

CEP Discussion Paper No 1562

August 2018

**Teaching Assistants, Computers and Classroom
Management: Evidence from a Randomised Control Trial**

**Helen Johnson, Sandra McNally, Heather Rolfe
Jenifer Ruiz-Valenzuela, Robert Savage,
Janet Vousden, Clare Wood**

Abstract

Many students still leave school without a good grasp of basic literacy, despite the negative implications for future educational and labour market outcomes. We evaluate a programme that involves changing how resources are used within classrooms to reinforce the teaching of literacy. Specifically, the programme involves training teaching assistants to deliver a tightly structured package of materials to groups of young children. Further, we compare the effectiveness of computer-aided instruction using available software with the paper equivalent. We implement the experiment in the context of a Randomised Control Trial in English schools. Both interventions have a short-term impact on children's reading scores, although the effect is bigger for the paper intervention and more enduring in the subsequent year. This paper shows how teaching assistants can be used to better effect within schools, and at a low cost.

Key words: literacy, ICT, teaching assistants
JEL: I21

This paper was produced as part of the Centre's Education & Skills Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We would like to thank the Education Endowment Foundation (EEF) for commissioning this research, which was implemented by the research team at Coventry University (involving Johnson, Vousden, Savage, Wood and their research teams) and independently evaluated by the team at Centre for Economic Performance, LSE (involving Ruiz-Valenzuela and McNally) and the National Institute for Economic and Social Research (Rolfe). We thank the team of research assistants at Coventry University that facilitated the fieldwork, the technical support provided by Annie Wade and Philip Abrami, and others at Concordia's [Centre for the Study of Learning and Performance](#) (CSLP). We also thank participating schools and teaching assistants. We thank Robert Slavin for evaluating the training materials and Morag Stuart for implementing the Non-ICT version of the program evaluated here. We thank participants at seminars at Trondheim University, the Centre for Economic Performance (London School of Economics); and at the CESifo Area Conference on Economics of Education, the Royal Economic Society conference, and the II Workshop on Empirical Research in Economics of Education (Universitat Rovira i Virgili).

Helen Johnson, Coventry University. Sandra McNally, University of Surrey and Centre for Economic Performance, London School of Economics. Heather Rolfe, National Institute for Social and Economic Research. Jenifer Ruiz-Valenzuela, Centre for Economic Performance, London School of Economics. Robert Savage, University College London. Janet Vousden, Coventry University. Clare Wood, Nottingham Trent University.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means without the prior permission in writing of the publisher nor be issued to the public or circulated in any form other than that in which it is published.

Requests for permission to reproduce any article or part of the Working Paper should be sent to the editor at the above address.

1. Introduction

A significant number of children leave primary school with low levels of literacy. Despite much effort to improve basic skills in England, about 11% of children still leave primary school without having achieved the ‘expected level’ set out in the National Curriculum. This is a long-standing problem in England as it is in many other developed countries. According to an international OECD study, about a fifth of adults in England have low levels of literacy and the problem has not improved amongst young adults compared to older generations (unlike most other countries).¹ The potential implications include lower subsequent educational performance and poor labour market outcomes (e.g, see Vignoles 2016).

Gaps in cognitive ability emerge early and intervening early allows children to more effectively build later skills (Cunha et al. 2006). Such arguments are often used to argue for greater investment in early years’ education. However, what matters is not only how much is invested but what it is used for. There is a large body of evidence showing that teacher quality matters and a small but growing literature showing how interventions can boost teachers’ skills (e.g. Taylor and Tyler, 2012) or a combination of a change in pedagogy and teacher training (e.g. Machin et al. 2018).² However, there are plenty of examples where intervention does not work even if it is provided early, such as many contexts involving the use of ICT-aided instruction (e.g. see review by Bulman and Fairlie, 2016). Less is known about the effect of teaching assistants on student outcomes, even though they are widely used within English schools. Furthermore, reading has proved less malleable than maths in the face of popular policy interventions such as charter schools and test-based accountability (as argued by Jacob, 2017).

¹ OECD PIAAC study, analysed by Kuczera et al. (2016).

² Examples of studies showing the importance of teacher quality include Aaronson et al., 2007; Araujo et al., 2016; Chetty et al., 2014a, 2014b; Hanushek et al., 2005.

In this paper, we evaluate the effects of two methods of small group tuition for 5 year-old pupils in English schools, both of which are delivered by teaching assistants. The interventions have been developed by a team of UK educational psychologists as a balanced, structured reading program that contains a systematic phonics aspect, in line with recommendations in the UK and other English speaking countries. It can be delivered in the context of an ICT program (*ABRACADABRA* or *ABRA*), which is widely used in Canada and North America (Abrami et al., 2010), or in more traditional paper form (Non-ICT).³ There are reasons for thinking that the pedagogy might be effective *a priori*. First, it is based on four decades of scientific psychological theory and evidence from a series of meta-analyses of ‘what works’ in literacy. Second, there is some previous evaluation support based on small-scale studies (see Section 2). This study is not only larger scale and the first one to evaluate the interventions by a team of independent researchers, but also compares the delivery of the intervention via the ICT program to a paper (Non-ICT) version of the same intervention. In either case, the core part of this intervention is the training of teaching assistants who are already employed by the school and then the implementation of the small group teaching (which takes place outside of core literacy classes). Specifically, pupils are put together in small groups (3 to 5 pupils) and receive 15 minutes of teaching four times per week over 20 weeks. In a broader context, we can think of this intervention as measuring the effectiveness of redeploying resources within a school rather than the provision of new resources. Furthermore, what is being manipulated is how teaching assistants are being used within a school, holding teacher quality constant. Teaching Assistants (TAs) are used in most English schools and they often do not have high-level qualifications. Studies about their effectiveness are mostly correlational.⁴ This study shows how Teaching Assistants might be used within schools to

³ More specifically, *ABRA* provides a balanced suite of online activities (alphabetics, fluency, comprehension, and writing) to support reading that can be tailored for context specific purposes.

⁴ The Education Endowment Foundation have an evidence summary about TAs. One of the references to how they may be effectively deployed refers to this study, which they commissioned.

good effect. It also contributes to the literature that gets inside the ‘black box’ of what is happening inside the classroom. We learn about how the training and use of TAs can be used to affect the educational outcomes of young people.

The study is conducted as a Randomised Control Trial. Schools are randomly assigned to receive the treatment. Within treated schools, pupils are randomly assigned amongst three conditions: ICT program (*ABRA*); Non-ICT program (paper equivalent of *ABRA*) and a control group. Within treatment schools, teaching assistants are also randomly assigned to receive training in the ICT and Non-ICT condition and therefore to teach students in one or other group within their school. This design enables us to distinguish between the effects of the underlying pedagogy (common to both) and the effects of the mode of intervention (technology or paper-based). It also enables us to observe whether spillovers occur within treated schools by comparing results with different control groups (i.e. pupils not receiving the treatment in treated schools; pupils not receiving the treatment because they are in control schools). We consider the effects of the intervention at the end of the school year in which it was implemented and also one year later.

Our results show a large initial effect of the program, which is higher for the Non-ICT intervention (0.18σ and 0.27σ for the ICT and Non-ICT interventions respectively). The spillover effect within schools is substantial and only a little lower than the effect of the ICT intervention. As TAs are with classes at other times of the school day, the most plausible reason for the spillover effect is that the TA is better able to do his/her job generally, thus affecting all students. However, one year later the spillover effect disappears and there is substantial fade-out of effects for pupils assigned to either the ICT or non-ICT intervention. The point estimates suggests an effect of about one-third of the initial effect (in either case). However, the

<https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/teaching-assistants/>

experiment is under-powered to detect small effects on the main reading test. On the other hand, there is a significant effect for the non-ICT treatment if one considers administrative measures of performance the following year.⁵ Pupils assigned to the non-ICT treatment are more likely to achieve the ‘expected level’ in reading by 6 percentage points (which may be compared to a mean of 74 percent in the control group). There are also effects for writing and a smaller effect for maths one year after the end of the intervention. Given the low cost of the intervention, effects of this magnitude are likely to be cost-effective.

The rest of the paper is structured as follows. In Section 2, we give a brief overview of relevant literature. In Section 3, we describe the intervention in detail and in Section 4 we explain the methodology. In Section 5, we present the results. We discuss potential mechanisms in Section 6 before concluding in Section 7.

2. Literature: a brief overview

There have been efforts in many different countries to change approaches to teaching literacy, both for the benefit of children generally as well as those who have initial reading difficulties. Slavin et al. (2011) reviews developments over the last 25 years in research, policy and practice relating to programs for elementary-aged children who are struggling to learn to read. For example, ‘Reading Recovery’, developed in New Zealand in the 1970s is one of the best-known and well-researched programmes, and has been disseminated throughout the English speaking world. This involves individualised instruction for 30 minutes a day for 12-20 weeks with a specially trained teacher. In the US, successive administrations have encouraged interventions aimed at struggling readers. For example, in the 1990s, the Clinton administration’s ‘America Reads’ initiative encouraged the creation of programmes for volunteer tutors to work with

⁵ This is part of the formal National Curriculum for all children. Key Stage 1 assessments take place at the end of Year 2, when children are aged 7.

struggling readers. 'Reading First' was the Bush administration's initiative for children in early years of schooling, focused on high-poverty, low-achieving schools with a particular focus on small group interventions for struggling readers. In the UK, there have been various national initiatives designed to improve literacy for all children, such as the National Literacy Strategy in the 1990s and the change in national policy to recommend 'synthetic phonics' to all primary schools in the 2000s (see for example Machin et al., 2008, 2018). In the late 2000s, the UK government has also supported 'Reading Recovery' (described above) for low attaining students.

