are sight word reading and text exposure activities that help to consolidate these skills in reading. These latter tasks at the higher end include reading text with appropriate speed and judging the correct prosody and emotional tone of a passage (and as such are classified as fluency activities here in this document).

B) Comprehension progressions

With the exception of sequencing which has two levels and which follows the same 90% progression rule articulated above for alphabetics), most comprehension activities are more open-ended and not all of them readily lends themselves to simple ordered progression rules, (though the accuracy on 3 successive occasions is still a useful heuristic where applicable, and the 90% rule can apply to Vocabulary, Story Elements and (to an extent) to Story Response).

More generally then, the progression for text-level skills as described in detail in the 20-week intervention) is essentially in the development of individual word knowledge (vocabulary) and easier 'receptive' tasks (no spoken response needed) of story ordering (e.g. 3 element stories) through after success at the first level, to longer and more challenging 5-element stories and comprehension monitoring activities to more difficult 'expressive' (verbal response) tasks of answering specific questions about a text and finally to key skills of predicting and summarizing parts of texts, and responding as a reader to texts. Comprehension tasks thus also graduate children to tasks of increased difficulty as they succeed.

Some task ordering is determined by logic: Prediction before a text (or part of a text) is read, Summarizing, and story response after reading text or text element.

Some activities (e.g. Vocabulary) can occur in both places, and some are best-used in real time during reading (like Comprehension Monitoring), so are ordered around each of the texts thus.

C) Fluency Progressions

The progression for fluency tasks is similar to that for the comprehension tasks in that they generally do not lend themselves to progression rules in the same way that phonic tasks in section A do.

Additional considerations

In all cases, for the precise progression of activities see the 20-week curriculum and the associated descriptions of each activity. Note that in all case children experience a rewarding (within technology) activity that follows the one they have completed (e.g. shooting a hockey puck at a goal after answering a comprehension question well), and children are often given some motivating choice of (selected) tasks at the end of a week of work.

It is important that the tool is used flexibly, to meet pupil needs, progression and discontinuation rules go some way towards this end of making sure the tool is not one size fits all and targets optimal levels for children who will arrive at different levels and progress at different speeds. Ability grouping should allow progressions to be similar across groups though an element of judgment of where the majority are operating will be needed. The progressions in the curriculum document assume a low-level of skills but it may clearly emerge that skills have been mastered in a given group (e.g. basic blending, all letter-sounds) and some testing of appropriate level needed to make sure pupils are not held up. Some of the information in this document can be given as a 'fact sheet' to TAs to aid comprehension.

Record Keeping

TAs will be encouraged to keep an activity log, detailing the activity of each session (see Appendices). The ICT intervention has built in to it a record keeping facility for both teachers and researchers. The teaching assistants can login to a teacher area of the software and access a teacher report that keeps track of the activities, levels, and performance data. There is a more detailed level of tracking data that can be accessed by researchers that tracks pupils, activities, levels, performance, and time spent on each activity.

References

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z. and Shanahan, T. (2001), Phonemic Awareness Instruction Helps Children Learn to Read: Evidence From the National Reading Panel's Meta-Analysis. Reading Research Quarterly, 36: 250–287. doi: 10.1598/RRQ.36.3.2.

National Reading Panel, (2000) Repot of the sub-groups. http://www.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf

Appendices

Example characters in the resource

Julie, the main character

These characters animate activities for children and provide a linking theme to the resource that children can engage with.

Example screenshots from ICT intervention activities

An example of a page that can be read by (or, by using the bird icon, to) child: The Little Red Hen story

An example of the phonic assistance given in stories if one clicks on a word (phoneme-graphemes are given for decodable 'regular' words, letter names for exception words).

Blending activities

An example of the leveling of activities in phonics, this one is for the Blending Train activity.

An example of the auditory blending activity for word ‘wet’ where bubbles are for the phonemes and the 4 choice answer pictures below are indicated. Children can click to hear the phonemes ‘w’-‘e’-‘t’ as often as they need.

Word changing Activity

Here, an example page from the Word Changing activity, here a child is supported in turning the word ’make’ into ‘made’ by fishing for the correct letter.

Comprehension Activity

This is the 3-item story ordering activity for Comprehension. Children click on the story sections 1-3 to hear them and then place them into the appropriate order.

ICT-led TA Literacy Lesson
Activity Log

TA's Name:	Year Level:
Date and time:	Activity:
Number of children participating:	ANY ABSENT CHILDREN?
Location of lesson:	Time taken to introduce the activity:

Please check the specific activities you undertook during this lesson.

Word Level Activities	**Specific activity name**	**Level and (group) success rate (90%+?)**	**Text Level Activities**
□ Blending □ Decoding □ Letter-sounds □ Segmenting □ Sight words			□ Story ordering □ Reading with expression □ Tracking □ Prediction □ Summarizing □ Sequencing □ Comprehension Monitoring □ Vocabulary development □ Story response □ Story elements Success level / rate?:

Additional comments on the above:

Performance Probes

How do you feel the technology lesson progressed today? (Were all the children engaged? What aspects were you particularly pleased / displeased with?), Can children progress? (what evidence is there for this?)

_____

_____

_____

Did you experience any technical difficulties? How was this handled? How much time was taken to resolve the difficulties?

_____

_____

_____

_____

_____

Did the session go as prepared? What is your comfort level with the programme and activities that were being used during this session?

_____

_____

_____

_____

_____

Please share any other observations you may have had regarding teaching and the children in regards using the technology i.e. class management, organization, teaching style, unforeseen interruptions, ability level differentiation.

_____

_____

_____

_____

____

_____

____

_____

____

_____

____

Appendix B

Non-ICT Intervention content and delivery

Content

The non-ICT intervention is defined by a 20 week programme of lesson plans based on pencil and paper activities, with additional resources such as magnetic letters and cards. The activities are linked to a series of texts. All resources and lesson plans are provided in a bespoke pack supplied to each TA at a training event prior to the intervention.

The intervention is a balanced one in that it encompasses both low level decoding activities and higher text level comprehension activities.
The content of the lessons are divided into 3 categories:

Decoding tasks (D) – to develop the ability to interpret symbols such as letters.

Comprehension tasks (C) – to develop skills in interacting with the text to gain meaning from the story.

Reading tasks (R) – to develop levels of fluency when reading text.

Activities are designed to be suitable for children with a range of abilities by dividing the activities into different levels, and pupils progress through the levels in a systematic manner.

Delivery

The lessons plans are designed to be delivered to groups of 4 pupils with each session lasting 15 minutes. Each lesson is carefully divided into timed sections and therefore each teaching element is designed to be achievable in the time give. Being able to remain within the allotted amounts of time is important as Teaching Assistants will have a number of groups to work through in an afternoon.

The lessons plans are designed to be delivered by Teaching Assistants who will have received a full days training on the content and delivery

of the materials. The Teaching Assistants will continue to be supported by just-in-time support by members of the research team should they need assistance during the intervention period.

Training

Each of the Teaching Assistants will be provided with a resource pack at the October training session which will contain all the paper based materials they require for the 20 weeks programme. This includes daily lessons plans for each of the 15 minute sessions and copies of the books they will share with the pupils.
Training will consist of a one day workshop to go through the intervention and programme. Plus a half day follow-up training after the TAs have had a chance to practice some sessions on non-participating children. Training will be sensitive to the fact that not all TAs will have received prior training in teaching literacy in the UK.

It is acknowledged that the Teaching Assistants who will be delivering this intervention will have differing levels of experience in teaching literacy skills that are required for the activities. Therefore a Key Facts sheet will be provided (see Appendix 1). The information in this sheet will also be covered in the training session prior to the intervention beginning.

An important part of the training which will be provided for Teaching Assistants before they begin to deliver the programme will focus on ensuring they feel confident to start pupils on the correct decoding levels, as well as how to progress through the levels.

Skills taught in the intervention

The content of the lessons are divided into 3 categories, and each category has a number of activities in the lessons plans related to it.

Decoding

The decoding related tasks in the programme acknowledge that pupils will be entering with differing levels of skills (for example some pupils may begin the programme having consolidated all the initial letter sounds whereas others may still have gaps in their letter sound knowledge). These lesson plans are designed to provide flexibility for the Teaching

Assistants to start pupils on decoding activities which are appropriate for their current level. This will ensure that pupils do not become bored completing work they have already consolidated or frustrated through being offered activities they cannot complete.

In terms of the *Alphabet* activities (which introduce different phonemes and are used on a daily basis) if it is agreed that the pupils in the group have already consolidated these skills the amount of time allotted to this activity will be reduced to 1 minute to act as a quick and familiar starter activity for each session. As pupils progress through the Alphabet activities it may be necessary for the Teaching Assistant to increase the amount of time again if pupils begin to encounter phonemes they have not consolidated. If time is reduced on this activity it is added to the either the Comprehension or Reading activity section (this is noted on the lesson plans).

Activities	Description	Resources
Letter Bingo	Given the pronunciation of a letter name, the pupils must determine if they have the letter on their bingo card.	Photocopiable bingo boards and letters are provided in TA resource pack.
Alphabet	Pupils must identify initial/ middle or final sounds in words.	Alphabet letter flashcards are provided in the TA resource pack.
Blending	Given the phonemic breakdown, the pupils will identify a word.	For both *Blending and Decoding* activities TAs are provided with a range of suggested suitable activities which they can chose from depending on the needs of the group they are working with and the resources available in the school they are working in.
Decoding	Given the visual representation of a word, the pupil must say the sounds of its letters, then blend them together to read the word.	No resources required.
Same Sounds	The pupils will distinguish between letter sounds that are the same and different.	
Letter-Sound Search	Given the pronunciation of a letter sound, the pupils must identify and find its associated letter.	Photocopiable activity sheets are provided in the TA resource pack.
High Frequency Words	Given a high frequency word the pupils will read it correctly.	Photocopiable high frequency word flashcards are provided in the TA resource pack.
Word Changing	Pupils change individual letters in a word to form a new word.	The TA can use either a whiteboard or magnetic letters to create word changing words. A list of suitable words is given in the TA resource pack.

The programme is designed to ensure that pupils progress to the next level of activities once they have (as a group) achieved an 80-90% success (repeated 2 to 3 times in a row).

In terms of the *Blending* and *Decoding* activities pupils work through 7 levels:

LEVEL	ACTIVITY
1	2-phoneme words with clusters, short, long, and r-controlled vowel sounds
2	3-phoneme words with short vowel sounds
3	3-phoneme words with clusters, short, long, and r-controlled vowel sounds
4	4-phoneme words with blends mainly at the beginning, short, long, and r-controlled vowel sounds
5	4-phoneme words with blends mainly at the end, short, long, and r-controlled vowel sounds
6	5-phoneme words with blends at the beginning and end, short, long, and r-controlled vowel sounds
7	5-phoneme challenge words

Table 2: Blending and Decoding Levels

As with the Alphabet activities Teaching Assistants will be supported through the initial training to identify a suitable starting point for pupils. They will then work through the different levels recording the results of each session on the Tracking Sheet (see Appendix 2) to monitor when pupils have achieved the group success rate and need to move up a level. No specific blending activities are given in each lesson plan instead a list of appropriate activities are provided (see Appendix 3). This will ensure that Teaching Assistants are able to select activities which will be motivating to the groups they are teaching and which the school have appropriate resources for.

In terms of the *Word Changing* activities the pupils work through 4 different levels:

LEVEL	ACTIVITY
1	CVC words. Only the first letter is manipulated.
2	CVC words. All letters are manipulated.
3	CVC words with long vowel sounds. Only the first letter is manipulated.
4	CVC words with long vowel sounds. All letters are manipulated.

Table 3: Word Changing Levels

It is intended that pupils will all begin at Level 1 with the Teaching Assistant moving them through the levels when they have achieved the 80-90% success rate (repeated 2 to 3 times in a row). There are example words given for each Word Changing level which Teaching Assistants

can use and then adapt if necessary (see Appendix 4) and also a Work Changing Tracking table (see Appendix 5).

Comprehension

A range of comprehension activities are contained within the lesson plans. In order to support Teaching Assistants in the comprehension activities prompts sheets (known as Top Tips sheets) will be used (similar sheets will also be provided for the Reading activities). These will be same across both the Teaching Assistant led and the computer based programme conditions (for an example Top Tips Sheet see Appendix 6).

Vocabulary	The pupils will explore the meaning of single words.	Single pages from each story with the target words highlighted are provided in the TA resource pack.
Sequencing	After reading a story the pupils will place story events in their correct order.	A series of either 3, 5 or 8 story cards are provided in the TA resource pack for pupils to order.
Monitoring	After reading a story the pupils will identify words that do and do not make sense.	Pupils are either presented with a picture and sentence from the story with a word missing which they must identify from the options given or presented with questions about the story which must be discussed as a group and answered. These are provided in the TA resource pack.
Prediction	Based on information from the story, the pupils will predict future events.	The TA will read through a new story with the pupils asking the prediction questions provided in the TA resource pack.
Story Response	Pupils are given open-ended questions about the story that	A list of question to ask pupils is provided in the TA resource pack.

	they will discuss with one another.	
Story Elements	The pupils must answer questions about events in the story.	A list of questions to ask pupils is provided in the TA resource pack.

Most of the comprehension activities are not explicitly structurally levelled like the decoding activities, with the exception of the Sequencing activity.
However, children still make progress through the comprehension activities through the incremental complexity of the sequence of activities as presented through the 20 week plan (available on request), and the complexity of the stories. For example, early activities are limited to single word level semantic work (Vocabulary) whereas later activities such as predicting future events in texts (Prediction) are more challenging.

Reading

Throughout the 20 week programme a range of books are used to support the reading activities (see attached file of books). The titles of the books are as follows:

Weeks 1 & 2 – I Can Move Like A….
Weeks 3, 4 & 5 – How a Bean Sprouts
Weeks 6, 7 & 8 – The Dove and the Ant
Weeks 9, 10 & 11 – The Frogs and the Well
Weeks 12, 13 & 14 – Henny-Penny
Weeks 15, 16 & 17 – The Little Red Hen
Weeks 18, 19 & 20 - Waterfall

Activities	**Description**	**Resources**

Tracking	The pupils will read with one-to-one correspondence and without skipping words	The pupils with read from the stories provided in the TA resource pack. A list of tracking tips is provided for the TAs.
Accuracy	After listening to a passage being read, the pupils will read the same accurately.	The pupils will use the stories provided in the TA resource pack.
Expression	The pupils will listen to a passage being read and decide whether it is being read with good or bad expression. The pupils will then read the passage.	The TA will read from the stories provided in the TA resource pack.
Speed	The pupils will read the text at an appropriate speed.	The pupils will use the stories provided in the TA resource pack.

Each pupil in the group will require access to a paper copy of the book. A range of reading activities are provided throughout the 20 week programme. As with the Reading activities Teaching assistant will be provided with a 'Top Tips' sheet for each activity and these sheets will be the same across both the Teaching Assistant led and computer based programme conditions.

Additional considerations

Within the lessons plans from week 12 onwards Teaching Assistants will be given more flexibility in deciding which activities to present to pupils depending on levels already achieved. Teaching Assistants can choose between Decoding, Word Changing or Tracking activities. Training on how to choose the correct activities will be given at the training event. In addition to this, support will also be provided around week 10 to staff by a Senior Research Assistant either through face-to-face meetings in

school settings or through a newsletter to ensure they feel confident in choosing the correct activities.

It is acknowledged that there may come a point within the 20 week programme where pupils have completed all the levels on the decoding activities and that Teaching Assistants feel they have successfully completed the comprehension and reading activities. Assessment will be provided to ensure that Teaching Assistants feel confident when to discontinue pupils from the programme. This assessment will be covered at the training event in October and further assistance can be given to the Teaching Assistants by a Senior Research Assistant if required. The discontinuation criteria for both the ICT and non-ICT interventions will be the same.

Record Keeping

Teaching Assistants will be encouraged to keep a short record of every session they complete with pupils (see Appendix 7). This will be particularly important if they have to deviate from the lesson plans for any reason.

APPENDIX 1 – KEY FACTS SHEET

Key Facts

- The English language uses approximately 44 sounds (20 vowels and 24 consonants)
- The English Alphabet has 26 letters
- English uses combinations of letters to represent single sounds
- There are 140 ways of combining letters to create the 44 sounds.

Letters: a b c d e f g h I j k l m n o p q r s t u v w x y z

Phonemes: /b/ /k/ /d/ /f/ /g/ /h/ /j/ /l/ /m/ /n/ /p/ /r/ /s/ /t/ /v/ /w/ /y/ /z/ /sh/ /ch/ /th/ /th/ /ng/ /zh/ /a/ /e/ /i/ /o/ /u/ ai/ /ee/ /igh/ /oa/ /oo/ /oo/ /ow/ /oi/ /ar/ /or/ /ur/ /air/ /ear/ /ure/ /er/

A *phoneme* is the smallest unit of sound in a word

A *grapheme* is letter(s) representing a phoneme

Phonemes are represented by letters (graphemes)

- A child need to learn the letters that make up each sounds, this is known as phoneme-grapheme representation
- Phonemes can be in the initial, medial or final position of a word.

A phoneme can be represented by one or more letters

- A single phoneme represented by 2 letters or more e.g.

  ch – ai – n

A Diagraph is two letters which make one sound

- A consonant diagraph contains two consonants

  sh ck th ll

- A vowel diagraph contains at least one vowel also known as long vowel phoneme

  ai ee ar oy

- Three letters, which make one sound

  igh dge

The same phoneme can be represented/spelt in more than one way

- This is particularly common among the vowels

  rain, may, lake

The same spelling may represent more than one phoneme

- This is where children need to learn to use the skill of making sense of text

  mean, deaf

Adjacent Consonants (formally known as blends) are letter combination where each letter makes an individual phoneme

sp	**st**	**sk**	**sl**
tr	**ft**	**nt**	**lt**
mp	**un**	**lp**	**dr**
cl	**sw**	**cr**	**sm**

Split Diagraphs are diagraphs in which the two letters are not adjacent

a_e	**e_e**	**i_e**	**o_e**	**u_e**

APPENDIX 2

Blending and Decoding Tracking

Blending Tracking

Pupils work through the levels. Once they have achieved a correct response rate of 80-90% as a group for three consecutive entries you can move onto the next level.

Week and Day	Level	Number of Correct Examples	Number of Incorrect Examples	Achieved (Y/N)	Notes

APPENDIX 3 – BLENDING ACTIVITIES

Teaching Assistants can chose from the range of activities depending on what they feel would be most suitable for their groups and the resources available in school. These different activities will be explored further in the training event.

Magnetic letters

- Facilitator places magnetic letters on a board with a gap between each phoneme. Support the pupils to sound out each phoneme and then push them together to create a word. If there are enough resources pupils could have their own board.

Jigsaws

- Facilitator to write letters onto jigsaw pieces. Support the pupils to sound out each letter. Then complete jigsaw and sound out full word.

Train Carriages

- Write each phoneme on a piece of card and attach to toy train carriage. The facilitator should help the pupil to sound out each phoneme and blend them together to create a word.

I Spy

Play *I Spy* by sounding out the name of the object you are looking at.

- Pupils have to blend the sounds together to determine the object
- E.g. *I spy a /b/ /a/ /ll/ (ball).*
- If pupils require additional support you can add a clue to begin e.g. *I spy something that is round.*

***This is only a sample of the activities which will be suggested to Teaching Assistant. Further examples will be added.*

APPENDIX 4 – WORD CHANGING EXAMPLES

LEVEL 1

fan / man / can
cat / sat / mat
mad / dad / sad
let / met / vet

LEVEL 2

big / bag / bad
hit / hat / cat
yet / get / got
Dan / dad / did

LEVEL 3

cage / sage / page
bake / cake / rake
mice / nice / rice
dine / mine / fine

LEVEL 4

tide / time / tame
dine / fine / find
mine / pine / pile
like / line / fine

APPENDIX 5

Word Changing Tracking

Pupils work through the levels. Once they have achieved a correct response rate of 80-90% as a group for three consecutive entries you can move onto the next level.