Slavin et al. (2011) review the considerable body of research amongst educationalists/psychologists that now exists on such reading programmes. Among their findings it is observed that small group tutorials can be effective, but not as effective as one-to-one instruction by teachers or paraprofessionals; teachers are more effective than paraprofessionals and volunteers as tutors; and traditional computer-assisted instruction programs have little impact on reading. This finding on the ineffectiveness of computer-assisted programs chimes well with the studies by economists who have evaluated this. Examples of relatively large-scale studies with a strong methodological design include those by Angrist and Lavy (2002), Rouse et al. (2004), and Berlinski and Busso (2017). These studies find no effect of teaching with ICT on pupil learning. A review by Bulman and Fairlie (2016) finds studies of ICT and computer-aided instruction in schools to produce mixed evidence with a pattern of null results, with notable exceptions of studies of developing countries and computer-aided instruction that target maths rather than language.

However, the fact that computer-aided instruction is often found to have zero effect does not mean this need always be the case. One would expect this to be influenced by the underlying pedagogy, the quality of the research design and the training of teachers/teaching

assistants that deliver the intervention; as well as the classroom context.⁶ Presumably, the reason why many schools use such programs is because they believe they are effective. The program being evaluated here (*ABRA*)⁷ has some support from small efficacy Randomised Control Trials (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 3 groups: a synthetic phonics intervention group, an analytic phonics intervention group and a classroom control group. The intervention groups were both using the *ABRA* computer program. The authors find that both interventions have a significant impact on literacy. Savage et al (2013) describe a classroom-level Randomised Control Trial (RCT) with just over 1000 pupils, and where the intervention is performed by teachers, also finding improvements in literacy for treated pupils.⁸ Our study differs from Savage et al (2013) along several dimensions. First, the size of the trial in terms of pupils is doubled. Second, this is the first evaluation that has been conducted by a team of independent researchers. Third, the intervention compares an ICT and Non-ICT technology programs that are equivalent in content, but which differ in the use of technology to deliver the program. Thus, we are able to assess whether the use of technology (i.e. software with graphics, sounds, and cartoon animations designed to appeal to young children) adds value when applying the same underlying pedagogy in the same context (i.e. teaching assistants, in the same schools, undertaking a paper version of the same program). Finally, and most importantly, the research design in this paper includes a clean control group with pupils in schools that do not receive and do not know about the existence of the web-based program while the intervention is in place. Thus, we can measure the extent of spillovers in treatment schools by comparing effects in the treatment group against students within two different control conditions (i.e. students not

⁶ Some studies suggest that technology does have potential to have a positive impact when implemented appropriately (e.g. Archer et al. 2014).

⁷ <http://www.concordia.ca/research/learning-performance/tools/learning-toolkit/abracadabra.html>

⁸ The effect size is in the region of 0.3-0.4 standard deviations, which varies by outcome measure.

assigned to receive the treatment within the treatment schools and students within the control schools).

3. The Intervention

Two literacy interventions are evaluated here and both consist of small group tuition for Year 1 pupils in English schools (i.e. pupils of age 5-6): one uses an ICT program (*ABRA*) and the other is identical (i.e. used materials that replicate the ICT intervention) but without using the computer program to deliver the content. Both methods were independently peer reviewed in advance of this study, and teaching assistants (TAs) were trained in the different approaches by academics who are experts in these areas.⁹ Table 1 gives a summary of the topics covered by the training approaches. The reading program consists of a balanced 20-week schedule of 15 minutes lesson plans, consisting of activities to develop phonics, fluency, and comprehension skills.

The ICT intervention, *ABRA*, is a modular game-based literacy intervention that is fixed in content (new activities cannot be added). The games are linked to a series of electronic texts (mainly ‘stories’, some non-fiction) suitable for beginner readers. The activities are aimed at phonics, word reading fluency, and text comprehension and there was a 20-week schedule of lessons planned for this study.¹⁰ There are extension activities for some of the tasks within *ABRA*, and these can be found in the “teacher area” of the website. Full details of the program are described in McNally et al. (2016).

⁹ Professor Robert Slavin (University of York, UK and Johns Hopkins University, Baltimore) reviewed plans for how the teaching assistants were to be trained in the different approaches and made recommendations on how the comparability of the different methods could be improved in advance. The training with the use of *ABRA* was provided by Professor Robert Savage (McGill University) and the training with the non-ICT methodology was provided by Professor Morag Stuart (University College London).

¹⁰ There are also activities for writing, but the implementation team chose not to include these in the 20-week schedule.

The Non-ICT intervention also covered the same 20-week schedule of lesson plans. The paper activities used materials such as magnetic letters and cards and a series of storybooks. To facilitate a clean comparison between the two delivery methods, the Non-ICT activities (especially developed for this study) were matched to each *ABRA* activity using the same stories, vocabulary items, questions, words and letter sounds in all the activities. Thus, the Non-ICT version was identical in content to the ICT version and only differed in terms of the delivery method.

Training occurred after schools had been randomised to the treatment and control conditions (discussed below) and after baseline testing of students in all schools. After school randomisation, treated schools provided the names of the teaching assistants that would participate in the intervention. For each school, a TA was assigned randomly to the ICT and Non-ICT condition before the training event.¹¹ Training within the ICT and Non-ICT condition was closely matched in terms of content but tailored for each specific mode of treatment delivery. Each TA was trained for 1.5 days (in a given approach) prior to the start of the intervention, in groups of 12-13 people. This consisted of a one-day training, ‘homework’ practice tasks and a further half-day of consolidation training. On average, each TA also received approximately 0.6 days of further post-training ‘just-in-time’ support from the project team (a mix of in-person, phone, and email support).

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) that has guided them on the activities to be performed 4 days per week during the 15-minute sessions. 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

¹¹ A small number of big schools had two TAs per condition.

weeks of treatment to observe how the intervention was delivered and to provide support for the TAs. The TAs were visited again about half way through the intervention.

During training, TAs received a list of pupils assigned randomly to them. Prior to the start of the intervention, TAs had some flexibility in arranging the small groups of pupils (around 3 to 4 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 3-4 pupils according to whether they were likely to be able to work well together. This was guided by ability, behaviour, special needs and personality. The process evaluation revealed 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. Schools were asked to deliver the programs during literacy-based lessons but not core literacy instruction, including phonics work. This is because the intervention was designed to complement (and not substitute for) normal classroom delivery of literacy. The process evaluation suggests this was faithfully adhered to by schools.¹² The broader context of English schools' approach to literacy is very phonics orientated and prescribed (e.g as discussed in Machin and McNally, 2018). If this intervention is found to benefit children's learning, then this shows that there is value in augmenting standard classroom practice with a wider range of reading activities than are currently used.

4. Methodology

The methodology is based on a Randomised Control Trial with two stages: (1) where 50 schools are randomised to treatment and control; (2) where pupils within treated schools are randomly assigned to one of three conditions: ICT, Non-ICT and a control group of students

¹² More details on the process evaluation can be found in McNally et al (2016).

within treated schools.¹³ The design of the experiment is illustrated in Figure 1 and the detail is explained below. An additional layer of randomisation is given by the random assignment of teaching assistants to either the ICT or Non-ICT condition within treated schools.

A. Participant selection

The implementation team at Coventry University first selected all schools with primary-aged children in the geographical areas near to them, covering schools in the West Midlands. The aim was to recruit about 60 schools, on the basis of power calculations made prior to the evaluation.¹⁴ They approached all 1682 eligible schools that included a Year 1 group in the school. A particular effort was made to encourage schools with disadvantaged intakes to participate during the recruitment stage.¹⁵ The participant schools are those that signed up for the intervention and actually implemented the baseline test for Year 1 students. Randomisation was conducted only after this baseline test had been completed. This applies to 50 schools.¹⁶

Five schools subsequently dropped out of the intervention, all of them in the treatment group. Of these, three dropped out immediately after randomisation took place and two dropped out later in the year.¹⁷ However, we were able to collect post-intervention data for 4 of these 5 schools that dropped out, and administrative (Key Stage 1 data) is available for all 50

¹³ The trial was registered under the title 'An Evaluation of Teaching Assistant-Based Small Group Support for Literacy' <http://www.isrctn.com/ISRCTN18254678>. It was conducted according to a protocol set out before the research was conducted. There were only a few small deviations from this protocol that are explained fully in the EEF report (please see McNally et al (2016) and the protocol description here): [https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Digital - Small Group Support for Literacy.pdf](https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Digital_-_Small_Group_Support_for_Literacy.pdf).

¹⁴ The calculations to decide on the sample size included in the protocol were performed using the Optimal Design (OD) Software (Spybrook et al, 2011) and is explained further in McNally et al (2016).

¹⁵ The remit of the commissioner (the Education Endowment Foundation) is especially focused on raising the attainment of disadvantaged students.

¹⁶ A further 7 schools originally agreed to take part, but 6 pulled out before baseline testing due to changed circumstances and 1 pulled out after baseline testing (but before randomisation) because they found the process too disruptive.

¹⁷ Two of the schools that dropped out immediately after baseline testing did so because they could not see how to integrate the intervention with their current literacy provision and worried that the children might get confused. One school dropped out during the intervention because of staffing issues and the other because of a change in the head teacher.

participating schools. This enables us to perform an Intention to Treat (ITT) analysis using most of the original randomised schools, though we also show results that estimate the Treatment on the Treated (TOT).¹⁸ Our full sample consists of 48 schools (or 50 when using the outcome variable from administrative data), half of which were randomly assigned to receive the treatment.¹⁹ Schools were told that they would either receive the treatment in 2014/15 or 2015/16. Thus, the control schools received the treatment in 2015/16. Importantly, the treatment is focused on Year 1 students and thus the cohort of interest to us (i.e. those in Year 1 in 2014/15) will never receive the treatment in control schools.²⁰ This enables us to consider the effects of the intervention one year later.