Week and Day	Level	Number of Correct Examples	Number of Incorrect Examples	Achieved (Y/N)	Notes

APPENDIX 6 – READING PROMPTS

An example of a Reading Prompt sheet. These will be further developed. The Reading Prompt Sheets will be the same in both the Teaching Assistant led and computer package conditions.

Top Tracking Tips

- When accessing stories the facilitator first should read the required pages.
- Feel free to ask questions about the illustrations, other prediction question, their opinion, their familiarity with the story.
- The pupils will not be able to read many words and we do not want them to feel pressure to do so.
- When finished reading the selected pages go back to the beginning. For each page ask if anyone can identify a word or two, or maybe even a letter if necessary.
- Read each page. Ask the prediction / summary questions. Encourage each pupil to comment.
- Use up any extra time by reviewing and discussing what has been read.

APPENDIX 7 – TEACHING ASSISTANT RECORD SHEETS

WEEK: **DAY:**	**PUPILS ATTENDED:**
ACTIVITIES COMPLETED	
ACTIVITY	**NOTES**
ANY ISSUES?	

Appendix C: Permission letters

Coventry University
Priory Street
Coventry CV1 5FB
Telephone 024 7688 7688

Dr Linda Merriman
Dean of Faculty

DATE.

Dear Headteacher

First of all, please allow me to apologise for contacting you somewhat out of the blue. My name is Prof. Clare Wood and I am the Director of the *Centre for Applied Research in Psychology* at Coventry University (www.coventry.ac.uk/CARP).

Our team recently applied for funding from the *Education Endowment Foundation* for a project which would evaluate the educational impact of training Teaching Assistants to deliver Small Group-based Support for Literacy. We have been successful in securing funding for this project, and so we are now looking to talk to local schools who may be interested in participating in the project. The project itself would commence in September 2014, and the intervention phase would last for 20 weeks, starting after the Autumn half term break.

I appreciate that you are likely to be very busy at this time of year, but if you think that you may be interested in discussing what would be involved from your point of view and how your school might benefit directly from participation, I would be very happy to meet with you or talk to you on the phone at a time which is convenient for you. I am very keen to involve as many Coventry schools as possible so that we have a strong local benefit and impact from the project, which is looking to recruit large numbers of children across the Midlands.

If you think that you may be interested, I can be contacted via email on c.wood@coventry.ac.uk or by telephone (Monday – Thursday) on 02476 888226. Many thanks for time and I very much hope to hear from you in due course.

Yours faithfully

Prof. Clare Wood
Professor of Psychology in Education

Coventry University
Priory Street
Coventry CV1 5FB
Telephone 024 7688 7688

Prof Guy Daly
Dean of Faculty

12 June 2014

Dear Parent

My name is Dr Janet Vousden and I am a member of the Centre for Research in Psychology, Behaviour and Achievement (http://www.coventry.ac.uk/CRPBA). We recently secured funding from the *Education Endowment Foundation* for a large-scale project which will evaluate the educational impact of training Teaching Assistants to deliver small group support for reading in Year 1. This project will run across 60 schools in the West Midlands and your child's school is participating in this project. I am therefore writing to you to make you aware of what is involved in the project and to offer you the chance to raise any questions about it with me directly, or with your school. Please feel free to contact me by email or phone if you have any concerns. I can be contacted via email on janet.vousden@coventry.ac.uk or by telephone (Monday – Thursday) on 02477 659510.

I have attached an information sheet which explains in simple terms what is involved. We hope that as many pupils as possible will be able to participate but we also want to offer you the chance to opt out of the project, if you so wish. We will also be asking the individual children if they are happy to take part, and we will respect their wishes if the project is not for them.

Please return the reply slip at the bottom of this letter to your school teacher as soon as possible and no later than the **week before the end of term** to notify me of your wishes. If we do not hear from you by this date we will assume that you have no objections and your child may be asked to take part.

Yours faithfully

Dr Janet Vousden
Research Fellow in Reading Development

Year 1 Reading Project

☐ I confirm that I have read the participant information sheet sent by Dr Janet Vousden in relation to the Year 1 Reading project.

☐ I agree to my child being asked to participate in the project.

☐ I do not want my child to be asked to participate in this project.

Child's Name _____

Parent's Name and Signature _____

Participant Information Sheet

Study title:
An Evaluation of Teaching Assistant-Based Small Group Support for Literacy

What is the purpose of the study?
We wish to evaluate the impact of two different approaches to supporting Year 1 children's reading attainment through the provision of small group reading support delivered by trained Teaching Assistants. This project is being conducted across a large number of schools in the Midlands.

Why have we been approached?
We are looking to recruit Year 1 pupils from schools in the Midlands where the project team is based. This is the only reason why you have been approached.

Do we have to take part?
No, participation is entirely voluntary and there are no consequences if you or your child decides not to take part.

What will happen to my child if s/he take part?
If your child takes part, they will be assessed on his/her reading and language skills by standard tests provided by Hodder, an education publishing company. This will happen for all children at the beginning of Autumn Term (2014), at the end of Summer Term (2015) and again at the end of Summer Term (2016) to evaluate the impact of the study. Your school will be randomly allocated to either take part in the intervention in 2014/2015, or to be in a control group that will not be offered the intervention(s) until 2015/2016. If your child takes part, they will either: a) be randomly selected to experience either one of two different types of small group support for reading or normal classroom activities in 2014/2015, or b) experience either one of two different types of small group support for reading in 2015/2016. One small group support group will complete computer-based activities, the other will complete non-computer-based activities. The intervention will last for 20 weeks.

What are the possible disadvantages and risks of taking part?
There are few disadvantages and risks. The reading assessments do take a little time to complete but we will ensure that they are completed as a time when it will cause minimal disruption to your child's school work. The assessments may feel challenging and some children may feel a little self-conscious about completing them (for example, if their reading is not as good as they would like it to be). However, all results will remain confidential to the research team and we will do our best to put your child at ease throughout. If your child is allocated to a control group, this may feel like a relative disadvantage, but it should be noted that there is no disadvantage insofar as the children in these groups will continue to receive the high quality literacy support that they would normally receive from their classroom teachers.

What are the possible benefits of taking part?
This project will enable schools to discover the best way of using Teaching Assistants to deliver more individualised teaching support for reading. We will provide opportunities for parents who are interested to find out more about the interventions that we are running and how they can also support their children's work. We expect the results of this research to benefit your children's school in future years, with consequences for how Teaching Assistants are subsequently used in the school to help support children's reading.

What if something goes wrong?
You or your child can indicate to the teacher or the reading group leader if either of you no longer wishes to take part, and you can leave the study without question. If you are unhappy with the conduct of the study, you can contact me directly in the first instance using the number at the end of this sheet. You are free to withdraw at any point during the study, and for up to one month following the completion of

the study. You can do this by contacting me and giving me your child's name and the name of his/ her school. If you are unhappy with the conduct of the research team, you can contact me (Dr Janet Vousden) in the first instance, or Prof. Neil Forbes, who is chair of the University Ethics Committee (address: Coventry University, Priory Street, Coventry, CV1 5FB).

Will my taking part in this study be kept confidential?
Yes. Pupils' test responses and any other pupil data will be treated with the strictest confidence. The reading test responses will be scored by the service provided at Hodder and accessed by the team at London School of Economics. Named data will be matched with the National Pupil Database and shared with the Coventry University team, the independent evaluators at London School of Economics and the Education Endowment Foundation. We will not use your child's name or the name of the school in any report arising from the research. Reading test sheets will be stored in a locked cabinet and destroyed as soon as the scores have been inputted into a computer file. All consent forms will be stored in a separate, locked location and will be destroyed on completion of the project.

What will happen to the results of the research study?
The reading test data will be used as the basis of a report to be compiled by colleagues from the London School of Economics, which will be submitted to the Education Endowment Foundation. Reports based on the reading scores will be presented at academic conferences and it may also be written up for publication in peer reviewed academic journals. Crucially, all participating schools and families will be informed as to the outcomes of the project overall, and we will host special events which families and non participating schools can attend to find out more about the project.

Who is organising and funding the research?
The research is organised by Dr Janet Vousden, who is a Research Fellow in the Centre for Applied Research in Psychology. It is funded by the Education Endowment Foundation.

Who has reviewed the study?
The Education Endowment Foundation and the Coventry University Ethics Committee have reviewed and approved this study.

Contact for Further Information
Dr Janet Vousden
Psychology Department
Coventry University
Priory Street
Coventry
CV1 5FB.

Tel: 024 77 659510
Email: janet.vousden@coventry.ac.uk

Appendix D: Schools included in the process evaluation

School	Location	Size (pupils in intervention)	Ofsted	Number of groups	Number of pupils in intervention	Number of TAs	TA interviews	ICT session observations	Non-ICT session observations
A	Leicester	93	Requires improvement	7 Non ICT 8 ICT	27 Control 28 ICT 26 Non-ICT	2 ICT 2 Non-ICT	2 Non-ICT TA 2 ICT TA	2	2
B	Coventry	17	Good	2 Non ICT 1 ICT	4 Control 6 ICT 5 Non-ICT	1 ICT 1 Non-ICT	1 Non-ICT TA 1 ICT TA	1	1
C	Leicester	45	Requires improvement	2 Non ICT 3 ICT	7 Control 9 ICT 8 Non-ICT	1 ICT 1 Non-ICT	1 Non-ICT TA 1 ICT TA/LSA	3	0
D	Birmingham	120	Outstanding	10 Non ICT 10 ICT	35 Control 36 ICT 38 Non-ICT	2 ICT 2 Non-ICT	1 Non-ICT TA 2 ICT TA	3	1
E	Coventry	60	Good	3 Non ICT 3 ICT	15 Control 13 ICT 11 Non-ICT	2 ICT (job share) 1 Non-ICT	1 Non-ICT TA 2 ICT TA	1	1
F	Coventry	60	Good	3 Non ICT 3 ICT	8 Control 10 ICT 9 Non-ICT	1 ICT 1 Non-ICT	1 ICT TA 1 Non-ICT TA	3	0
Total		395		55		17	16	13	5

Appendix E: Padlock rating

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	Power	Attrition*					
5	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%					
4	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%	4		Adjustment for Balance [0]		
3	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%					
2	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%			Adjustment for threats to internal validity [0]		
1	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	51-50%					
0	No comparator	MDES > 0.6	<50%					

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = 4 padlocks
- **Reason for adjustment for balance** (if made): none as the majority of pupil characteristics and all of the outcome measures are balanced at baseline.
- **Reason for adjustment for threats to validity** (if made): none required
- **Final padlock score:** initial score adjusted for balance and internal validity = 4 padlocks

*Attrition should be measured at the pupil level, even for cluster trials.