B. Randomisation

School-level randomisation was conducted within pairs of schools. Initially a number of variables based on administrative data on schools was used to assign each school to its closest pair. These variables included the size of the relevant cohort; the Key Stage 1 average point score (i.e. based on teacher assessment for students at age 7) for the relevant cohort in the preceding academic year (2013), and a measure of the percentage of pupils classified as being eligible to receive free school meals.²¹ Within each pair, one of the schools was randomly allocated to be in the treatment group, with the other allocated to the control group. We then randomised students in treated schools to one of three groups: (1) the ICT treatment; (2) the Non-ICT treatment and; (3) control pupils in treatment schools. Finally, and as mentioned above, an additional layer of randomisation is given by the random assignment of the teaching

¹⁸ 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 (except when the outcomes are defined using Key Stage 1 administrative data, where we can use the full sample of 50 schools).

¹⁹ Results are very similar if we use the 48 schools for all outcome variables.

²⁰ Furthermore, only 10 of the 25 control schools actually elected to take up the treatment for their Year 1 cohort in 2015/16.

²¹ In addition, infant schools were paired together (i.e. those catering for pupils of age 4-7; the majority of primary schools cater for pupils of age 4-11).

assistants participating in the intervention in treated schools, to either the ICT or Non-ICT conditions.

C. Data and outcome measures

The primary outcome was measured (pre and post-treatment) by the Progress in Reading Assessment (PIRA) test. This is an age-standardised test that evaluates the general reading ability of pupils.²² In particular, it assesses reading ability in the following areas: phonics, literal comprehension and reading for meaning, which are the areas that the intervention targets. The secondary outcomes assess more specific components of reading and are not discussed here (results available on request). These tests took place at the beginning (September 2014, all before randomisation) and end (July 2015) of the academic year 2014-2015. We also analyse the primary outcome one year after the completion of treatment (test administered in July 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 to. Furthermore, the RAs were blind to the nature of the study – i.e. 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 has been scored by Hodder Education. All other tests have been 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 have 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.

²² More information on the PIRA test can be found here: <https://www.hoddereducation.co.uk/pira>

One year subsequent to the intervention, pupils get to the end of ‘Key Stage 1’ and receive teacher assessments. The national curriculum in England is organised around ‘Key Stages’, within which various goals are made out for children’s learning and development and this ends with a formal assessment. Although pupils are assessed by their own teachers at the end of Key Stage 1, there is extensive guidance on how the assessment should be made and it is moderated. As the pupils are in a different school year, the assessment is not made by the same teachers who taught them during the year of this intervention. The results of the teacher assessment are available in administrative data (the National Pupil Database).

The outcome variables are as follows: (1) PIRA test at endline (i.e., July 2015); (2) PIRA test one year later (July 2016) and (3) Key Stage 1 Reading one year later. The last of these measures is a binary variable, which indicates whether students are at or above the expected level as defined by the National Curriculum. We standardize the PIRA test score to have mean zero and standard deviation of one.²³

We also incorporate administrative data on pupils as additional control variables: eligibility for free school meals, gender and whether the pupil achieved a good level of development in the Foundation Stage Profile (FSP GLD), which is assessed by teachers when children are at age 5 (their first year of school) in all schools across the country according to standardised criteria.²⁴ In this Foundation Stage Profile, pupils are assessed in relation to 17 early learning goals.

The final distribution of pupils in treatment schools before the start of treatment was as follows: ICT treatment (360 pupils), Non-ICT treatment (350 pupils), and control pupils in

²³ The raw PIRA test score is a continuous variable that can take values from 0 – 25. The age standardised scores range from 70 – 130.

²⁴ The variable used here 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.
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/488745/EYFS_handbook_2016_-_FINAL.pdf.

treatment schools (373 pupils) (see Appendix Table 1). There were 1158 pupils in the control schools. Because of school and pupil attrition, our analysis is based on 80 to 95% of the originally randomised sample, depending on the outcome measure analysed (see section below and Appendix Table 1 for further details on the level of missing data for the three different outcome variables and across different groups). More details about balance of predetermined characteristics for those observed at endline (for each of the outcome variables) are given in Section 5.

D. Empirical Approach

To estimate the intention-to-treat (ITT) impact, we estimate a regression where the outcome variable is regressed against dummy variables for whether individuals were originally randomised to the ICT or Non-ICT treatment groups (relative to the control group). In our main regression, we also include a dummy for assignment to the control group within treated schools (CT). We control for the school pair in which schools were originally randomised and the baseline test results. We also report results from an augmented regression where we control for predetermined characteristics of students. Given the randomised nature of the intervention, the point estimates should not be greatly affected by the inclusion of additional controls. However, we would expect it to be important for the precision of estimates given a limited number of school clusters. Thus, our most detailed ITT specification can be described as follows:

$$Y_{ist} = \beta_1 ICT_{ist} + \beta_2 NonICT_{ist} + \beta_3 CT_{ist} + \beta_4 Y_{ist-1} + \beta_5 X_{ist-1} + \rho_s + \varepsilon_{ist} \quad (1)$$

Where Y_{ist} is the test outcome for person i in school s at time t . As discussed above, we also run this regression using outcomes measured one year later. We are interested in the effects of being assigned to the ICT or Non-ICT treatment (i.e. β_1 and β_2) conditional on baseline scores (Y_{ist-1}), a vector of personal predetermined characteristics described by X_{ist-1} (which includes gender, eligibility to receive free school meals prior to treatment and whether the pupil

achieved a good level of development in the Foundation Stage Profile), and the school pair ρ_S . Standard errors are clustered at the level of the school (i.e. the first stage of randomisation). We are also interested in establishing whether there is any spillover effect of the treatment to control students within treated schools (i.e. β_3).

We estimate this regression for different subgroups using the ‘clean’ control group only.²⁵ These subgroups are defined on the basis of free school meal status; gender; above median attainment on pre-test (i.e. PIRA test at baseline). This is of interest in that the effects of the treatment may be heterogeneous between pupils with different characteristics.

Given that 5 schools in the treatment group dropped out (3 immediately after randomisation, and 2 during the intervention), we also estimate instrumental variable regressions, using the initial random allocation of students as instruments for the final treatment received. We perform this exercise using the ‘clean’ control group. See the ‘Note on Methodology’ in Appendix for further detail.

5. Results

A. Balance at baseline

Table 2 shows characteristics of treatment and control schools in terms of the number of teaching assistants (TAs), teachers, the ratio of TAs to teachers, teacher qualifications, salaries and the size of the Year 1 cohort. There is no difference between those schools assigned to treatment and control in these respects. There are about 50 pupils on average within the Year 1 group, which implies about two classes per school. The ratio of TAs to teachers is very close to the national average and close to 0.8 for both treated and control schools. This implies that on average, there is almost one TA per teacher.

²⁵ The main regressions are shown including and excluding control students in treatment schools. The coefficients on the ICT and Non-ICT treatments are almost the same. Having made the point about spillover effects, we only use the ‘clean’ control group for subsequent analysis on subgroups etc. (i.e. we exclude control students within treatment schools).

Table 3 shows characteristics of TAs within treatment schools that are assigned to the ICT and Non-ICT conditions. The information in Panel A of Table 3 is available for all teaching assistants in treated schools (except for the 3 schools that dropped out immediately after randomisation); and for slightly less TAs in Panel B. As TAs were randomly assigned to the ICT and Non-ICT condition, it is not surprising to see that for the most part, their characteristics are similar on average within each condition. The average TA is in her/his early 40s with about 10 years of experience as a TA.²⁶ The percentage with qualifications of ‘level 3 or more’ (corresponding to at least upper secondary education) is 84 percent for those assigned to the ICT condition and 67 percent for those assigned to the Non-ICT condition.²⁷ Information from the TA baseline survey shows that most TAs use information technology (IT) professionally both for the teaching of literacy and numeracy and over 40 percent use IT professionally every day or for every lesson. For the most part TAs feel comfortable using IT for teaching. This applies to 68 percent of those TAs assigned to the ICT condition and 47 percent of TAs assigned to the Non-ICT condition.

Table 4 shows characteristics of students assigned to control and treated schools (columns 1 and 2, respectively); and then within treated schools, those assigned to the ICT, Non-ICT or control condition (columns 3, 4 and 5, respectively). The characteristics are those used in the regression analysis: the student’s gender; eligibility for free school meals; whether he/she has achieved a ‘good’ level of development as measured by teachers in the previous year for the Foundation Stage Profile (described above); and the baseline PIRA reading test. There is almost no difference between any of the groups with respect to any of these characteristics. The one exception is whether pupils were assessed as having a ‘good level of

²⁶ Only 3 out of the 52 TAs are male (1 in the ICT and 2 in the Non-ICT condition).

²⁷ In terms of tertiary education, 28% of TAs in the ICT condition have a Higher Education degree; and 8% of the TAs in the Non-ICT condition.

development' within the Foundation Stage Profile.²⁸ On average, this is higher in control schools (at 54 percent) compared to treatment schools (at 48 percent). Otherwise, the groups are fairly well balanced.

We analyse whether attrition is a threat to validity to our estimates by checking balance at endline, for each of the three outcome variables. The results are very similar to those found at baseline and for the three outcomes and are available upon request. Therefore, attrition has not worsened balance on observables across the different conditions. Nonetheless, we show results with and without controlling for baseline characteristics for the main specifications.

B. Main results for reading

Estimates of the 'Intention to Treat Effects' are shown in Table 5. Columns (1) and (2) show estimates of equation (1) for all students. Columns (3) and (4) exclude control students within treatment schools. In each case, we show a specification with minimal controls (i.e. the school pair dummies and the baseline reading score) and an augmented version (including controls for gender, eligibility for free school meals and whether the pupil achieved a 'good level of development' in the Foundation Stage Profile at age 5). The simple specification is shown in columns (1) and (3) and the augmented specification is shown in columns (2) and (4). We show three panels of results, with Panel A being the 'intention to treat' effect within the same school year (i.e. about two months after the end of treatment). Panel B shows results when the outcome variable is the same reading test administered one year later. Panel C shows results when the outcome variable is defined as a binary variable indicating whether the student achieves the 'expected level' in the Teacher Assessment that is conducted one year after the intervention (in line with national requirements described above).