Appendix F: Cost rating

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

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

Addendum – added March 2018

Executive Summary

The project

The original report "*ABRA: Online Reading Support*" summarises the findings of the ABRA randomised controlled trial from the year that the interventions were implemented. This addendum report analyses the impact of the ABRA interventions one year after the end of interventions This analysis was specified in the original trial protocol as a method to assess whether any impact on reading outcomes seen in the year of the intervention translated into impact on Key Stage 1 (KS1) national tests.

As in the original report, we analyse the effects of two methods of small group tuition for Year 1 pupils in English schools on measures of literacy. One of these methods used an ICT programme (Abracadabra) which has some supporting evidence from small randomised controlled trials (RCT) efficacy studies in other countries (see, for instance, Comaskey, Savage and Abrami, 2009; Savage et al., 2009; and Wolgemuth et al., 2011) and a bigger effectiveness trial (Savage et al., 2013). The other method is identical in content (i.e., used materials that replicate the ICT intervention), but does not use the computer programme (Non-ICT intervention). Both the ICT and Non-ICT intervention were delivered by teaching assistants (TAs), during a 20-week program and focused on phonics, fluency and comprehension activities that were connected to and drew most content from, a series of texts. Pupils in the treatment conditions were put together in small groups (3 to 5 pupils), and received 15 minutes of intervention 4 times per week. There were 51 participating schools and 2241 pupils at randomisation.

Summary of previous results

A summary of the short-term impacts described in the original report is as follows: both the ICT and Non-ICT treatments were found to have positive results for literacy that were unlikely to have occurred by chance, although the size of the Non-ICT effect was found to be considerably higher (1.6 times bigger). FSM children benefited more strongly from the two treatments than non-FSM pupils. Pupils with below median pre-test outcomes seemed to benefit more from the ICT treatment than those with above median pre-test outcomes, whereas both below and above median pupils benefited from the Non-ICT treatment equally. Finally, our results suggested that there were *spillover* effects within treatment schools as the within--school control group had been affected by the treatment relative to pupils in control schools. This is discussed in the original report. Successful implementation resulted from a well-designed and well-delivered training programme which emphasised fidelity and consistency, enforced by on-going support from the project team.

Summary of new results

The results of this addendum report indicate that an impact on attainment was still observable a year after the intervention, as measured by nationally prescribed teacher assessments at the end of Key Stage 1 (end of Year 2). The greater benefit for FSM pupils remained.

In particular, the Non-ICT treatment has had an average positive effect on reading and writing attainment one year after the intervention, whereas the ICT treatment is only found to influence positively the KS1 writing outcome. However, these average effects differ by subgroup as discussed in the conclusions below.[22]

In line with the results from the first report, the results in this addendum report suggest that FSM students in both treatments groups (particularly those in the Non-ICT treatment group) are more likely to be

[22] Note that the subgroup analysis needs to be interpreted with greater caution as the randomisation was not stratified by the characteristics considered here: gender, disadvantage, prior attainment.

observed working at the expected reading (and writing) standard a year after the end of the intervention, compared to the control group.

Students with above median prior attainment seem to experience a greater impact on both reading and writing outcomes, but these effects are mainly driven by higher ability FSM students. In terms of gender, females appear to benefit from the interventions to a greater extent than males (although only females in the Non-ICT treatment show positive significant effects in terms of reading a year after the end of the intervention). Finally, the *spillover* effects (from treated students to control students in treatment schools) seem to have vanished a year after treatment.

Key conclusions
1. Pupils originally allocated to the Non-ICT intervention are on average 6 percentage points more likely to be observed working at the expected (or higher) **reading** standard a year after the end of the intervention. This can be compared to the mean of the outcome variable in control schools, where 74% of students are working at this level This corresponds to an effect size of 0.14, or 2 months' progress.
2. In terms of **writing**, average effects are also found for pupils who were allocated to the ICT and non-ICT interventions. The effect sizes are 5 percentage points and 9 percentage points respectively, one year after the intervention. This can be compared to the mean of the outcome variable in control schools, where 62.5% of students are working at this level**.** These results correspond to effect sizes of 0.10 and 0.18 respectively, around 2 months progress in each case.
3. The analysis by **subgroup** reveals that **FSM** students appear to reap larger benefits from the interventions: For FSM pupils the Non-ICT intervention increases the probability of achieving the expected reading standard by 11.5 percentage points (This corresponds to an effect size of 0.26). Both the ICT and Non-ICT interventions are found to increase the probability of achieving the expected writing standards, by 11 and 16 percentage points, respectively (These results correspond to effect sizes of 0.23 and 0.33).
4. Students with above median prior attainment in the baseline test are more likely to benefit from both interventions, both in terms of reading and writing outcomes. These results are driven by FSM students that scored above the median at the baseline test.
5. The initial *spillover* effects found for children in the control category of treatment schools disappear when analysing outcomes a year after the end of the intervention.

1. Introduction

Intervention

This addendum report analyses the impact of two literacy related interventions one year after they ended. The original report ("*ABRA: Online Reading Support*") summarised the findings of this randomised controlled trial on literacy outcomes within the same academic year as the intervention. Specifically, in the original report we analysed the effects of two methods of small group tuition for Year 1 pupils in English schools on measures of literacy. One of these methods used an ICT programme (Abracadabra), which has some support from small RCT efficacy studies in other countries (see background evidence on the original report). The other method is identical in content (i.e., used materials that replicate the ICT intervention), but does not use the computer programme (Non-ICT intervention). The analysis of the short-term outcomes revealed that both the ICT and Non-ICT treatments had positive results for literacy that were unlikely to have occurred by chance, although the size of the Non-ICT effect was considerably higher. We refer the reader to the original report for a more in-depth description of the intervention and the short-term results.

Evaluation objectives

The evaluation is designed to answer the following questions, as described in the protocol.

1. Do the interventions have an impact on the treatment groups relative to the control group on outcomes within one academic year?
2. How do the interventions compare to each other in terms of effectiveness?
3. Is there evidence of heterogeneity in the effects of the interventions on the following groups: free school meal status; gender; above median attainment on pre-test.[23]
4. Is there any evidence of *spillover* effects on the control group?
5. Do any effects persist one year later?

Questions 1 to 4 were dealt with in the original report. We address question 5 in this addendum report. As stated in the protocol, we will use Key Stage 1 reading and writing measures (that are measured one year after the end of both treatments, using teacher assessment).

Additionally, Coventry University also tested students one year after treatment using the general reading ability test, which was used in the original analysis (PIRA, described in the next section). Although this analysis of PIRA outcomes is not covered in the protocol, these results are presented in the Appendix of this addendum report.

There is no process evaluation in this addendum report. Please refer to the original report on the findings pertaining to this intervention.

Project team

The project team who developed and delivered the intervention are based at Coventry University and led by Professor Clare Wood and Dr Janet Vousden. The other team members are Rob Savage, Helen Johnson, Sabrina Ammi, Claire Pillinger, Sam Waldron and Rosa Kwok. The ICT 20-week programme was developed by Dr Robert Savage, Associate Professor at McGill University, based around the ABRA programme. The ABRA programme is a freely available computer-based programme, widely used in Canada (Abrami et al., 2010; Wolgemuth et al., 2013; Abrami et al., 2014, Cheong et al., 2016).

[23] Heterogeneity according to the first language spoken by the pupil cannot be assessed due to a lack of data. This is because the consent forms sent to parents ahead of testing were opt-out consent forms. In order to be able to receive sensitive NPD data, parents should have given explicit consent (opt-in consent forms).

The evaluation team is formed by Professor Sandra McNally and Dr Jenifer Ruiz-Valenzuela at the Centre for Economic Performance, London School of Economics, and Dr Heather Rolfe at the National Institute of Economic and Social Research (NIESR) who led the work on the process evaluation for the original report.

2. Methods

Trial design

The results in this addendum report are based on a randomised controlled trial (RCT) with randomisation taking place at the school (first stage) and at the pupil (second stage) level. Teaching Assistants were also randomised into different treatments. In the first stage, recruited schools were paired according to predetermined characteristics (see original report). Each school within each pair was randomised to either cohort 1 (i.e. receives the treatment(s) in Year 1) or cohort 2 (receives the treatment(s) in Year 2).[24] In the second stage, within treatment (cohort 1) schools, children in Year 1 were randomly allocated into three groups:

- (A) Support from Teaching Assistants using the computer programme (ICT intervention);
- (B) Support from Teaching Assistants using other methods (Non-ICT intervention); and,
- (C) Control group within treatment schools

That is, the design included two control groups. One group consisted of pupils in purely control schools, and the second group included pupils in group (C) described above (i.e. pupils in the control group of treatment schools). This design allowed us to investigate the existence of *spillover* effects from treatment to control pupils in those schools randomly assigned to treatment (i.e., those in cohort 1). Finally, teaching assistants in treatment schools were randomised to deliver the ICT or the Non-ICT intervention. For full details see the main report.