²⁸ The p-value is 0.01. There is one other difference where the p-value is less than 0.10 (i.e. 0.09). There are fewer females within the control condition in treated schools compared to the two treatment conditions (i.e. 45% compared to about 51%).

In each case, the point estimates of the effects are slightly higher in the augmented specification. Unsurprisingly, the estimated effect of assignment to the ICT and Non-ICT conditions is approximately the same whether or not we exclude control students within treatment schools. This is because we include a binary variable for whether or not students are assigned to that group (in columns 1 and 2).

We first consider the short-term effects of the intervention on the reading test conducted at the end of the same school year (Panel A, Table 5). The effect of being assigned to the ICT condition moves from 0.14σ to 0.18σ from the simple to the augmented specification. The effect of being assigned to the Non-ICT condition moves from 0.25σ to 0.27σ . Both interventions have a significant effect; although the impact of the Non-ICT intervention is about 50% bigger (and the p-value of the difference between assignment to the ICT and Non-ICT intervention is just over 0.10). However, the effect of being assigned to the control condition within treatment schools (captured by the CT dummy in Table 5) is almost the same as being assigned to the ICT condition (and is not significantly different). Thus, there is a substantial spillover effect. We leave possible interpretations of the spillover effects to Section 6 where we discuss mechanisms.

Panels (B) and (C) enable us to consider the effects of the intervention in the next school year. By this time, pupils will have been exposed to another full year of teaching with a different teacher and different teaching assistants. In Panel B, the outcome variable is the PIRA reading test.²⁹ Any spillover effect disappears as the point estimate is close to zero for being assigned to the control condition within treatment schools. The magnitude of the intention to treat effect of being assigned to the ICT or Non-ICT condition reduces considerably. In the augmented specification, the point estimate is 0.08σ and 0.10σ for the ICT and non-ICT condition respectively. However, the standard errors remain roughly the same as in Panel A,

²⁹ This is the Year 2 Summer version of the test, to take into account that students are a year older.

which is almost as high as the estimated effects. Thus, at conventional levels of significance, we are unable to say whether or not the intervention continued to have an effect on pupils when using the PIRA test.

In Panel C, we show results where the outcome variable is whether or not the pupil achieved the ‘expected reading level’ according to the (‘Key Stage 1’) Teacher Assessment. The baseline (in the control group) is 74 percent. Again, there is no evidence of a *spillover* effect (with the point estimate being close to zero). Estimates of the intention to treat effect are 0.02 and 0.06 (i.e. 2 and 6 percentage points) in the ICT and Non-ICT conditions respectively within the augmented specification. This is significantly different from zero in the case of the Non-ICT condition. Thus, these results give firmer evidence that the effect of the intervention did endure for the Non-ICT condition.

Appendix Table 2 shows the impacts of the ICT and Non-ICT conditions when we scale up the results to show the ‘Treatment on the Treated’ effects. The regressions in the table use the ‘clean’ control group.³⁰ Point estimates increase slightly to 0.22σ and 0.34σ when using the PIRA at endline outcome variable for the ICT and Non-ICT conditions, respectively (column 1); to 0.09σ and 0.12σ one year later (though not statistically significant); and to 0.02 and 0.07 (i.e. 2 and 7 percentage points) when using the binary variable capturing whether the student has achieved the expected reading level at the end of Key Stage 1 (column 3). The estimated impacts are close to the ITT results because the assignment to treatment and the final treatment received were not very different in most cases (as can be seen by the magnitude of the main coefficients in the ICT and Non-ICT first stages in Panels B and C).

It is difficult to compare the reading test to the teacher assessment because the latter is a binary variable and the former is a continuous variable. Of course, they are also different types of assessment and may give different results for that reason. To make results more

³⁰ We exclude control students in treated schools in the remaining tables throughout the paper.

comparable, we do two things: (1) we convert the reading test to a binary variable based on how the teacher assessment indicator corresponds to the average reading test score (at endline and endline+1, respectively) within control schools³¹; (2) we convert the teacher assessment indicator to a more continuous measure. The extent to which we can do the latter is limited because there are only 4 possible values.³² Results are reported in Table 6. Column (1) shows results where the outcome is the PIRA reading test at the end of the same school year. Columns (2) and (3) show results where the outcome is measured one year later either in the same reading test (column 2) or in the teacher assessment (column 3). Here we report coefficients on the other variables because it is interesting to notice how the magnitudes of the coefficients are similar for the two different assessments measured at the same time (i.e. columns 2 and 3). With regard to the main coefficients of interest, a comparison between columns 2 and 3 shows that results are very similar if we try to measure the reading test and the teacher assessment on a comparable scale, whether this is binary (Panel A) or more continuous (Panel B)³³. Comparing point estimates for the outcome variable in the same year as the intervention (column 1) compared to one year later (columns 2 or 3) suggests that the effect one year later might be around one-third of the original effect.

C. Results for other subjects

Although the intervention was targeted on activities particularly important for reading, it might also impact on other subjects. There is an obvious connection between reading and writing. Machin and McNally (2008) show that there is a strong relationship between reading demands of tests in maths and reading. Specifically, an analysis done on the age 11 reading and maths test showed that the reading demand of the maths test (based on text difficulty) is nearly 70

³¹ We refer the reader to the notes in Table 6 for more detail on how we construct the binary variables at endline and endline+1 (with information from the continuous PIRA at endline and PIRA at endline+1, respectively).

³² The level assigned in the KS1 reading teaching assessment can be transformed into a point score using scales available from the Department for Education. We then standardise this variable using the mean and standard deviation in the whole sample of available students for this outcome.

³³ The results are very similar if we use probit/logit regressions for binary outcome variables.

percent of what it is in the reading assessment. We do not have test outcomes for other subjects immediately after the intervention but we do have Teacher Assessments for reading, writing and maths in administrative data at the end of the subsequent year when pupils are age 7.

Table 7 shows results for reading, writing and maths respectively where the outcome variable is one if the pupil achieves at least the ‘expected level’ in these subjects. The effect is only statistically significant in the case of reading and writing and for the Non-ICT treatment only. Specifically, the effect of assignment to the Non-ICT condition increases the probability of achieving the ‘expected level’ in reading and writing by 0.06 and 0.08 (i.e. 6 and 8 percentage points respectively). The point estimate for maths is also positive (0.05) but not statistically significant. Assignment to the ICT condition does not show effects that are statistically significant. However, point estimates are 0.02, 0.04 and 0 for reading, writing and maths, respectively, and thus show a pattern of results that is consistent with estimates for the Non-ICT condition, and with the overall short-term results.

D. Results across the distribution

It may be that effects vary across the distribution as some students may benefit more than others from this additional help from teaching assistants. In Table 8, we show results from quantile regressions using the reading test administered at the end of the intervention and one year later. These results show that the Non-ICT intervention has an effect throughout the distribution and a stronger effect above the median, when the outcome variable is measured at endline (Panel A). In contrast, the effect of the ICT intervention is more pronounced in the bottom 25th percentile of the distribution, both in terms of size and precision, particularly when measured one year after the end of the intervention. These are the only students for whom an effect seems to significantly endure beyond the first year for the ICT condition. The patterns found in Panel B for the Non-ICT condition suggest that, a year after the end of treatment, there is still a sizable impact throughout the distribution, especially at the median and the 75th percentile (although

most of these effects are not statistically significant). Some of these findings also resonate with the heterogeneous impacts discussed below.

E. Heterogeneity

In Table 9, we show results where each treatment dummy is interacted by an individual characteristic: whether the pupil is eligible to receive free school meals (FSM) (panel A); gender (panel B); and whether he/she is above or below the median of the baseline test (panel C). In each case, we include four “treatment” variables defined according to the ICT/Non-ICT treatment status and the characteristic under study. We exclude students in treatment schools who were assigned to the control condition. We show three columns of results: the reading test at the end of the intervention year (column 1), the same reading test at the end of the subsequent year (column 2) and a binary variable for whether the pupil achieved the ‘expected level’ in the Key Stage 1 teacher assessment (also one year after the intervention).

The short-term effect of the intervention was much stronger for FSM pupils compared to non-FSM pupils. For FSM students, the effect was about half of a standard deviation for both the ICT and non-ICT conditions. This would close the gap between FSM and non-FSM students (as this is about 0.27σ whereas the effect of the non-ICT intervention was 0.22σ for non-FSM pupils). The group for whom the intervention was least effective was non-FSM students assigned to the ICT condition (where the point estimate is 0.11σ and not statistically significant). However, these effects all diminish one year after the intervention. The point estimates suggest that the group least likely to benefit are still the non-FSM students assigned to the ICT condition whereas effects are more likely to endure for FSM students.

In panel B, we show effects by gender. Although point estimates for the short-term effect suggest a slightly bigger effect for girls than boys, the difference is not statistically significant. There is fade-out for all groups. However, the point estimates suggest that girls assigned to the Non-ICT condition benefit most in the short-term (column 1) and also in the

longer term if we consider the indicator variable for whether pupils achieve the expected level in reading (column 3). Girls assigned to the Non-ICT condition are more likely to achieve this standard by 9 percentage points whereas the point estimates are smaller and not statistically significant for girls assigned to the ICT condition or for boys assigned to either condition.

Finally, in panel C, we show results according to whether the pupil scored above or below the median of the baseline PIRA test. The first column suggests that the short-term effect of the Non-ICT intervention was about the same, regardless whether the pupil was above or below the median. The magnitude of the effect is also similar to those assigned to the ICT intervention if they scored below the median in the baseline test. A lower point estimate (which is not statistically significant) is found for pupils above the median who were assigned to the ICT intervention. Although these effects fade out in the subsequent year, the same pattern of effects is observed for the reading test (column 2). The teacher assessment outcome (column 3) suggests a somewhat different pattern because in this case the point estimates are higher (and statistically significant) only for those students who scored above the median in the baseline test. However, differences are not strong enough to generate a statistically different effect compared to students who scored below the median. In further investigation (not reported), we find that the significant effects found in Panel C, Column 3 (for both the ICT and Non-ICT condition) are driven by students scoring above the median who are also eligible to receive free school meals.