Outcome measures 1 year after treatment

As stated in the protocol, the primary outcome measures analysed in this addendum report are constructed from Key Stage 1 outcomes for the academic year 2015-2016. At the end of KS1 teachers use teacher assessment judgements to report on the progress of their pupils. These assessment judgements also take into account a pupil's performance in national curriculum tests (often referred to as SATs) for the English reading measure. In particular, we derive measures from the *KS1_READ_OUTCOME* and *KS1_WRIT_OUTCOME* variables. As described in the National Pupil Database (NPD) tables[25], students can be awarded any of the following seven categories in the teaching assessments that take place in Year 2:

- A = Absent (or there is not enough information to provide a teaching assessment judgement)
- D = Disapplied (i.e., when a pupil has been disapplied from the national curriculum, including statutory assessment requirements)
- BLW = Below the standard of the interim pre-Key Stage- corresponds with P-scales or NOTSEN
- PKF = Pre-Key Stage foundations for the expected standard
- WTS = Working towards the expected standard
- EXS = Working at the expected standard
- GDS = Working at a greater depth within the expected standard

The two variables that we will use in the analysis as our outcome variables are constructed in the same way for both the reading and writing outcomes. In particular, we construct a dummy variable that is equal to one if the student is in categories EXS (working at the expected standard) or GDS (working at a greater depth within the expected standard); and that is zero otherwise (i.e., for the BLW, PKF, and

[24] Note that the equivalent cohort in control schools will not get the treatment at any stage. Cohort 2 is a year younger. This enables us to consider the effects of the treatment on longer-term outcomes (i.e. Key Stage 1 teacher assessments).

[25] https://www.gov.uk/government/publications/national-pupil-database-user-guide-and-supporting-information

WTS categories described above). The observations for students classified as absent (A) or disapplied (D) are set to missing.

In addition to these KS1 outcomes, Coventry University also tested students one year after the end of treatment, in July 2016, using the same Progress in Reading Assessment (PIRA, Year 2 Summer) test, which was used as the primary outcome in the original report. PIRA assesses reading ability in the following areas: phonics, literal comprehension and reading for meaning.[26] The PIRA tests were administered by a team of Research Assistants (RAs) employed by Coventry University who did not know to what condition the children had been allocated.

Participant selection, sample size and randomisation

We refer the reader to the Methods section in the original report, and in particular, to the subsections on participant selection, sample size and randomisation which describe how participating schools were selected, the details on the sample size calculations, and the randomisation procedures employed.

Analysis

Our analysis for this addendum report follows the protocol and estimates Ordinary Least Squares (OLS) regressions for each outcome variable where the outcome is regressed against dummy variables for whether the pupils are in the ICT treatment or the Non-ICT treatment.[27] The preferred specification in this analysis is the same as in the original report (where we controlled for the pairing dummies, as well as for prior attainment (as given by the PIRA baseline test), gender, FSM eligibility and a dummy variable indicating whether the pupil had a good level of development in the Foundation Stage Profile (FSP GLD)).[28] Standard errors in the regression analysis are always clustered at the school level. This takes into account the fact that students are clustered in schools, and therefore outcomes might be correlated.

The derived KS1 outcomes are dummy variables, and therefore the coefficients of the different explanatory variables should be read in terms of percentage points (the tables also present the effect size in terms of standard deviations, see the table notes for details).

The PIRA test outcome variable has been standardised to have mean zero and standard deviation (SD) one, using the mean and standard deviation of the available PIRA sample a year after the end of the treatments. In this case, this allows interpreting the coefficients of the explanatory variables in terms of standard deviations of the outcome variable.

Our analysis has been on an Intention-to-Treat basis: as explained in the original report, 51 schools were randomised into treatment and control schools. Out of these 51 schools, one was an independent school that was paired to a blank school. This independent school is not part of the analysis. From the remaining 50 schools, 25 were randomised to the treatment group and 25 schools were randomised to the control group. Five schools dropped out after randomisation. The Key Stage 1 variables, however, are available for all the 50 schools, so this allows us to carry out an intention-to-treat analysis. In the main analysis in this addendum report, we will use all the available observations for the KS1 outcomes, although we will compare the results with those obtained when we restrict the sample to the available observations for only those schools that we could include in the primary analysis from the original report. The sample for the main analysis (the one with controls and all available observations) includes 2111

[26] More information on the PIRA test can be found here: https://www.hoddereducation.co.uk/pira

[27] Marginal effects evaluated at mean values using logit regressions in the preferred specification offer very similar conclusions to those presented using OLS regressions.

[28] FSP GLD is a dummy variable that indicates whether the pupil has achieved a good level of development in the Foundation Stage Profile. This is the case if the pupil achieved a level of 2 or 3 in each of COM (Communication), PHY (Physical development), PSE (Personal, Social and Emotional Development), LIT (Language and Literacy) and MAT (Mathematical development) results.

students. This corresponds to 95% of the students included in the randomisation (or 94% if we include pupils from the independent school in the total number of students).

For the original PIRA tests, for four out of the five schools that dropped out, we were able to obtain post-test data and therefore they were included in our original short-term analysis. This is also the case with the PIRA tests carried out by Coventry University a year after the end of the intervention. The team was able to test students in 48 schools in July 2016. so the sample of schools for the analysis using the standardised PIRA outcome variables includes the same number of schools as were included in the PIRA analysis in the original report.[29] We miss some students who were included in the sample in the original report for the preferred specification, though. Specifically, we do not have PIRA test data (a year after the end of the treatments) for 99 students. The final sample for the analysis of PIRA outcomes one year after the intervention consists of 1785 students (or 80% of the students included in the randomisation).

The original report checked whether baseline characteristics were balanced for the final sample to discern whether the small amount of attrition (i.e. one treatment school) created imbalance in the different intervention and control groups. We also checked whether those pupils with missing data at follow-up (post-treatment stage) had different baseline characteristics than the pupils for which we had data at post-test. Since we did not find significant differences, we did not impute missing data and worked with the available final sample. We do the same here, since the main outcome variables (i.e. those defined on the basis of Key Stage 1 outcomes), are available for all participating schools and only missing for a very small number of students.

Following the protocol, we also explore here whether there are heterogeneous effects of treatment on the main KS1 outcome variables according to free school meal status, gender, and above median attainment on pre-test. For example, for FSM, we do this by interacting each treatment variable with a dummy variable that indicates whether the pupil is eligible to receive FSM[30]. Finally, we study the existence of *spillover* effects from treatment pupils to control pupils in treatment schools in the same way as we did in the first report.

[29] Because it was not possible to obtain post-intervention outcomes for one of the treatment schools that dropped out of the intervention, we also had to drop the relevant paired comparison group school. Hence we have 48 schools in the analysis that uses PIRA outcomes and 50 schools in the analysis that uses administrative data one year later.
[30] This variable comes from the National Pupil Database, in particular, from the pupil Census. A pupil is recorded as eligible for Free School Meals on Census day. From 2012/13, this variable set to 'true' (i.e., 1) if a pupil has an FSM period with a start date and end date blank or end date on or after census date which means they are FSM eligible on Census day.

3. Impact evaluation

Outcomes and analysis

The impact of the different interventions on Key Stage 1 reading and writing is summarised in Tables 1 and 2, respectively. The analysis uses as outcome variables the dummy variables detailed above that are derived from Key Stage 1 reading and writing outcomes. As stated above, each of these two variables is equal to one if the student is working at the expected standard or above. Following the main table in the first report (Table 9), Tables 1 and 2 have the same format and show specifications in the three different columns, in each case making use of the maximum number of observations available.[31] In all the columns the results of regressing the dummy variable of interest on the ICT and Non-ICT treatment dummies, are shown. In addition, all the specifications include dummies for the randomisation pairs. Standard errors are clustered at the school level in all regressions.

Each cell in the table includes a number of figures. The top figure shows the value of the coefficient (which should be interpreted in terms of percentage points of the outcome variable). For the main treatment variables, the next row shows the coefficient presented as an effect size in terms of standard deviations. This figure is in italics. The figure in parentheses is the p-value. The pair of figures in brackets shows 95% confidence interval for the coefficients. The control group in all regressions consists of all control pupils (i.e. control pupils in treated schools and control pupils in control schools). Results using pupils in control schools only will be shown later in the section.

The first column in Tables 1 and 2 show the results of a specification that only controls for the treatment and pairing dummies. The second column includes a control for the standardised baseline PIRA test. The third column further controls for gender, FSM and whether the pupil achieved a good level of development in the Foundation Stage Profile (FSP_GLD). This is the preferred specification.

Similarly to results reported in the main table of the original report, the results in Table 1 show that as we progress from columns (1) to (3), there is a small increase in the point estimates of the treatment effects as well as an increase in precision. However, a year after the end of treatment, only the Non-ICT intervention seems to have a positive and significant impact on the Key Stage 1 teacher assessed reading variable. In particular, in column 3 (our preferred specification) we show that students that were randomly assigned to the Non-ICT treatment are on average 6 percentage points more likely to be observed working (at least) at the expected reading standard a year after the end of the intervention. As mentioned earlier in this section, the control group in these regressions consists of all control pupils. The magnitude of the effect is sizable. This can be better seen by comparing the 6-percentage point increase to the mean of the outcome variable in control schools (where 74% of students are working (at least) at the expected reading standard a year after the end of the intervention) as shown in the penultimate cell of the table.

The initial significant positive effect found in the first report for the ICT treatment disappears when using the Key Stage 1 reading outcome a year after the end of treatment. The sign of the ICT coefficient is still positive, but the magnitude is small and it is not precisely estimated (i.e., the standard errors are too big to be able to say with confidence that the estimated coefficient is actually different from zero). The final row in Table 1 shows the p-value from testing the null hypothesis that the ICT and Non-ICT coefficients are equal. The p-value in the first column suggests that we can marginally reject the null hypothesis that the effects of the two different treatments are equal (at the 10% level) –although the p-value increases slightly in columns 2 and 3.