6. Mechanisms

The training of teaching assistants both for the ICT and Non-ICT condition had a positive effect on the educational outcomes of pupils in the short-term. There is some evidence that effects endure, particularly in the case of the Non-ICT intervention. It would appear that the latter

intervention is effective for most groups of students whereas the ICT intervention is more selective in who it benefits.

In considering mechanisms, we first discuss how to interpret differences between the treatment and control group. Then we discuss how we might interpret the spillover effect (evident in the short-term but not one year later). Finally, we discuss possible reasons for why the Non-ICT version of this intervention appears to be more effective than the ICT version.

The intended interpretation of this RCT is that differences between the treatment and control group of schools can only be attributed to the effect of training teaching assistants in the use of the pedagogy applied here. A threat to this interpretation would exist if treatment schools actually increased the hours devoted to literacy as a result of the intervention (potentially at the cost of other activities for which we have no measure of outcomes). Table 10 shows results from a survey of treatment and control schools that was undertaken at the end of the school year in which the intervention took place.³⁴ This shows that the hours devoted to literacy instruction was approximately the same in treatment and control schools and that schools were also similar to each other with regard to the use of computers and other forms of IT to support teaching.

Another threat to the interpretation of findings would be if there was a ‘Hawthorne effect’, whereby treatment schools improve relative to the control group simply because the fact of there being *any* intervention is an impetus to increase effort. This would certainly be a potential explanation for a large spillover effect within treatment schools. While one cannot rule out some effect from being put under the spotlight, the strongly heterogeneous effects of the interventions would move against such an interpretation. For example, the effects of the intervention are much stronger for pupils from disadvantaged backgrounds compared to others.

³⁴ The results of this exercise are informative but need to be taken with caution since the data is only available for 29 schools (out of 50 schools that were randomised).

This is particularly evident in the results after the first year of the intervention. Thus, the most obvious interpretation of the intervention is that the training of teaching assistants in the use of this particular pedagogy, along with its practical implementation, was effective for students.

However, the results show a strong spillover effect to control students within treatment schools. Even though this does not last beyond the year of the intervention itself, the strong magnitude of this spillover effect in the short term is something of a puzzle. A suspicion might be that the parents or teachers of students in the control condition might have found out about the methods used by the teaching assistants and started using the resources more broadly. However, the (independently conducted) process evaluation suggests that this is extremely unlikely. Firstly, it was not straightforward even to apply the intervention to the treatment groups. Logistical issues that affected the majority of TAs included taking pupils to and from sessions; space within the school and the short length of sessions. Secondly, the external process evaluation did not find that schools were compensating for the program by delivering additional help to pupils in the control group. Finally, the identity of the computer program was suppressed throughout the evaluation and known only to TAs and students that saw the name of the program when actually using it.³⁵

It seems more likely that the spillover effect arises from the training to TAs, which might have affected their other activities with the Year 1 group as a whole. Most TAs on the project were drawn from those working with Year 1 pupils. As the pupils did plenty of other literacy activities outside the intervention time, there would have been opportunity for TAs to use any new skills they had learnt to help pupils informally at other times. Feedback from TAs given in the context of the process evaluation was that they perceived it to have improved their

³⁵ The intervention was closely monitored by the implementation team throughout (with TAs receiving visits) and fidelity to the design was strongly emphasised. TAs were asked to keep the interventions distinct by not sharing information about the content and delivery of the two programs. Process evaluators found only a low level of awareness among TAs for the training program that they were not trained to implement (in a post-treatment survey answered by 35 TAs, only 17% of the TAs answered that they saw the intervention of the other TA within their school).

skills in small group tuition. Moreover, data from a post-treatment survey (answered by more than 70% of the TAs) shows that 74% of TAs had a better or much better understanding of phonics after the intervention, and 69% of TAs were confident or very confident to deliver small group teaching after the intervention.

Also, it is possible that the reduced number of students in the class (albeit for short periods) might have helped the class teachers with other students. Or it might be the case that the teacher was able to advance the whole class more quickly on account of the fact that two-thirds of the year group were exposed to this intervention, which complemented core literacy instruction. In any case, the spillover effect does not last into the subsequent year and the Non-ICT intervention has a more enduring impact than the ICT intervention (at least on average). So why might the Non-ICT intervention have been more effective?

We first consider whether compliance was different for teaching assistants assigned to either type of intervention. Table 11 shows scores for daily record keeping and the use of levels (which indicates the extent to which TAs were moving pupils through different layers of the program adequately). These measures suggest a high level of compliance for TAs assigned to both treatments. Even though those assigned to the Non-ICT condition perform slightly better on daily record keeping, it would be hard to believe that this could explain the stronger and more enduring effect for pupils being assigned to the Non-ICT treatment. Also, although TAs were allowed to decide how to group pupils assigned to each condition, there was no difference in the size of groups or their composition between the ICT and Non-ICT condition. This is shown in Table 12.

Although one might think that technical problems could jeopardise the ICT intervention, in practice any technical problems with implementing the ICT intervention were minor and occasional. Furthermore, the process evaluation found that both interventions were extremely popular with TAs and with pupils. The training for interventions was also equally

well received.³⁶ The process evaluation found that the Non-ICT intervention was perceived to have greater adaptability to different ability levels by TAs. This may lie at the heart of the differential effectiveness because it is consistent with the fact that the Non-ICT intervention shows stronger effects throughout the ability distribution. Thus, it might be that when confronted with different levels of ability and progression, the TAs and pupils found it easier to use books and magnetic letters to advance learning rather than the medium of a computer screen. This is consistent with the large body of research (cited above) suggesting that computer-aided instruction is not in and of itself any better than what it replaces.³⁷

This study shows that teaching assistants can be deployed very effectively to supplement classroom teaching with small, short tutorial sessions, using a highly structured evidence-based approach. Most of the TAs already had some experience of using literacy programmes with small children, but their feedback suggested that this intervention was unlike anything 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 the delivery. The TAs in this study reported feeling well prepared for the intervention in terms of training and well supported throughout by the implementation team.

7. Conclusion

In this study, we get inside the ‘black-box’ of the education production function from within the classroom. The experiment provides an opportunity to evaluate whether teaching assistants can be effectively deployed to complement the work of the teacher. This study shows a context of how teaching assistants (who are employed by most primary schools) can be used to better

³⁶ The qualitative methods used in the process evaluation are documented in McNally et al. (2016).

³⁷ An additional disadvantage of the computer program in this particular context is that there were Canadian English pronunciations, which might have affected the learning experience of students.

effect to improve the literacy of young children. Teaching training has been shown to be important in other contexts (e.g. Angrist and Lavy, 2001). Here we show that training of teaching assistants can also be an effective way to improve student outcomes.

Further, we are able to distinguish the effects of the training of TAs and pedagogy from the effect of the medium of delivery of the intervention (whether ICT or Non-ICT). Although both modes of delivery show positive effects on pupil outcomes, the Non-ICT mode of delivery has a stronger and more enduring effect. This shows that although computer-aided instruction can be useful, it does not (in and of itself) add value to such pedagogical approaches.

Given that both interventions were delivered by TAs, who are not very highly qualified (or highly paid), the per-pupil costs of delivering this intervention were modest. We estimated that the per-pupil cost (including the training of TAs; support provided during the project etc.) was about £25. This low per pupil cost implies that effects do not have to be very large before the intervention becomes cost effective. Although there is some evidence of fade-out, the one year follow up does suggest that effects endure (at least beyond the year of the intervention). This is most evident with respect to the effect of the Non-ICT intervention on the probability of being at or above the 'expected level' at age 7 in teacher assessments of reading and writing. Finally, this is an intervention that disproportionately benefits students from a lower socio-economic background. Although this is most evident for short-term outcomes, it is also true for outcomes measured one year later. Thus, such an intervention helps to level the playing field between pupils from different socio-economic groups.