[31] The results are very similar if we restrict the samples in all columns to have the same number of observations given by the sample with controls in column 3.

We repeat the analysis shown in Table 1 using the sample of 48 schools for which data was available for the outcome measure used in the first report. We do this in order to see whether having a slightly bigger sample in Table 1 than the one in the original makes any difference to our results. The results are similar to the ones reported in Table 1 and can be found in Appendix Table 1 of this Addendum report.

Table 1: Intention to treat analysis for the KS1 reading outcome. All available observations

Outcome: KS1_read_outcome	(1) Pairs	(2) + baseline test	(3) +FEM FSM & FSP_GLD
ICT	-0.006 *-0.013* (0.857) [-0.072,0.060]	0.019 *0.043* (0.456) [-0.032,0.070]	0.024 *0.055* (0.329) [-0.025,0.073]
NON-ICT	0.053* *0.120** (0.074) [-0.005,0.112]	0.058** *0.131*** (0.015) [0.012,0.104]	0.061** *0.137*** (0.013) [0.014,0.108]
Std baseline PIRA test		0.224*** (0.000) [0.197,0.250]	0.162*** (0.000) [0.135,0.189]
Female			-0.036* (0.072) [-0.075,0.003]
FSMeligible_SPR14			-0.087*** (0.001) [-0.136,-0.038]
FSP_GLD			0.222*** (0.000) [0.171,0.274]
Constant	0.736*** (0.000) [0.591,0.881]	0.757*** (0.000) [0.672,0.841]	0.677*** (0.000) [0.588,0.767]
Mean outcome in control schools (%) Students Schools	73.04 2188 50	73.93 2129 50	74.06 2111 50
Ho: ICT-NONICT=0 (p-value shown)	0.105	0.161	0.148

Note: Standard errors clustered at the school level. P-values in parentheses (p < 0.10; ** p < 0.05; *** p < 0.01), and 95% confidence intervals shown in brackets. ICT and non-ICT are the intention to treat dummies. All regressions include pairing dummies (schools were paired for randomisation). PIRA at baseline is standardised to have mean 0 and SD 1. The outcome variable takes value 1 whenever the pupil is working at or above the expected level for the Key Stage 1 Read variable (ks1_read_outcome). For the main treatment variables (ICT and NON-ICT), the first figure in each cell is interpreted in terms of percentage points; whereas the second in terms of standard deviations.*

A similar pattern to that observed for the KS1 reading outcomes emerges in Table 2, when we use the dummy variable derived from the Key Stage 1 writing outcome. As we progress from columns (1) to (3), there is a small increase in the point estimates of the treatment effects and a marked increase in precision. In our preferred specification in column 3, the ICT treatment is significant at the 10% level. Although not very precisely estimated, the coefficient for the ICT treatment variable suggests that students in the ICT treatment still benefit from the treatment a year after the end of the intervention in terms of improved writing skills. In particular, students in the ICT treatment are almost 5 percentage points more likely to be observed working (at least) at the expected writing standard a year after the end of the intervention. This effect is almost doubled, however, for students that received the Non-ICT treatment (almost a 9-percentage point increase on average) (As mentioned earlier in this section, the control group in these regressions consists of all control pupils). In addition, the estimates associated with the two interventions are significantly different from each other for the three specifications (see the last row of Table 2).

We present a similar table in Appendix Table 2 using the same specifications as for Table 2 but using the sample of 48 schools available in the first report. In this analysis, the estimates of the Non-ICT treatment are similar for the preferred specification including controls; but the ICT coefficient is no longer significant in this reduced sample. However, given a higher level of attrition in the sample used in the first report, our preferred sample and specification is the one presented in column 3 of Tables 1 and 2 (given by all available KS1 observations); and we will continue to use it throughout the analysis in this section.

Table 2: Intention to treat analysis for the KS1 writing outcome. All available observations

Outcome: KS1_write_outcome	(1) Pairs	(2) + baseline test	(3) +FEM FSM & FSP_GLD
ICT	0.014 *0.028* (0.715) [-0.061,0.088]	0.041 *0.085* (0.157) [-0.016,0.099]	0.048* *0.099** (0.096) [-0.009,0.104]
NONICT	0.075** *0.155*** (0.035) [0.006,0.145]	0.082*** *0.169**** (0.006) [0.025,0.139]	0.088*** *0.182**** (0.004) [0.029,0.147]
Std baseline PIRA test		0.260*** (0.000) [0.237,0.284]	0.181*** (0.000) [0.153,0.208]
Female			0.002 (0.908) [-0.036,0.040]
FSMeligible_SPR14			-0.081*** (0.003) [-0.131,-0.030]
FSP_GLD			0.281*** (0.000) [0.211,0.351]
Constant	0.650*** (0.000) [0.429,0.870]	0.670*** (0.000) [0.523,0.817]	0.537*** (0.000) [0.389,0.686]
Mean outcome in control schools Students Schools	61.59 2188 50	62.4 2129 50	62.51 2111 50
Ho: ICT-NONICT=0 (p-value shown)	0.076	0.054	0.051

Note: Standard errors clustered at the school level. P-values in parentheses (p < 0.10; ** p < 0.05; *** p < 0.01), and 95% confidence intervals shown in brackets. ICT and non-ICT are the intention to treat dummies. All regressions include pairing dummies (schools were paired for randomisation). PIRA at baseline is standardised to have mean 0 and SD 1. The outcome variable takes value 1 whenever the pupil is working at or above the expected level for the Key Stage 1 Write variable (ks1_writ_outcome). For the main treatment variables (ICT and NON-ICT), the first figure in each cell is interpreted in terms of percentage points; whereas the second in terms of standard deviations.*

Additional analysis

Coventry University collected PIRA test data in July 2016 –a year after the end of the intervention. This is the same type of test that was used in the analysis presented in the original report, repeated one year later, and using the version of test appropriate for children a year older (Year 2 Summer). Although analysis using this July 2016 PIRA data was not specified in the protocol, it is of interest because it allows direct comparison of impact immediately after the intervention and one year on. The estimated effects are too imprecise to provide information on the effectiveness of the interventions on this outcome measure.

More detail is provided in Appendix Table 3.

Sub-group analysis

Following the format of the first report and the protocol, Tables 3 and 4 explore whether there are heterogeneous effects of treatment on the main KS1 outcome variables according to free school meal status, gender and whether the pupil has a score above median attainment on the pre-test. For example,

when we consider whether effects are heterogeneous by FSM status, we interact each treatment variable with a dummy variable that indicates whether the pupil is eligible to receive FSM. All regressions control for the set of controls included in the preferred specification (see Analysis section) but only the sub group coefficients are shown in Tables 3 and 4). To ease comparability, the first column reproduces the main results (i.e. column 3 of Tables 1 and 2, respectively, for the coefficients of interest).

The presence of heterogeneous effects has also been explored by running separate regressions for each condition (i.e. FSM and NON-FSM pupils; males and females; and above and below median baseline attainment). The results remain virtually unchanged.

The results in the second column of Table 3 for the KS1 reading outcome are in line with those found in the first report for the short-term impacts of the interventions. A year after treatment we find that the impact of the intervention is still bigger for disadvantaged pupils (as measured by the FSM eligibility indicator) for both treatments (although only the Non-ICT coefficient is significant). Compared to all students in the control group, FSM students allocated to the Non-ICT treatment are 11.5 percentage points more likely to be observed working at the expected reading standard (or higher) a year after the end of the intervention. The coefficients for the other subgroups are not statistically significant, although the point estimates are consistent with the possibility that non-FSM students in the Non-ICT treatment also do better in reading a year after the end of the treatment.

The results in Column 3 suggest that the Non-ICT treatment is more beneficial to females than to males for this reading outcome. Finally, the results in Column 4 show that it is the students above median ability (for both the ICT and the Non-ICT treatment) for whom effects are evident one year after the intervention.[32] This is different from findings in the first report where positive effects were found for every sub-group apart from those of above-median ability exposed to the ICT intervention. In Appendix Table 4 we show (using a triple interaction) that the higher results for above median students are driven by FSM students that had above median attainment in the pre-test.

Table 4 shows the results of a similar analysis when using the dummy variable obtained from the writing Key Stage 1 variable. Dividing the groups in terms of FSM status, all subgroups show a significant impact of the interventions on writing skills (except for students originally assigned to the Non-FSM ICT group). This is also true for gender: interventions have a positive and significant impact on all subgroups apart from males exposed to the ICT intervention. Finally, the results are very similar to the KS1 reading variable when it comes to differential impacts by prior attainment at the PIRA baseline test. Again, above median FSM students are driving these results (see Column 2 in Appendix Table 4).

[32] 'Abiity' refers to reading (PIRA) test scores measured at baseline, before the intervention was implemented.