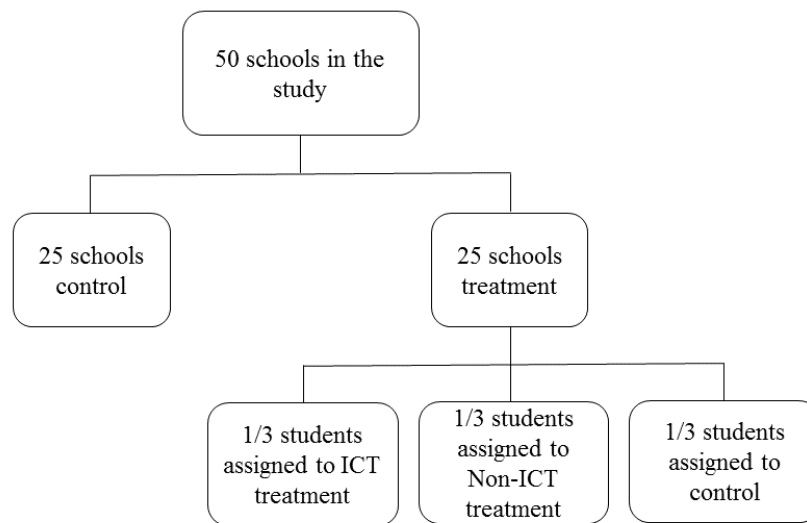
References

- Aaronson D., Barrow, L., and Sander, W. 2007. "Teachers and Student Achievement in the Chicago Public High Schools", *Journal of Labor Economics*, 25: 95–135
- Abrami, P.C., Savage, R.S., Deleveau, 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. 2001. Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools. *Journal of Labour Economics* 19(2): 343-369
- Angrist, J., and Lavy, V. 2002. "New Evidence on Classroom Computers and Pupil Learning", *Economic Journal*, 112: 735-765
- Araujo, M., Carneiro, P., Cruz-Aguayo, Y., and Schady, N. 2016. "Teacher Quality and Learning Outcomes in Kindergarten", *Quarterly Journal of Economics*, 131: 1415-53
- Archer, K., Savage, R.S., Sanghera-Sidhu, S., Wood, E., Gottardo, A., and Chen, V. 2014. "Examining the effectiveness of technology use in classrooms: A tertiary meta-analysis", *Computers and Education*, 78:140-149
- Berlinski, S. and Busso, M. 2017. "Challenges in Educational Reform: An Experiment on Active Learning in Mathematics", *Economics Letters*, 156:172-175
- Bulman, R., and R.W. Fairlie. 2016. "Technology and Education: computers, software and the Internet". In E. A. Hanushek, S. J. Machin and L. Woessmann (Eds.). *Handbook of the Economics of Education*. Volume 5: 239-280
- Chetty, R., Friedman, J.N., and Rockoff, J. 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates", *American Economic Review*, 104: 2593-632
- Chetty, R., Friedman, J.N., and Rockoff, J. 2014b. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood", *American Economic Review*, 104: 2633-79
- Comaskey, E. M., Savage, R. S., & 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: 92–108
- Cunha, F., J. Heckman, L. Lochner and D. Masterov 2006. "Interpreting the evidence on life cycle skill formation" in E. Hanushek and F. Welch, *Handbook of the Economics of Education*, Vol 1. North Holland.
- Hanushek, E., Rivkin S., and Kain, J. 2005. "Teachers, Schools and Academic Achievement", *Econometrica*, 73: 415–58
- Jacob, B. 2017. "When Evidence is Not Enough: Findings From a Randomized Evaluation of Evidence-Based Literary Instruction (EBLI)", *Labour Economics*, 45: 5-16
- Kuczera, M., Field, S. and Windisch, H. 2016. *Building Skills for All: A Review of England*. OECD.
- Machin, S. and S. McNally. 2008. "The Literacy Hour", *Journal of Public Economics*, 92: 1441-62
- Machin, S., S. McNally and Viarengo, M. 2018. "Changing How Literacy is Taught: Evidence on Synthetic Phonics", *American Economic Journal: Economic Policy*, 10(2): 217-41
- McNally, S., Rolfe, H., and Ruiz-Valenzuela, J. 2016. "ABRA: Online Reading Support Evaluation report and executive summary". Report for the Education Endowment Foundation.

- https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_ABRA.pdf
- 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., & Deault, L. 2009. "A randomized controlled trial study of the ABRACADABRA reading intervention program in grade 1", *Journal of Educational Psychology*, 101(3): 590
- Savage, R., Abrami, P. C., Piquette, N., Wood, E., Deleveaux, G., Sanghera-Sidhu, S., & Burgos, G., 2013. "A (Pan-Canadian) cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program", *Journal of Educational Psychology*, 105(2): 310
- Slavin, R.E., C. Lake, S. Davis, Madden, N.A. 2011. "Effective Programs for Struggling Readers: A Best-evidence Synthesis", *Educational Research Review*, 6: 1-26
- Spybrook, J., Bloom, H., Congdon, R., Hill, C., Martinez, A., Raudenbush, S. 2011. *Optimal Design Plus Empirical Evidence: Documentation for the "Optimal Design" Software*. http://sitemaker.umich.edu/group-based/optimal_design_software
- Taylor, E.S., and Tyler, J.H. 2012. "The Effect of Evaluation on Teacher Performance", *American Economic Review*, 102(7): 3628-51
- Vignoles, A. 2016. "What is the Economic Value of Literacy and Numeracy? Basic Skills in Literacy and Numeracy are Essential for Success in the Labor Market", *IZA World of Labor*, 229.
- Wolgemuth, J. R., Savage, R. S., Helmer, J., Lea, T., Harper, H., Chalkiti, K., Bottrell, C., & 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: 727-750

Tables and Figures

Figure 1: Design of the Experiment



Note. The focus of the analysis is on state schools. Within each school, teacher assistants were also randomised to the ICT and Non-ICT condition, respectively.

Table 1. Content of 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/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 feedback experience of delivering sessions and activities
- structured sessions to trouble-shoot and share good practice

Note. An in-depth description of the content of both interventions can be found in Appendix A and B in McNally et al (2016).

Table 2: Baseline Characteristics – characteristics of treatment and control schools

	Control Schools (1)	Treatment schools (2)	Difference in Means (P-Values & Observations) (3)
Total number of teaching assistants (Full-time equivalent)	12.40 (6.848)	12.31 (7.743)	0.960 (50)
Total number of teachers (Full-time equivalent)	15.65 (6.899)	16.31 (10.13)	0.759 (50)
Ratio of teaching assistants to all teachers	0.772 (0.223)	0.758 (0.262)	0.695 (49)
Teachers with Qualified Teacher Status (%)	97.34 (4.643)	98.22 (3.378)	0.455 (50)
Mean Gross Salary of all Teachers (in 000s)	36.28 (1.890)	35.59 (2.133)	0.248 (50)
Size of the Year 1 cohort	51.44 (20.02)	52.76 (27.33)	0.712 (50)

Note. Data comes from the School Workforce Dataset (November 2014), except data on the size of the year 1 cohort, that was collected from the implementation team directly from the school records. Columns 1 and 2 show means (first row) and standard deviations (in parentheses). P-values are calculated using pairing fixed effects and robust standard errors (column 3).

Table 3: Characteristics of TAs assigned to each condition

	ICT (1)	Non-ICT (2)	Difference in Means (P-Values & Observations) (3)
Panel A. Information from Curriculum Vitae of Teaching Assistants			
Age TA in first term academic year 2014-2015	42.46 (11.76)	42.57 (8.417)	0.970 (49)
Years of teaching assistant experience	9.800 (7.331)	10.46 (7.271)	0.747 (52)
TA has any qualification of level 3 or more	0.840 (0.374)	0.667 (0.480)	0.154 (52)
Panel B. Information from TA baseline surveys			
Use of IT (professionally) for literacy	0.955 (0.213)	0.868 (0.347)	0.336 (42)
Use of IT (professionally) for numeracy	0.955 (0.213)	0.816 (0.398)	0.17 (42)
Use IT professionally every day or lesson	0.409 (0.503)	0.457 (0.513)	0.769 (40)
TA feels comfortable or very conf. using IT for teaching	0.682 (0.477)	0.474 (0.512)	0.185 (42)

Note. The information in this table comes from data collected via standardised curriculum vitae sheets and other pre-information survey. Columns 1 and 2 show means (first row) and standard deviations (in parentheses). P-values are calculated using robust standard errors (column 3). [Results are very similar when we also include school fixed effects or when we cluster the standard errors at the school level. Due to the low number of observations and clusters, and the fact that in the second panel we miss information for some of the TAs in some categories, we show the results without including school fixed effects and without clustering standard errors at the school level]. Observations have a weight of 1 if there is only one teaching assistant per group; and 0.5 when there are two teaching assistants per group (due to replacements or TA sharing in a very small number of cases).

Table 4. Balance checks at baseline: students

	Baseline Variable Means and Standard Deviation					Difference in Means (P-values & Observations)					
	Control Schools (1)	Treatment schools (2)	ICT (3)	Non-ICT (4)	Control in Treatment schools (5)	[2] vs [1] (6)	[3] vs [1] (7)	[4] vs [1] (8)	[4] vs [3] (9)	[5] vs [3] (10)	[5] vs [4] (11)
Panel A. Individual characteristics											
Female	0.498 (0.500)	0.494 (0.500)	0.516 (0.500)	0.513 (0.501)	0.455 (0.499)	0.555 (2221)	0.466 (1511)	0.677 (1501)	0.963 (696)	0.087 (720)	0.106 (710)
FSM	0.216 (0.411)	0.229 (0.420)	0.219 (0.414)	0.232 (0.423)	0.236 (0.425)	0.527 (2203)	0.665 (1498)	0.587 (1486)	0.779 (692)	0.8 (717)	0.952 (705)
FSP GLD	0.543 (0.498)	0.482 (0.500)	0.482 (0.500)	0.500 (0.501)	0.466 (0.500)	0.010 (2210)	0.057 (1505)	0.27 (1492)	0.605 (693)	0.633 (718)	0.381 (705)
Panel B. Baseline test											
Std PIRA	0.0328 (1.000)	-0.0513 (0.998)	-0.0510 (1.019)	-0.0412 (0.959)	-0.0609 (1.015)	0.233 (2160)	0.230 (1464)	0.155 (1459)	0.661 (677)	0.710 (701)	0.923 (696)

Note. The sample for variables in Panel A includes all available observations in the National Pupil Dataset/survey records. The sample for the variable in Panel B includes all students sitting the baseline PIRA test. The variable in Panel B is standardised using the mean and standard deviation of all available observations at baseline. 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. PIRA is the progress in Reading Assessment test, our primary outcome. Standard deviations are in parentheses in columns 1-5 and the available observations for the respective samples are in parentheses in columns 6-11. P-values are calculated using pairing fixed effects (columns 6-8) and school fixed effects (columns 9-11). Standard errors are clustered at the unit of randomisation: i.e., at the school level in columns 6-8, and at the student level in the within school comparisons (i.e. robust standard errors are used in columns 9-11).

Table 5: Intention to Treat effects: Main results

	<i>All students</i>		<i>Excluding control students in treated schools</i>	
	(1)	(2)	(3)	(4)
A. Outcome: PIRA test at <i>endline</i>				
ICT	0.144 (0.087)	0.179** (0.079)	0.150 (0.090)	0.186** (0.081)
NONICT	0.246*** (0.082)	0.272*** (0.075)	0.259*** (0.083)	0.284*** (0.076)
CT	0.116 (0.082)	0.167** (0.074)		
Students	1901	1884	1591	1576
Schools			48	
P value: ICT=NONICT	0.104	0.102	0.086	0.092
P value: ICT=CT	0.579	0.821		
P value: NONICT=CT	0.017	0.039		
B. Outcome: PIRA test at <i>endline + 1</i>				
ICT	0.053 (0.073)	0.077 (0.072)	0.055 (0.075)	0.078 (0.073)
NONICT	0.072 (0.084)	0.094 (0.079)	0.081 (0.086)	0.101 (0.082)
CT	-0.021 (0.078)	0.015 (0.073)		
Students	1799	1785	1501	1488
Schools			48	
P value: ICT=NONICT	0.752	0.789	0.650	0.703
P value: ICT=CT	0.16	0.271		
P value: NONICT=CT	0.113	0.156		
C. Outcome: Key Stage 1 Reading at <i>endline + 1</i> (at or above the expected reading level)				
ICT	0.008 (0.025)	0.019 (0.025)	0.008 (0.024)	0.018 (0.025)
NONICT	0.048* (0.027)	0.055** (0.027)	0.048* (0.028)	0.055* (0.028)
CT	-0.021 (0.024)	-0.006 (0.025)		
Students	2129	2111	1770	1756
Schools			50	
P value: ICT=NONICT	0.163	0.146	0.160	0.148
P value: ICT=CT	0.217	0.335		
P value: NONICT=CT	0.001	0.007		
Mean outcome in control schools	0.739	0.741	0.739	0.741
<i>Controls variables:</i>				
Baseline PIRA test	✓	✓	✓	✓
Gender, FSM, FSP GLD		✓		✓

Note. Outcome variables: PIRA test at *endline* is the standardised score of the PIRA test taken at the end of treatment. PIRA test at *endline + 1* is the standardised score of the PIRA test taken a year after the end of treatment. KS1 reading at *endline + 1* is a dummy variable that equals 1 if the student is at or above the expected reading level at the end of Key Stage 1. ICT and NONICT are the treatment dummies. CT is a dummy that equals 1 for pupils in the control group of treatment schools. All available students are used in columns 1 and 2. In columns 3 and 4, students that were in the control group of treated schools are excluded. All regressions control for randomisation pair dummies. 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. Standard errors are clustered at the school level.

Table 6: Intention to Treat effects: Binary versus Continuous outcome measures

	PIRA endline (1)	PIRA endline+1 (2)	KS1 read endline+1 (3)
A. Binary outcome variables			
ICT	0.070* (0.038)	0.037 (0.038)	0.018 (0.025)
NONICT	0.127*** (0.040)	0.046 (0.036)	0.055* (0.028)
Std PIRA baseline	0.210*** (0.013)	0.191*** (0.016)	0.160*** (0.015)
Female	-0.042* (0.024)	-0.006 (0.020)	-0.037* (0.021)
FSM	-0.046 (0.029)	-0.043 (0.034)	-0.071** (0.028)
FSP GLD	0.223*** (0.030)	0.148*** (0.034)	0.224*** (0.027)
P value: ICT=NONICT	0.152	0.801	0.148
Mean outcome in control schools	0.453	0.535	0.741
B. Continuous outcome variables			
ICT	0.186** (0.081)	0.078 (0.073)	0.042 (0.061)
NONICT	0.284*** (0.076)	0.101 (0.082)	0.087 (0.070)
Std PIRA baseline	0.478*** (0.024)	0.468*** (0.028)	0.472*** (0.031)
Female	-0.044 (0.039)	0.011 (0.045)	-0.055 (0.043)
FSM	-0.161*** (0.056)	-0.188*** (0.053)	-0.184*** (0.049)
FSP GLD	0.517*** (0.052)	0.367*** (0.068)	0.457*** (0.062)
P value: ICT=NONICT	0.092	0.703	0.360
Students	1576	1488	1756
Schools	48	48	50

Note. Binary outcome variables: PIRA dummy: equals 1 if the student has a PIRA endline score equal or bigger than the mean PIRA endline score observed for students in control schools working at the KS1 expected reading level. PIRA+1 dummy: equals 1 if the student has a PIRA endline+1 score equal or bigger than the mean PIRA endline+1 score observed for students in control schools working at the KS1 expected reading level. KS1 reading at endline + 1 is a dummy variable that equals 1 if the student is at or above the expected reading level at the end of Key Stage 1. ICT and NONICT are the treatment dummies. *Continuous outcome variables:* PIRA test at endline is the standardised score of the PIRA test taken at the end of treatment. PIRA test at endline + 1 is the standardised score of the PIRA test taken a year after the end of treatment. Continuous KS1 endline+1 is the standardised score of the KS1 read variable transformed into a continuous measure (please refer to the main text for more details on this transformation). ICT and NONICT are the treatment dummies. We exclude from the regressions those students that were in the control group of treated schools. All regressions control for randomisation pair dummies. 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. Standard errors are clustered at the school level.

Table 7: Results for Other Subjects, one year later

	Key Stage 1 Reading (1)	Key Stage 1 Writing (2)	Key Stage 1 Maths (3)
ICT	0.018 (0.025)	0.042 (0.033)	0.004 (0.032)
NONICT	0.055* (0.028)	0.082** (0.036)	0.048 (0.032)
P value: ICT=NONICT	0.148	0.050	0.035
Control mean	0.741	0.625	0.715
Students	1756		
Schools	50		

Note. Outcome variables: Key Stage 1 Reading(Writing/math) is a dummy variable that equals 1 if the student is at or above the expected reading (writing/math) level at the end of Key Stage 1. ICT and NONICT are the treatment dummies. The control group is formed by control students in control schools only. All regressions control for FSM and female dummy, FSP GLD, standardised baseline PIRA tests, and the randomisation pairs. Standard errors clustered at the school level.

Table 8: Distributional Effects - Reading

	0.1Q (1)	0.25Q (2)	0.50Q (3)	0.75Q (4)	0.90Q (5)
A. Outcome variables defined at endline (i.e., using PIRA at endline)					
ICT	0.038 (0.111)	0.214** (0.107)	0.184* (0.094)	0.199** (0.087)	0.188 (0.147)
NONICT	0.188** (0.085)	0.220** (0.098)	0.228** (0.089)	0.273*** (0.101)	0.462*** (0.111)
Students	1576	1576	1576	1576	1576
Schools	48	48	48	48	48
P-value Parente-Santos Silva test for intra-cluster correlation	0.009	0.000	0.000	0.003	0.602
A. Outcome variables defined at endline+1 (i.e., using PIRA at endline+1)					
ICT	0.097** (0.049)	0.124* (0.073)	0.047 (0.088)	0.079 (0.051)	0.001 (0.028)
NONICT	0.081 (0.055)	0.080 (0.089)	0.128 (0.092)	0.139* (0.077)	0.073 (0.098)
Students	1488	1488	1488	1488	1488
Schools	48	48	48	48	48
P-value Parente-Santos Silva test	0.639	0.093	0.001	0.246	0.269

Note. Outcome variables: PIRA test at endline is the standardised score of the PIRA test taken at the end of treatment. PIRA test at endline +1 is the standardised score of the PIRA test taken a year after the end of treatment. ICT and NONICT are the treatment dummies. All regressions control for FSM and female dummy, FSP GLD, standardised baseline PIRA tests, and the randomisation pairs. We exclude from the regressions those students that were in the control group of treated schools. Standard errors clustered at the school level, except when the Parente-Santos Silva test for intra-cluster correlation cannot reject the null of no intra-cluster correlation. In the latter case, we show robust standard errors.

Table 9: Heterogeneous effects

Outcome: Standardised PIRA	PIRA at endline (1)	PIRA at endline +1 (2)	KS1 reading at endline +1 (3)
A. FSM			
ICT*FSM	0.464*** (0.138)	0.216* (0.117)	0.044 (0.060)
ICT*NOFSM	0.117 (0.081)	0.045 (0.078)	0.011 (0.025)
NONICT*FSM	0.505*** (0.102)	0.126 (0.091)	0.094** (0.045)
NONICT*NOFSM	0.219*** (0.080)	0.095 (0.096)	0.044 (0.034)
FSM	-0.275*** (0.077)	-0.225*** (0.067)	-0.087** (0.040)
Ho: ICT (FSM-NOFSM)=0	0.006	0.19	0.722
Ho: NONICT (FSM-NOFSM)=0	0.000	0.317	0.063
B. Gender			
ICT*Female	0.211** (0.092)	0.018 (0.083)	0.014 (0.026)
ICT*Male	0.159* (0.092)	0.146 (0.095)	0.023 (0.040)
NONICT*Female	0.345*** (0.094)	0.097 (0.087)	0.093** (0.035)
NONICT*Male	0.219** (0.092)	0.105 (0.124)	0.013 (0.043)
Female	-0.075 (0.045)	0.036 (0.049)	-0.045 (0.032)
Ho: ICT (Fem-Male)=0	0.561	0.216	0.831
Ho: NONICT (Fem-Male)=0	0.243	0.957	0.156
C. Above/below median prior attainment (based on PIRA baseline test)			
ICT*(> median)	0.078 (0.077)	0.046 (0.087)	0.053** (0.025)
ICT*(< median)	0.288*** (0.105)	0.108 (0.088)	-0.015 (0.040)
NONICT*(> median)	0.272*** (0.075)	0.127 (0.089)	0.066** (0.029)
NONICT*(< median)	0.301** (0.117)	0.076 (0.104)	0.040 (0.048)
Std Pira baseline	0.498*** (0.026)	0.469*** (0.034)	0.153*** (0.017)
Ho: ICT (Above-Below)=0	0.034	0.523	0.153
Ho: NONICT (Above-Below)=0	0.816	0.609	0.644

Note. Number of students (schools) in columns 1, 2 and 3, respectively is: 1576 (48), 1756 (50) and 1488 (48). Outcome variables: PIRA at endline is the standardised score of the PIRA test taken at the end of treatment. PIRA at endline +1 is the standardised score of the PIRA test taken a year after the end of treatment. KS1 reading at endline + 1 is a dummy variable that equals 1 if the student is at or above the expected reading level at the end of Key Stage 1. The control group is formed by control students in control schools only (i.e., control students in treatment schools are excluded from the regressions). All regressions control for FSM and female dummy, FSP GLD, standardised baseline PIRA tests, and the randomisation pairs. Standard errors are clustered at the school level.

Table 10: A comparison between treatment and control schools, post-intervention

	Control Schools	Treatment schools	Difference in Means (P-Values & Observations)
	(1)	(2)	(3)
Hours of literacy instruction per week	7.372 (1.697)	8.049 (2.790)	0.39 (48)
Computers are used to support literacy teaching	0.750 (0.442)	0.726 (0.456)	0.863 (48)
Smartboards are used to support literacy teaching	0.967 (0.183)	0.964 (0.190)	0.962 (48)
Projectors are used to support literacy teaching	0.467 (0.509)	0.393 (0.500)	0.651 (48)
Tablets are used to support literacy teaching	0.628 (0.493)	0.750 (0.443)	0.413 (48)

Note. The information in this table comes from data collected via surveys at endline (i.e., end of Year 1). Columns 1 and 2 show means (first row) and standard deviations (in parentheses). P-values are calculated using robust standard errors (column 3). [Results are very similar when we also include randomisation-pairing dummies to calculate p-values; or when we calculate them using standard errors clustered at the school level. Due to the low number of observations and clusters, we show the results without including pairing dummies and without clustering standard errors at the school level]. Observations have a weight of 1 if there is only one Year 1 teacher replying to the questionnaire per school; and 0.5 when there are two Year 1 teachers replying to the questionnaire per school.

Table 11: Compliance according to intervention type

	ICT	Non-ICT	Difference in Means (P-Values & Observations)
	(1)	(2)	(3)
Score based on daily record keeping by the TA (1 to 10)	8.130 (2.916)	9.478 (1.229)	0.047 (46)
Score based on TA use of the levels (1 to 10)	6.457 (2.147)	7.022 (1.880)	0.347 (46)
Number of weeks the TA kept records (maximum=20)	18.28 (3.304)	19.42 (1.865)	0.158 (46)

Note. The information in this table comes from data collected by the implementation team. Researchers at the implementation team gave scores for daily record keeping and use of levels at the end of the implementation. Columns 1 and 2 show means (first row) and standard deviations (in parentheses). P-values are calculated using robust standard errors (column 3). Results are very similar when we also include school fixed effects or when we cluster the standard errors at the school level. Due to the low number of observations and clusters, and the fact that in the second panel we miss information for some of the TAs in some categories, we show the results without including school fixed effects and without clustering standard errors at the school level. There is only one case with two teaching assistants per group in this data. For this particular case, we consider the average score among the two teaching assistants (all the other cases have 1 observation per teaching assistant or group of teaching assistants).

Table 12. Group size and composition by treatment condition

	ICT	Non-ICT	Difference in Means (P-Values & Observations)
	(1)	(2)	(3)
Average group size	3.597 (0.520)	3.69 (0.667)	0.35 (148)
<i>Within group standard deviations for:</i>	ICT	Non-ICT	<i>Difference in SD by group and treatment conditions (P-values)</i>
FSM	0.316	0.34	0.59
Female	0.425	0.426	0.988
Standardised baseline PIRA	0.592	0.566	0.649

Note. P-values calculated by regressing the average group size in each small group (or the SD for each small group for the variables FSM, Female and Standardised baseline PIRA) on a dummy for the NON-ICT group, with robust standard errors. Results are very similar when we also include school fixed effects or when we cluster the standard errors at the school level. Due to the low number of observations and clusters, we show the results without including school fixed effects and without clustering standard errors at the school level. The number of observations in these regressions is 148, which corresponds to the number of small groups formed by the teaching assistants overall (i.e., in both ICT and NON-ICT conditions). There is no information on the groups for the 3 schools in the treatment group that dropped out immediately after randomisation.

APPENDIX

Table A1: Attrition

	Control Schools (1)	Treatment schools (2)	ICT (3)	Non-ICT (4)	Control in Treatment schools (5)
Students initially allocated to...	1158	1083	360	350	373
<i>Fraction students in each group with....</i>					
Missing baseline PIRA	0.030	0.024	0.033	0.020	0.019
Missing endline PIRA	0.047	0.153	0.150	0.171	0.139
Missing endline Key Stage 1 Reading at t+1	0.020	0.028	0.028	0.034	0.021
Missing endline PIRA at t+1	0.108	0.189	0.186	0.211	0.172

Note. Key Stage 1 data is available for all schools that were included in the randomisation. Five schools in the treatment group dropped out after randomisation (3 right after randomisation, 2 during the intervention). Post-intervention tests right at the end of the intervention and at t+1 were conducted in all schools but 1.

Table A2. IV estimates

A. Outcome:	PIRA at endline	PIRA at endline+1	KS1 read at endline+1
	(1)	(2)	(3)
ICT	0.224** (0.094)	0.094 (0.084)	0.023 (0.031)
NONICT	0.343*** (0.090)	0.121 (0.094)	0.072** (0.035)
Students	1576	1488	1756
Schools	48	48	50
B. Main coefficient in ICT first stage			
Randomised to ICT	0.842*** (0.066)	0.842*** (0.066)	0.760*** (0.084)
F-test of excluded instruments	104.340	103.380	60.900
C. Main coefficient in NON-ICT first stage			
Randomised to NONICT	0.830*** 0.069	0.834*** (0.067)	0.753*** (0.040)
F-test of excluded instruments	89.520	93.930	57.380

Note. Outcome variables: PIRA at endline is the standardised score of the PIRA test taken at the end of treatment. PIRA at endline +1 is the standardised score of the PIRA test taken a year after the end of treatment. KS1 reading at endline + 1 is a dummy variable that equals 1 if the student is at or above the expected reading level at the end of Key Stage 1. ICT and NONICT are the treatment dummies. The control group is formed by control students in control schools only (i.e., control students in treatment schools are excluded from the regressions). All regressions control for FSM and female dummy, FSP GLD, standardised baseline PIRA tests, and the randomisation pairs. Standard errors are clustered at the school level.

Note on Methodology

The first stages for whether students are in the final ICT or final Non-ICT treatments are as follows:

$$ICT\ Final_{ist} = \gamma_1 ICT_{ist} + \gamma_2 NonICT_{ist} + \gamma_3 Y_{ist-1} + \gamma_4 X_{ist-1} + \rho_s + \varepsilon_{ist} \quad (A1)$$

$$NonICT\ Final_{ist} = \pi_1 ICT_{ist} + \pi_2 NonICT_{ist} + \pi_3 Y_{ist-1} + \pi_4 X_{ist-1} + \rho_s + \varepsilon_{ist} \quad (A2)$$

Where $ICT\ Final_{ist}$ ($NonICT\ Final_{ist}$) is a dummy variable equal to 1 if students received the *complete* 20-week ICT (Non-ICT) intervention, and equal to 0 otherwise. The second stage equation is then given by:

$$Y_{ist} = \theta_1 ICT\ Final_{ist} + \theta_2 NonICT\ Final_{ist} + \theta_3 Y_{ist-1} + \theta_4 X_{ist-1} + \rho_s + \varepsilon_{ist} \quad (A3)$$

We estimate (A3) by two stage least squares, using the initial random allocations, ICT_{ist} and $NonICT_{ist}$, respectively, as instruments for $ICT\ Final_{ist}$ and $NonICT\ Final_{ist}$, and the other variables as instruments for themselves.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

- | | | |
|------|--|---|
| 1561 | Daniel J. Graham
Stephen Gibbons | Quantifying Wide Economic Impacts of Agglomeration for Transport Appraisal: Existing Evidence and Future Directions |
| 1560 | Daniel S. Jacks
Dennis Novy | Market Potential and Global Growth over the Long Twentieth Century |
| 1559 | Alan Manning
Paolo Masella | Diffusion of Social Values through the Lens of US Newspapers |
| 1558 | Jan David Bakker
Stephan Maurer
Jörn-Steffen Pischke
Ferdinand Rauch | Of Mice and Merchants: Trade and Growth in the Iron Age |
| 1557 | Giuseppe Berlingieri
Sara Calligaris
Chiara Criscuolo | The Productivity-Wage Premium: Does Size Still Matter in a Service Economy? |
| 1556 | Christian A.L. Hilber
Olivier Schöni | The Economic Impacts of Constraining Second Home Investments |
| 1555 | Filippo Di Mauro
Fadi Hassan
Gianmarco I.P. Ottaviano | Financial Markets and the Allocation of Capital: The Role of Productivity |
| 1554 | Keith Head
Yao Amber Li
Asier Minondo | Geography, Ties and Knowledge Flows: Evidence from Citations in Mathematics |
| 1553 | Nicholas Bloom
Kalina Manova
Stephen Teng Sun
John Van Reenen
Zhihong Yu | Managing Trade: Evidence from China and the US |

1552	Gabriel M. Ahlfeldt Wolfgang Maennig Steffen Q. Mueller	The Generation Gap in Direct Democracy
1551	Keith Head Tierry Mayer	Brands in Motion: How Frictions Shape Multinational Production
1550	Natalie Chen Dennis Novy	Currency Unions, Trade and Heterogeneity
1549	Felipe Carozzi	The Role of Demand in Land Re-Development
1548	Nick Bloom Brian Lucking John Van Reenen	Have R&D Spillovers Changed?
1547	Laura Alfaro Nick Bloom Paola Conconi Harald Fadinger Patrick Legros Andrew F. Newman Raffaella Sadun John Van Reenen	Come Together: Firm Boundaries and Delegation
1546	Jörg Kalbfuß Reto Odermatt Alois Stutzer	Medical Marijuana Laws and Mental Health in the United States
1545	Maria Molina-Domene	Specialization Matters in the Firm Size-Wage Gap
1544	Jonathan Colmer	Weather, Labor Reallocation and Industrial Production: Evidence from India
1543	Andrew B. Bernard Esther Ann Bøler Swati Dhingra	Firm-to-Firm Connections in Colombian Imports

The Centre for Economic Performance Publications Unit

Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk

Website: <http://cep.lse.ac.uk> Twitter: @CEP_LSE