Table 3: Analysis of heterogeneous effects of the interventions on the KS1 reading outcome

	(1)	(2)	(3)		(4)	
Outcome: KS1_read_outcome	Final specif	FSM interaction	Outcome: KS1_read_outcome	Gender interaction	Outcome: KS1_read_outcome	Pre-test interaction
ICT	0.024 *0.055* (0.329)					
NONICT	0.061** *0.137*** (0.013)					
ICT_FSM		0.061 *0.139* (0.302)	ICT*Female	0.025 *0.055* (0.355)	ICT*(> median)	0.055** *0.125*** (0.012)
ICT_NOFSM		0.014 *0.032* (0.564)	ICT*Male	0.024 *0.055* (0.510)	ICT*(< median)	-0.004 *-0.010* (0.915)
NONICT_FSM		0.115*** *0.260**** (0.003)	NONICT*Female	0.095*** *0.215**** (0.005)	NONICT*(> median)	0.072** *0.162*** (0.011)
NONICT_NOFSM		0.045 *0.101* (0.129)	NONICT*Male	0.023 *0.053* (0.517)	NONICT*(< median)	0.048 *0.108* (0.257)
Students Schools	2111 50	2111 50	Students Schools	2111 50	Students Schools	2111 50
Ho: ICT FSM - ICT NOFSM=0		0.543	Ho: ICT Female - ICT Male=0	0.998	Ho: ICT Above - ICT Below=0	0.18
Ho NONICT FSM - NONICT NOFSM=0		0.003	Ho NONICT Female - NONICT Male=0	0.165	Ho NONICT Above - NONICT Below=0	0.649

Note: Standard errors clustered at the school level. P-values in parentheses (p < 0.10; ** p < 0.05; *** p < 0.01). ICT and non-ICT are the intention to treat dummies. All regressions include pairing dummies (schools were paired for randomisation) and all other controls. PIRA at baseline is standardised to have mean 0 and SD 1. The outcome variable takes value 1 whenever the pupil is working at or above the expected level for the Key Stage 1 Read variable (ks1_read_outcome). Sample with all observations available. For the main treatment variables (ICT and NON-ICT, for each of the subgroups), the first figure in each cell is interpreted in terms of percentage points; whereas the second in terms of standard deviations.*

Table 4: Analysis of heterogeneous effects of the interventions on the KS1 writing outcome

	(1)	(2)	(3)		(4)	
Outcome: KS1_write_outcome	Final specif	FSM interaction	Outcome: KS1_write_outcome	Gender interaction	Outcome: KS1_write_outcome	Pre-test interaction
ICT	0.048* *0.099** (0.096)					
NONICT	0.088*** *0.182**** (0.004)					
ICT_FSM		0.111** *0.229*** (0.044)	ICT*Female	0.081*** *0.167**** (0.006)	ICT*(> median)	0.096*** *0.198**** (0.001)
ICT_NOFSM		0.031 *0.063* (0.311)	ICT*Male	0.012 *0.025* (0.758)	ICT*(< median)	0.004 *0.009* (0.918)
NONICT_FSM		0.160** *0.331*** (0.018)	NONICT*Female	0.114*** *0.235**** (0.002)	NONICT*(> median)	0.098*** *0.203**** (0.001)
NONICT_NOFSM		0.067** *0.139*** (0.025)	NONICT*Male	0.061* *0.127** (0.087)	NONICT*(< median)	0.075 *0.156* (0.110)
Students Schools	2111 50	2111 50	Students Schools	2111 50	Students Schools	2111 50
Ho: ICT FSM - ICT NOFSM=0		0.110	Ho: ICT Female - ICT Male=0	0.067	Ho: ICT Above - ICT Below=0	0.032
Ho NONICT FSM - NONICT NOFSM=0		0.015	Ho NONICT Female - NONICT Male=0	0.19	Ho NONICT Above - NONICT Below=0	0.629

Note: Standard errors clustered at the school level. P-values in parentheses (p < 0.10; ** p < 0.05; *** p < 0.01). ICT and non-ICT are the intention to treat dummies. All regressions include pairing dummies (schools were paired for randomisation) and all other controls. PIRA at baseline is standardised to have mean 0 and SD 1. The outcome variable takes value 1 whenever the pupil is working at or above the expected level for the Key Stage 1 Write variable (ks1_write_outcome). Sample with all observations available. For the main treatment variables (ICT and NON-ICT, for each of the subgroups), the first figure in each cell is interpreted in terms of percentage points; whereas the second in terms of standard deviations.*

The last part of the analysis deals with the fact that there could be *spillover* effects from treatment pupils to control pupils in treatment schools. Tables 5 and 6 reproduce the same structure that was used in the first report to address this issue, for the dummies derived from the KS1 reading and writing outcomes, respectively. As explained in the first report, the protocol specified that we would run a regression where the control group would be formed by pupils in the same schools, and a second regression where the control group would be composed by pupils in control schools only. We could then check whether the treatment effects were larger in the second regression than in the first one. This would be indicative of *spillover* effects. Although we do this, we also run a regression where we simply include an additional dummy variable for pupils that are in the control group of treatment schools (denoted by CT). The advantage of this approach is that it produces similar insights without having to reduce the number of clusters (which is necessary when running regressions for treatment schools only: 25 schools).

Both Tables 5 and 6 present a similar picture. Even though we did find indications of *spillover* effects from treated to control students in treatment schools immediately after the end of the intervention, these effects do not seem to persist one year later. The results when using either of the KS1 derived dummies

as the outcome variable show that the coefficients on the CT dummy are close to zero in both cases. The regression results are in line with those already discussed for Tables 1 and 2 in all specifications regardless of the control group used.

Table 5: Spillover effects using the KS1 reading outcome

Outcome: KS1_read_outcome	(1) Final specif	(2) +Dummy for Control students in treated schools	(3) Control group: control students in treated schools	(4) Control group: Control students in control schools
ICT	0.024 *0.055* (0.329)	0.024 *0.054* (0.382)	0.022 *0.050* (0.397)	0.024 *0.055* (0.375)
NONICT	0.061** *0.137*** (0.013)	0.060** *0.136*** (0.030)	0.058** *0.131*** (0.015)	0.061** *0.138*** (0.033)
CT		-0.001 *-0.002* (0.965)		
Std baseline PIRA test	0.162*** (0.000)	0.162*** (0.000)	0.162*** (0.000)	0.163*** (0.000)
female	-0.036* (0.072)	-0.036* (0.072)	-0.022 (0.331)	-0.038* (0.082)
FSMeligible_SPR14	-0.087*** (0.001)	-0.087*** (0.001)	-0.053 (0.100)	-0.082*** (0.004)
FSP_GLD	0.222*** (0.000)	0.222*** (0.000)	0.236*** (0.000)	0.224*** (0.000)
Constant	0.677*** (0.000)	0.678*** (0.000)	0.631*** (0.000)	0.687*** (0.000)
Students Schools	2111 50	2111 50	1020 25	1756 50
P-values: Ho: ICT-NONICT=0 Ho: ICT-CT=0 Ho: NONICT-CT=0	0.141	0.141 0.337 0.007	0.175	0.142

Note: Standard errors clustered at the school level. P-values in parentheses (p < 0.10; ** p < 0.05; *** p < 0.01). ICT and non-ICT are the intention to treat dummies. CT is a dummy that equals 1 for pupils in the control group of treatment schools. All regressions include pairing dummies (schools were paired for randomisation). PIRA at baseline is standardised to have mean 0 and SD 1. The outcome variable takes value 1 whenever the pupil is working at or above the expected level for the Key Stage 1 Read variable (ks1_read_outcome). Sample with all observations available. For the main treatment variables (ICT and NON-ICT), the first figure in each cell is interpreted in terms of percentage points; whereas the second in terms of standard deviations.*

Table 6: Spillover effects using the KS1 writing outcome

Outcome: KS1_write_outcome	(1) Final specif	(2) +Dummy for control students in treated schools	(3) Control group: control students in treated schools	(4) Control group: Control students in control schools
ICT	0.048* *0.099** (0.096)	0.051 *0.105* (0.157)	0.033 *0.067* (0.127)	0.052 *0.108* (0.156)
NONICT	0.088*** *0.182**** (0.004)	0.091** *0.188*** (0.017)	0.072*** *0.150**** (0.002)	0.092** *0.191*** (0.019)
CT		0.012 *0.025* (0.747)		
Std baseline PIRA test	0.181*** (0.000)	0.181*** (0.000)	0.172*** (0.000)	0.188*** (0.000)
female	0.002 (0.908)	0.002 (0.896)	0.024 (0.312)	0.002 (0.915)
FSMeligible_SPR14	-0.081*** (0.003)	-0.081*** (0.003)	-0.064** (0.035)	-0.054* (0.063)
FSP_GLD	0.281*** (0.000)	0.282*** (0.000)	0.297*** (0.000)	0.278*** (0.000)
Constant	0.537*** (0.000)	0.534*** (0.000)	0.520*** (0.000)	0.524*** (0.000)
Students Schools	2111 50	2111 50	1020 25	1756 50
P-values: Ho: ICT-NONICT=0 Ho: ICT-CT=0 Ho: NONICT-CT=0	0.051	0.051 0.078 0.001	0.072	0.050

Note: Standard errors clustered at the school level. P-values in parentheses (p < 0.10; ** p < 0.05; *** p < 0.01). ICT and non-ICT are the intention to treat dummies. CT is a dummy that equals 1 for pupils in the control group of treatment schools. All regressions include pairing dummies (schools were paired for randomisation). PIRA at baseline is standardised to have mean 0 and SD 1. The outcome variable takes value 1 whenever the pupil is working at or above the expected level for the Key Stage 1 Write variable (ks1_write_outcome). Sample with all observations available. For the main treatment variables (ICT and NON-ICT), the first figure in each cell is interpreted in terms of percentage points whereas the second in terms of standard deviations.*

4. Conclusion

The objective of the addendum report is to assess whether effects of the interventions persist one year after the intervention ended. In line with the protocol we focus on teacher assessments of reading and writing at the end of Key Stage 1. A full discussion of the intervention and the evaluation is provided in the main report and we do not repeat this here.

Our results show that effects of the interventions do persist one year after they have ended. This is especially evident for the Non-ICT intervention. Specifically, pupils originally allocated to the Non-ICT intervention are more likely to be observed working at the expected level (or higher) in both reading and writing. The magnitude is 6 and 9 percentage points for reading and writing respectively. For the ICT intervention, the point estimate is very small with regard to reading, although the level of imprecision around the estimate implies that we are unable to rule out positive or negative effects. For writing, the average effect is positive; pupils exposed to this intervention are more likely to achieve at least the expected standard by 5 percentage points relative to the control group.

The subgroup analysis shows that pupils who benefited from the intervention were more likely to be in the following categories: FSM, female, above-median ability (as measured by the baseline test). Furthermore, within 'above median ability', FSM students drive the effect (even though this is a numerically small number of students). Finally, the regression results do not suggest that the spillover effect (originally evident for control pupils in treatment schools) persisted one year on.

A strength of this analysis is that the outcome measures are collected in administrative data and we are able to include almost all pupils (95% of those included in the randomisation stage) and schools, thus ruling out that any of the detected impacts are influenced by attrition. A limitation is that the outcomes are teacher assessed. However, given that in most schools the teacher will have changed between the year of the intervention and the year of assessment, we think it unlikely that assessments will be biased by the original treatment status of the schools. An even stronger test of whether effects persist will be whether students exposed to interventions perform better at the end of primary school in Key Stage 2 tests. However, to evaluate this, we will need to wait for another 4 years.

Future research and publications

Future research will analyse additional data collected by the implementation team in terms of pupil group composition (number of pupils, average ability of pupils, etc.), TA surveys, TA logs, etc. We will incorporate this in the analysis with the objective of understanding the higher impact of the Non-ICT intervention, for both the short-term and the impacts a year after treatment. This analysis will be published in the discussion paper series of the Centre for Economic Performance (there will be no additional EEF report). Moreover, we expect to publish academic paper(s) in collaboration with the team at Coventry University.

5. References

Abrami, P.C., Savage, R.S., Deleveaux, G., Wade, A., Meyer, E. & Lebel, C. (2010). The Learning Toolkit: The design, development, testing and dissemination of evidence-based educational software. In P. Zemliansky & D.M. Wilcox (Eds.), Design and implementation of educational games: Theoretical and practical perspectives (pp. 168-187). Hershey, PA: IGI Global. http://dx.doi.org/10.4018/978-1-61520-781-7.ch012

Abrami, P. C., Wade, A., Lysenko, L. V., Marsh, J., & Gioko, A. (2014). Using educational technology to develop early literacy skills in Sub-Saharan Africa. *Education and Information Technologies*, Advance online publication. http://dx.doi.org/10.1007/s10639-014-9362-4

Cheung, A., Mak, B.S.Y., Abrami, P.C., Wade, A. & Lysenko, L**.** (2016). The effectiveness of the ABRACADABRA (ABRA) web-based literacy program on primary school students in Hong Kong. *Journal of Interactive Learning Research. 27* (3), 219- 245

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 children', Journal of Research in Reading, 32, pp. 92–108.

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', Journal of 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', Journal of Educational Psychology, 105(2), p. 310

Wolgemuth, J. R., Savage, R. S., Helmer, J., Lea, T., Harper, H., Chalkiti, K., Bottrell, C. and Abrami, P. (2011) 'Using computer-based instruction to improve indigenous early literacy in Northern Australia: A quasi-experimental study', Australasian Journal of Educational Technology, 27, pp .727–750.

Wolgemuth, J., Savage, R., Helmer, J., Harper, H., Lea, T, Abrami, P. C., . . . Louden, W. (2013). ABRACADABRA aids indigenous and non-indigenous early literacy in Australia: Evidence from a multisite randomized controlled trial. Computers & Education, 67, 250-264. **http://dx.doi.org/10.1016/j.compedu.2013.04.002**

Appendix

Appendix Table 1: Intention to treat analysis for the KS1 reading outcome. All available observations from the sample in the original report

Outcome: KS1_read_outcome	(1) Pairs	(2) + baseline test	(3) +FEM FSM & FSP_GLD
ICT	-0.008 (0.789)	0.010 (0.676)	0.021 (0.374)
NONICT	0.057 (0.106)	0.066** (0.022)	0.076*** (0.007)
Std baseline PIRA test		0.223*** (0.000)	0.158*** (0.000)
Female			-0.038* (0.074)
FSMeligible_SPR14			-0.078*** (0.006)
FSP_GLD			0.235*** (0.000)
Constant	0.579*** (0.000)	0.659*** (0.000)	0.585*** (0.000)
Mean outcome in control schools (%) Students Schools	 73.18 1868 48	 73.18 1868 48	 73.18 1868 48

Note: Standard errors clustered at the school level. P-values in parentheses (p < 0.10; ** p < 0.05; *** p < 0.01). ICT and non-ICT are the intention to treat dummies. All regressions include pairing dummies (schools were paired for randomisation). PIRA at baseline is standardised to have mean 0 and SD 1. The outcome variable takes value 1 whenever the pupil is working at or above the expected level for the Key Stage 1 Read variable (ks1_read_outcome). This sample uses all the available observations from the PIRA sample in the original report.*

Appendix Table 2: Intention to treat analysis for the KS1 writing outcome. All available observations from the sample in the original report

Outcome: KS1_write_outcome	(1) Pairs	(2) + baseline test	(3) +FEM FSM & FSP_GLD
ICT	0.006 (0.866)	0.027 (0.348)	0.040 (0.173)
NONICT	0.058 (0.158)	0.068** (0.038)	0.080** (0.021)
Std baseline PIRA test		0.261*** (0.000)	0.177*** (0.000)
Female			0.008 (0.685)
FSMeligible_SPR14			-0.063** (0.037)
FSP_GLD			0.294*** (0.000)
Constant	0.394*** (0.000)	0.488*** (0.000)	0.348*** (0.000)
Mean outcome in control schools (%)	61.52	61.52	61.52
Students	1868	1868	1868
Schools	48	48	48

Note: Standard errors clustered at the school level. P-values in parentheses (p < 0.10; ** p < 0.05; *** p < 0.01). ICT and non-ICT are the intention to treat dummies. All regressions include pairing dummies (schools were paired for randomisation). PIRA at baseline is standardised to have mean 0 and SD 1. The outcome variable takes value 1 whenever the pupil is working at or above the expected level for the Key Stage 1 Write variable (ks1_write_outcome). This sample uses all the available observations from the PIRA sample in the original report.*

Appendix Table 3: Intention to treat and instrumental variable analysis for the PIRA test 1 year after treatment. All available observations from the sample in the original report

Outcome: Standardised measure of PIRA test one year after treatment	(1) Pairs	(2) + baseline test	(3) +FEM FSM and FSP_GLD	(4) IV regressions (2nd stage)
ICT	0.025 (0.677)	0.058 (0.336)	0.073 (0.229)	0.087 (0.212)
NONICT	0.053 (0.514)	0.077 (0.281)	0.090 (0.183)	0.108 (0.157)
Std baseline PIRA test		0.564*** (0.000)	0.448*** (0.000)	0.448*** (0.000)
Female			0.023 (0.627)	0.022 (0.64)
FSMeligible_SPR14			-0.189*** (0.000)	-0.199*** (0.000)
FSP_GLD			0.392*** (0.000)	0.396*** (0.000)
Constant	-0.257*** (0.000)	-0.042 (0.577)	-0.209*** (0.000)	-0.214*** (0.000)
Students Schools	1799 48	1799 48	1785 48	1785 48

Note: Standard errors clustered at the school level. P-values in parentheses (p < 0.10; ** p < 0.05; *** p < 0.01). ICT and non-ICT are the intention to treat dummies. All regressions include pairing dummies (schools were paired for randomisation). Both the outcome variable and PIRA at baseline are standardised to have mean 0 and SD 1, so the coefficients are interpreted in terms of SD of the outcome variable.*

The format of the Appendix Table 3 is very similar to Tables 1 and 2 in the addendum report. Columns 1 to 3 show the results of analogous specifications. The results in our preferred specification in Column 3 show positive impacts of both the ICT and Non-ICT treatments, but the standard errors are too big to make claims about the effectiveness of the interventions one year after they have finished.

We also try to explore local average treatment effects here by using instrumental variable methods, since four of the treated schools in this sample dropped out at some point after randomisation. We instrument the actual receipt of treatment by the initial allocation at the randomisation stage. As expected, initial assignment to treatment status strongly predicts whether the school remained in the treatment (not shown here). The instrumental variable results shown for the second stage in column 4 are in line with the Intention to Treat impacts shown in Column 3, although the estimates are slightly more precise, particularly for the Non-ICT outcome.[33] Although we have to take these results with

[33] The intention to treat regressions classify schools in the groups given at the randomisation stage. That is, even if a treatment (control) school drops out from the analysis, it is considered as a treatment (control) school in the regressions. This is possible because data was collected for all schools, even for those dropping out (with the exception of one of the schools). The schools are kept with the initial assignment because that initial assignment is random. Because the small number of drop-out schools did not receive the intervention, the 'intention to treat' effect might underestimate the average effect of actually receiving the treatment. The instrumental variable estimates shown in Appendix Table 3 (column 4), deal with this issue. They give an unbiased estimate of the effect of actually receiving the intervention.

caution, the point estimates suggest an impact of the ICT and non-ICT interventions of 8.7% and 10.8% of a SD when using the standardised PIRA test outcomes. However, the large p-values (particularly for the ICT intervention) would not allow us to exclude the possibility that the effect of the interventions did not last at all for this outcome measure. We do not report further analysis using this outcome measure because for the most part estimated effects are too imprecise to provide information on the effectiveness of the interventions on this outcome measure.

Appendix Table 4: Analysis of heterogeneous effects (triple interactions) of the interventions for the KS1 reading and writing outcomes

	(1) KS1_read_outcome	(2) KS1_write_outcome
ICT*ABOVE*FSM	0.211*** (0.000)	0.264*** (0.002)
ICT*ABOVE*NOFSM	0.032 (0.163)	0.070** (0.016)
ICT*BELOW*FSM	0.003 (0.972)	0.050 (0.428)
ICT*BELOW*NOFSM	-0.006 (0.895)	-0.013 (0.796)
NONICT*ABOVE*FSM	0.188*** (0.000)	0.222*** (0.000)
NONICT*ABOVE*NOFSM	0.049* (0.086)	0.073** (0.018)
NONICT*BELOW*FSM	0.071 (0.183)	0.121 (0.176)
NONICT*BELOW*NOFSM	0.040 (0.457)	0.058 (0.224)
Students Schools	2111 50	2111 50

Note: Standard errors clustered at the school level. P-values in parentheses (p < 0.10; ** p < 0.05; *** p < 0.01). All regressions include pairing dummies (schools were paired for randomisation) and other controls. The outcome variable takes value 1 whenever the pupil is working at or above the expected level for the Key Stage 1 Read/Write variable (ks1_read_outcome/ks1_write_outcome). Sample with all observations available.*

Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk