PRIFYSGOL BANGOR / BANGOR UNIVERSITY

Effects of implementation support on children’s reading outcomes following an online early reading programme: a cluster-randomised controlled trial

Roberts-Tyler, Emily; Roberts, Sarah; Watkins, Richard; Hughes, Carl; Hastings, Richard; Gillespie, David

British Journal of Educational Technology

DOI:
10.1111/bjet.13312

E-pub ahead of print: 24/03/2023

Peer reviewed version

Cyswllt i'r cyhoeddiad / Link to publication

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Roberts-Tyler, E., Roberts, S., Watkins, R., Hughes, C., Hastings, R., & Gillespie, D. (2023). Effects of implementation support on children’s reading outcomes following an online early reading programme: a cluster-randomised controlled trial. *British Journal of Educational Technology*. Advance online publication. https://doi.org/10.1111/bjet.13312

Hawliau Cyffredinol / General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

20. Mar. 2024

Effects of implementation support on children's reading outcomes following an online early reading programme: a cluster-randomised controlled trial

Authors & affiliations:

Emily J. Roberts-Tyler [a], Sarah E. Roberts [a], Richard Watkins[b],
J. Carl Hughes [a] , Richard P. Hastings[c], and David Gillespie[d]

[a] Collaborative Institute for Education Research, Evidence and Impact (CIEREI), School of Educational Sciences, Bangor, Wales, UK
[b]GwE, Regional School Improvement Service for North Wales
[c]Centre for Educational Development Appraisal and Research, University of Warwick
[d]Centre for Trials Research, Cardiff University

Corresponding Author: Emily J Roberts-Tyler, School of Educational Sciences, Bangor University, Normal Site, Bangor, Gwynedd, Wales, LL57 2PZ (e-mail: e.j.tyler@bangor.ac.uk).

Emily J. Roberts-Tyler is a Lecturer and applied Researcher in the School of Educational Sciences at Bangor University, and an active member of the Collaborative Institute for Education Research, Evidence and Impact (CIEREI). Her research interests are predominantly in the field of effective reading instruction, specifically in relation to improving reading outcomes for children with intellectual and developmental disabilities, including children who are non-verbal. Emily also

researches the use of fluency-based instruction to improve academic skills with diverse learners, and the effects and challenges of implementation fidelity in school settings.

Sarah E. Roberts is a doctoral researcher in the School of Educational Sciences at Bangor University, and an active primary school teacher. Her research interests are focused on supporting schools and parents to implement effective interventions to enhance learner outcomes.

Richard Watkins is the research and evaluation lead for the Regional School Effectiveness and Improvement Service for North Wales (GwE) and co-director of the Collaborative Institute for Education Research, Evidence and Impact, Bangor University (CIEREI). His research interests include science education, and the application of evidence-based teaching strategies in schools.

J. Carl Hughes is a Professor of Psychology, Director of the Collaborative Institute for Education Research, Evidence and Impact (CIEREI), and Head of the School of Educational Sciences at Bangor University. His research interests include evidence-based educational interventions, health and well-being in schools, and the application of behavioural psychology to enhance learning environments.

Richard P. Hastings is a Professor of Education and Psychology, and Cerebra Chair of Family Research, and the Director of the Centre for Educational Development Appraisal and Research at the University of Warwick. His research focuses on the fields of special educational needs and disability across the lifespan, and on the evaluation of psychological interventions in educational contexts.

David Gillespie is a Senior Research Fellow at the Centre for Trials Research at Cardiff University. His background and much of his current work relates to Medical Statistics, with broader methodological interests including trials and general research methodology, and causal modelling.

Practitioner notes

What is already known about this topic

- Well-designed computer or app-based instruction has a number of potential benefits (e.g., increasing accessibility and feasibility of high-quality instruction, reducing time and resources required for training expert delivery, saving instructional time).
- Implementation can still affect outcomes when using educational technology, and without follow-up support after training, implementation of educational interventions is often poor and outcomes reduced.
- The extent to which this is the case when the core element of an intervention is computer or app-delivered is not yet clear.

What this paper adds

- We found that providing implementation support for teachers and teaching assistants delivering Headsprout Early Reading (HER; an early reading programme accessible via a computer or an app) did not affect the reading outcomes of learners.
- We also found the implementation support did not affect delivery of the core, app-delivered element of the programme.
- However, there were notable differences in implementation of other aspects of the programme, particularly in relation to the role of the teacher or educational practitioner in managing the interplay between the app-based learning and teacher intervention for learners who require further support.

Implications for practice and/or policy

- These findings have implications for providing access to high quality instruction in early reading skills at scale, with minimal training.
- More broadly, the current study suggests that well-designed computer or app-based instruction can yield positive outcomes with minimal implementation support and training.
- However, the findings of this study identify some potential risk of an over-reliance on technology to facilitate the learning of all learners accessing the programme.
- Further research is required to ensure the interplay between learners' app-based learning and teacher intervention functions as intended to provide additional support for those who need it.

Abstract

Well-designed computer or app-based instruction has a number of potential benefits (e.g., increasing accessibility and feasibility of high-quality instruction, reducing time and resources required for training expert delivery, saving instructional time). However, variation in implementation can still affect outcomes when using educational technology. Research generally suggests that without follow-up support after training, implementation of educational interventions is often poor and outcomes reduced. However, the extent to which this is the case when the core element of an intervention is computer or app-delivered is not yet clear. This study investigated the effects of providing ongoing implementation support for Headsprout Early Reading (HER, an early reading programme accessible via a computer or an app), to determine whether such support leads to better outcomes. Twenty-two primary schools (269 learners) participated in a cluster-randomised controlled trial. Eleven schools received initial training followed by ongoing support across the school year, whereas the other 11 schools received initial training and technical support only. Pre and post-measures of reading skills were conducted using the York Assessment of Reading for Comprehension. We found no effect of implementation support on outcomes, and no effect of implementation support on delivery of the core element of HER. However, there were some effects of implementation support on the implementation of other HER elements relating to the responsiveness of educators to learners' learning within HER. These findings have implications for providing access to high quality online instruction in early reading skills at scale, with minimal training. More broadly, the current study suggests that well-designed computer or app-based instruction can yield positive outcomes with minimal implementation support and training. However, further research is required to ensure the interplay between learners' app-based learning and teacher intervention functions as intended to provide additional support for those who need it.

Considerable evidence supports the use of explicit, systematic phonics instruction for beginning readers (e.g., Rose, 2006; Rose, 2009; National Reading Panel, 2000; Coyne et al., 2004). Despite some inconsistent effects of computer-based instruction (Higgins, Xiao & Katsipataki, 2012), research generally indicates positive effects of computer-based phonics instruction on reading skills (e.g., National Reading Panel, 2000; Blok, Oostdam, Otter & Overmaat, 2002; Huffstetter et al., 2010; Twyman et al., 2011; Storey, McDowell & Leslie, 2017; Cheung & Slavin, 2012; Abrami, Lysenko & Borokhovski, 2020).

A key challenge in the evaluation of computer-based instruction and educational technologies is that the component parts may differ considerably in terms of the quality and appropriateness of the interface and aesthetic, the content and specific skills targeted, and the instructional design (Bishop & Santoro, 2006), bringing into question the validity and usefulness of combined comparisons of the effects. However, well-designed computer-based instruction has a number of potential benefits, including increasing the accessibility and feasibility of high-quality instruction, reducing the time and resources required for training expert delivery, and saving instructional time (Kulik & Kulik, 1991; Pennington, 2010). If they can be harnessed, these benefits have considerable implications for greater equity in education for disadvantaged learners and those with special educational needs internationally. Such learners have been disproportionately impacted by the Covid-19 pandemic (Twist, Jones, & Treleaven, 2022; Theis et al., 2021). However, there is now a significant opportunity to learn from the use and experiences of educational technology during the pandemic, and to focus our research efforts on how educational technology can be used to help address locally and globally relevant issues faced in practice (Kerres & Buchner, 2022).

Implementation conditions and outcomes of educational interventions

Despite evidence of the effects of many educational interventions, less experimental research focuses on the implementation conditions that affect the outcomes experienced by learners.

An important consideration in achieving and maintaining good implementation of evidence-based practices is the extent to which follow-up support is provided (Kretlow & Bartholomew, 2010). Training in evidence-based practices is necessary to disseminate new practices, and effective training to enhance educators' self-efficacy is a factor found to affect implementation and learner outcomes (Durlak & Dupre, 2008). However, when subsequently implementing new practices in situ, teachers often report challenges relating to having sufficient depth of understanding of the new practice, remembering how to implement the practice effectively, and the reality of using the new practice alongside many other practices (Klingner, Vaughn, Hughes & Arguelles, 1999). Research indicates that training *and* some form of performance feedback or coaching significantly affects the implementation of educational practices, improving the likelihood of higher levels of implementation fidelity and leading to better outcomes (e.g., Buzhardt, Greenwood, Abbott & Tapia, 2006; Yoon, Duncan, Lee, Scarloss & Shapley, 2007; Klingner, Vaughn, Hughes & Arguelles, 1999; McCollum, Hemmeter & Hsieh, 2011; Elish-Piper & L'Allier, 2011; Neufeld & Roper, 2003; Schechter et al., 2017; Owen et al., 2021) and also sustained implementation (e.g., Kretlow, Wood & Cooke, 2011; and Kretlow, Cooke & Wood, 2012; Klingner, Vaughn, Hughes & Arguelles, 1999).

In a review of 13 studies, Kretlow and Bartholomew (2010) investigated the impact of coaching teachers (including both trainees and qualified teachers) on the implementation of evidence-based practices, concluding that there is strong evidence for the effectiveness of coaching approaches in improving implementation. Of the studies reporting student outcomes, these were found to be positively influenced by coaching and improved implementation. Most studies included an initial training session, followed by either supervisory coaching (whereby teachers were observed and provided with face-to-face feedback), or side-by-side coaching (whereby teachers were observed, provided with face-to-face feedback, and asked to observe a model session delivered by the coach), or a combination of both methods. Both methods also typically included follow-up observations and further feedback. Such coaching allows for individualized follow-up support to be provided, as well as opportunities to model correct implementation and to reinforce appropriate implementation in situ (Joyce & Showers, 1995; Blakely, 2001).

As an example, Owen et al., (2021) investigated the effects of training plus ongoing implementation support on the fluency outcomes of learners receiving a maths fluency intervention. In a cluster-randomised controlled trial, 64 schools were randomly allocated to either receive initial training plus ongoing support (n=33; 294 learners) or initial training and no ongoing support (n=31; 281 learners). Ongoing support included three 1-hour visits, which were individualised depending on the support needs of the school, but typically included modelling, direct feedback, and troubleshooting implementation issues related to the learner or school. No support included initial training and email access to queries not related to implementation of the intervention. Results indicated that learners in the ongoing support arm of the trial demonstrated greater gains in fluency in the targeted maths skills.

Implementation of educational technology applications

Even though computer and app-based delivery methods might be considered to be all that is needed for intervention, the way in which a programme is implemented has still been found to affect outcomes when using educational technology (Savage et al., 2010; Savage et al., 2013; Abrami, Lysenko & Borokhovski, 2020). In a meta-analysis of 84 studies investigating the effects of educational technology applications on reading outcomes for pupils (age 4-18years), Cheung and Slavin (2012) found significant differences in reading outcomes according to levels of implementation as reported within each study, with no effects seen when implementation levels were rated as low, and significant positive effects seen when implementation levels were rated as medium or high. Their findings also indicated some evidence that programmes including a combination of technology-based and non-technology-based components might yield better outcomes, suggesting the role of the educator remains important in the effective use of educational technology.

Further, Outhwaite, Gulliford and Pitchford (2019) found that having a well-established routine (e.g., consistent timetabling of the intervention, dedicated classroom space, and a dedicated staff member responsible for implementation) was strongly correlated with learning outcomes following a 12-week maths app intervention, with such routine enabling more progress over time, and accounting for 41% of the variance in learning outcomes. Convergent findings were reported by Gulliford, Walton, Allison and Pitchford (2021) in a study combining qualitative data from two evaluations of a maths app intervention. Both the features of the app (e.g., accessibility, instructional demand, task structure, curriculum links) and child factors (e.g., language proficiency, attention, motivaton) were reported to be important factors in the outcomes for learners using the app. However, learning environment factors (e.g., the physical learning environment and implementation) and facilitator factors (e.g., the provision of pedagogical facilitation and further support) were also considered crucial. Gulliford et al., (2021) therefore also highlight the importance of the educator's role in ensuring an appropriate level of challenge and support to optimise the success of educational technology.

Headsprout Early Reading

Headsprout Early Reading is a computer-based, systematic phonics programme, including instruction in phonemic awareness, print awareness, phonics, segmenting and blending, and reading with comprehension (Layng, Twyman & Stikeleather, 2003; https://www.headsprout.com). HER utilises highly effective instructional principles (employed in Direct Instruction; Schieffer et al., 2002; Kinder, Kubina & Marchand-Martella., 2005), including teaching consistent elements before exceptions, basic strategies to mastery, and easy skills prior to more difficult skills. HER begins with highly stable phonetic elements: the first 33 elements introduced are regular in more than 85% of the words in which they appear. Fluency practice then allows for mastery of decoding strategies before introducing less stable elements. HER also employs sophisticated adaptive learning technology—instruction adapts to individual responses, providing additional instruction or practice, and high levels of response and feedback. HER includes the key instructional features reportedly found in effective computer-based instruction, including practice opportunities, self-correction and immediate corrective feedback, teacher-directed instruction, and contingencies for enhancing learner motivation and engagement, including those within the programme (such as frequent positive feedback, presentation of short cartoon sequences between tasks, visual progress maps, and additional reward strategies that can be implemented), and contingencies external to the programme (e.g., being able to read a story using newly acquired skills after only 5 lessons) (Kim, McKenna & Park, 2017; Layng et al., 2003).

In addition to empirically informed development (Layng et al., 2003), there is evidence suggesting HER can help improve reading skills, including typically developing children (Tyler et al., 2015a, Huffstetter et al., 2010; Twyman et al., 2011; Storey, McDowell & Leslie, 2017) children with ADHD (Clarfield & Stoner, 2005), and children with intellectual disabilities and/or autism (Roberts-Tyler et al., 2019; Grindle et al., 2013, 2021; Tyler et al., 2015b; Whitcomb et al., 2011).

The current literature generally suggests that without follow-up support after training, implementation of educational interventions is often poor and outcomes reduced (e.g., Mortenson & Witt, 1998; Kretlow & Bartholomew, 2010; Elish-Piper & L'Allier, 2010). However, there are relatively few studies investigating the extent to which this is the case for programmes in which the core element of the intervention is automated and provided via a computer or app-based programme using experimental designs. Further, there is a paucity of research employing robust designs to directly investigate the effects of implementation support on outcomes of educational interventions more generally.

The main aim of the current study was to examine the effects of providing ongoing implementation support for Headsprout Early Reading compared to initial training alone on reading outcomes for children. The associated main research question was: Does ongoing implementation support lead to better reading outcomes for learners enrolled in an online reading programme compared to initial training alone?

The exploratory aims of the study were to investigate: a) whether ongoing implementation support leads to better implementation of HER; b) whether there were any associations between process measures and our primary outcome measure.

Method

Schools and learners

School recruitment.

Twenty-two primary schools across four counties in North Wales were recruited to participate in this study. Both English medium and Welsh medium schools were invited to participate; 15 participating schools were English medium, and 7 were Welsh medium. In collaboration with GwE (the regional school improvement service), all primary schools in North Wales were provided with information about the study, and information sessions were arranged for schools who registered an interest. This provided an opportunity to access HER, discuss how it might work in their schools (e.g., how it might work alongside other reading programmes), and discuss the requirements of the project (e.g., direct cost, and information technology and staffing resources). None of the participating schools had previously used HER.

In addition to providing information sessions, the project was specifically designed to appeal to prospective participating schools, investigating some important pragmatic research questions. Considerations included the need to focus on national policy objectives, including effective use of school improvement funding and alignment with national priorities, such as the Literacy and Numeracy Framework (LNF) initiative and improving outcomes for disadvantaged pupils (Welsh Government, 2013, 2014, 2015). GwE has a central role in monitoring school's expenditure of the Welsh Government Pupil Development Grant (PDG; a core funding arm delegated directly to schools based on the number of pupils eligible for free school meals, eFSM). Through this important policy lever, we were able to align this project with school's PDG funding. This, together with the GwE's ability to access and communicate with large numbers of schools, enabled very rapid take-up of the project and a route to funding high-quality research in a '*collaborative stakeholder funding model*'; we were able to fund this RCT with resources internal to the system through a model where each partner contributed part of the funding through strategic use of PDG funding from participating schools, other local funding, and matched funding. For further discussion of this funding and recruitment model, see Tyler, Watkins, Roberts, Hoerger, Hastings and Hughes (2019).

Learner selection and recruitment.

All schools were asked to identify up to 15 learners in either Year 1 or Year 2 (aged 4-7 years; English medium schools) or Year 3 or Year 4 (aged 7-9 years; Welsh medium schools). A choice of year groups was offered to ensure that smaller schools could put forward learners eligible for the study. Welsh medium schools selected older learners due to the common practice that English reading instruction begins in Year 3 following the initial teaching of reading in Welsh. Because we were interested in evaluating the implementation of the programme with beginning readers, it was a requirement that learners had minimal preexisting English reading skills. Therefore, schools were advised to target learners with a standard score of less than 85 on national English reading tests, and given a description of basic reading skills that would be permitted within the inclusion criteria. Schools were also encouraged to target learners eligible for free school meals (eFSM learners) where possible, as required by PDG funding. School data on learners' reading tests were not obtained by the researchers. However, 142 participants (53%) reached a standard score lower than 85 on the pre-test YARC Early Word Recognition measure (68 in the ISG, and 74 in the SSG), providing an indication of the reading skills of participants recruited to the study.

Once schools had identified learners, parental consent was obtained for their participation in the study. The HER placement test was then used to determine eligibility. Learners placing beyond episode 19 were excluded from the study, due to more advanced reading skills than were described in the inclusion criteria. The schools were still free to use the programme with these learners. However, reading assessments were not conducted for learners who did not meet the inclusion criteria, and these were not included in the study.

Design

A cluster-randomised controlled trial design was employed, with randomisation occurring at the level of the school. An a priori sample size calculation was informed by effect sizes from previous evaluations of HER (Tyler et al., 2015a/b). Assuming a standardized mean difference effect size of .50, power of 80%, and 5% alpha, the total sample size required would have been 64 children. The design effect allowing for clustering at the school level of ICC .20, and an estimated mean cluster size of 10 was calculated as 2.8. Thus, 179 learners would be required. Allowing loss to follow-up of 10% would mean a required sample size of 197 learners (approximately 20 schools).

Overall, 22 schools were randomly allocated to either the standard support group (SSG; 11 schools; 4 Welsh medium, and 7 English medium schools; *n* = 129 learners) or the implementation support group (ISG; 11 schools; 3 Welsh medium schools, and 8 English medium schools; n = 140 learners). Schools were allocated on a 1:1 basis at a single point in time in a single block, following baseline data collection for all included children. The randomization scheme was generated by using the website randomization.com ⟨http://www.randomization.com⟩. The researcher responsible for randomisation was blinded to any school identification or reading assessment information.

Measures

Reading outcomes.

All learners were assessed using the York Assessment of Reading for Comprehension (YARC; Snowling et al., 2010) Early Reading and Passage Reading for Primary. The Early Reading subtest comprises four sections: Letter Sound Knowledge (extended test administered), Early Word Recognition, Sound Isolation, and Sound Deletion. Each subtest provides a raw score, from which an Ability score, Standard score, Percentile rank and Age equivalent are calculated.

Some learners' scores fell below the range for which standardised scores were provided (See Table 1). For these learners, estimated standardised scores were calculated based on chronological age and reading age output. To calculate these estimates, chronological age was divided by reading age and multiplied by 100.

Interobserver agreement was calculated for 30% of pre and post-test assessments across outcome measures and was 90% for scoring both pre and post-test reading assessments. Any differences were resolved by listening back to the assessment recordings.

Monitoring implementation

Key aspects of HER implementation were measured for all schools, including frequency of episodes delivered (average episodes per week), percentage of episodes that were correctly repeated, whether benchmark assessments were conducted and scored online, and self-reported/checklist data on fluency-building. Table 5 provides a summary of the implementation monitoring data for each group. See Procedure for further explanation of these components.

Confidence following training.

Questionnaires were distributed after initial training to gage the implementation confidence of staff following the initial training. Staff were asked to report their level of confidence pertaining to various aspects of the programme, including episodes, stories, benchmark assessments, fluency building, and generally accessing the programme resources.

Procedure

Pre-randomisation.

Three undergraduate students and one research assistant were trained to conduct baseline assessments using the YARC. The assessments were carried out over a period of four weeks at the beginning of the school year. All assessments were recorded using Dictaphones to enable any queries or issues to be identified and resolved where possible, and to enable Interobserver agreement to be calculated for a sample of the assessments (see earlier).

Staff training.

Participating schools were required to send two members of staff to the initial training session, to ensure programme delivery did not rely on the presence of one person in the school with knowledge of the programme. Fifty-two members of staff across the 22 schools attended the training; 27 from the ISG schools, and 24 from the SSG schools. For 17 of the 22 schools at least one teacher attended the training (for 15 of those schools, at least one teaching assistant also attended, and for three of those schools the headteacher attended). For the remaining five schools, either two or three teaching assistants attended.

Four training sessions were conducted, one for each cluster of nearby schools to minimise travel distances for school staff. Each training session lasted three hours, and, following discussion with schools, was scheduled according to the most practical time for staff to attend. Key aspects of implementation were covered explicitly in this training session, including:

Episodes. Staff were trained how to access episodes, how to check episode accuracy, and how to repeat episodes. They were also informed of some important elements of the instructional design (e.g., how the phonetic elements are sequenced, and how error correction and feedback is built into the programme), the importance of achieving at least 90% accuracy or repeating that episode, and the importance of completing at least three sessions a week.

Stories. From episode 5, Sprout stories begin to appear after every episode. Staff were advised to listen learners read these stories and were shown how to print them and how learners can access them through the Book room.

Benchmarks. After every tenth episode, a benchmark story automatically appears to be conducted with the learner using the benchmark tracker and scores input online. The scoring grades were explained, and the Benchmark implementation sheet was highlighted. Members of staff were encouraged to record the learners carry out the benchmark assessments and informed they would receive microphones in order to do this.

Fluency-building. If learners achieved ‘needs practice’ on their benchmark assessment or achieve less than 90% accuracy after having repeated an episode, staff were instructed to conduct fluency building with them. Staff were shown where to access the fluency building materials and how to conduct the fluency building using the fluency building tracker.

Three of the four sessions were carried out in ICT suites, where staff were able to experience the programme first hand, including logging onto the programme, navigating the site, and conducting part of an episode. The remaining session was conducted in a standard classroom, with fewer ICT resources available. This led to some staff having less opportunity for hands on practice during the session.

Each school was given a training pack including a paper copy of the Headsprout training PowerPoint slides, Headsprout Progress map, key implementation prompt sheets, and fluency building tracker sheets.

Support models.

Following training teachers and completion of baseline assessments, schools were randomised to either the Standard Support Group (SSG; offering only technical programme support) or the Implementation Support Group (ISG; including regular school visits, ongoing technical support, and practical advice via phone and email).

Our ISG model was designed from a pragmatic perspective, based on our previous pilot work with this programme and with schools in this region, with a focus on a feasible delivery model given resources available, but also drawing on key features and functions of coaching support in the literature. For example, during individualised visits, observations were followed by face to face feedback and modelling of correct implementation (Kretlow & Bartholemew, 2010; Noell et al., 2005; Owen et al., 2021).

ISG support visits. The intended support model for the ISG schools was to provide fortnightly visits to each school, equating to 11 visits across the 23-week implementation period. All schools received 8 of these visits, with the exception of one school which received 9. During the first visit, the research officer reminded staff of the key aspects of the training and ensured staff were ready to begin the programme. Subsequent visits were observational sessions, ensuring the members of staff were delivering the sessions correctly, with the correct resources required. Guidance was given as required – either as identified by the research officer or as requested by school staff. Each member of staff was observed assessing a benchmark story. Once each school was up and running with the delivery of the sessions, they were given a Learner Checklist to keep a record of whether the learner was present, which episode was completed and what score they achieved, or whether they conducted

Fluency Building activities. Following the initial 2-3 visits, an Implementation Checklist was created to keep track of whether the ISG schools were carrying out all aspects of HER as stated in the training and on the training documentation given to schools. The Implementation Checklist was divided into six categories, including: 1) Quality of Learning Environment and Resources (*assessed directly by the research officer, including aspects such as sufficient PC access, and an appropriate working area);* 2) Episodes *(assessed directly by the research officer and self-reported by school staff, including aspects such as sessions being timetabled, ensuring responses to speak-out-loud activities, and repeating episodes when necessary)* 3) Stories *(assessed via self-report, including how the stories were used, whether some were printed out for use in class or at home, and whether staff encouraged learners to engage in the meaning of the stories); 4)* Benchmark Assessments *(assessed directly by the research officer, including checking benchmarks had been completed and scores recorded online, and that appropriate action was taken if the benchmark was rated as 'Needs Practice')*; 5) Fluency Building *(assessed directly by the research officer, including checking activities had been timed, the target had been met 3 times, and that scores had been recorded on the fluency tracker sheet*); 6. Progress and Incentives *(assessed directly by the research officer, enquiring about which aspects of the progress and incentives are being used)*. This procedure was used with all members of staff, with observations recorded on the research officer's copy, and suggestions recorded on the staff copy. Twice during the project, the checklist was used to generate a formal report providing information on aspects being delivered well, and highlighting areas for development, which was shared with implementation staff and the headteacher of the school. The second report was completed after going through the checklist on the final visit, providing information on how well HER was being implemented at the end of the intervention period.

SSG support. All SSG schools were offered technical support. From the technical issues raised by the ISG schools both groups of schools in the trial were given a Technical Support FAQ sheet with solutions to the most common technical issues.

Programme delivery.

HER was intended to be implemented as a supplementary reading programme with beginning readers, therefore delivered alongside their standard reading instruction. The nature of 'standard reading instruction' varied across schools, but typically included a 45minute to 1 hour literacy session each day, and involved teacher and/or TA led phonics sessions for beginning readers, using a variety of commercially available and school-developed phonics programmes.

Post-test measures.

At the end of the school year, after 23 school weeks of HER intervention, we repeated assessments with all children, regardless of whether they finished the programme earlier and regardless of whether they had finished all episodes of the programme. Six trained individuals conducted the post-test assessments. All assessors were blind to trial arm. As with pre-tests, all assessments were recorded using a Dictaphone.

Ethical approval

Ethical approval was granted for this study through the University ethics committee of the first author.

Results

Twenty-two schools were randomised in total, with 11 schools randomised to ISG (SSG) and the remaining 11 to SSG (ISG). In total, 269 learners were included in the study, with 140 in ISG and 129 in SSG. Primary outcome data were available for 253 learners (94.1%) from all 22 randomised schools (See Figure 1 for CONSORT flow diagram).

Support given to the SSG schools consisted of a total of 8 enquiries for technical support and 3 enquiries for implementation support (which resulted in staff from SSG schools being referred back to the training slides). Ten of the ISG schools received 8 support visits, and the eleventh received 9 support visits.

Analysis

Analysis was conducted based on the intention-to-treat principle, with all participants and schools analysed in the groups to which they were randomised. Between-group mean differences in the YARC early word recognition standard score (primary outcome), and YARC standard scores for letter sound knowledge, sound isolation, sound deletion, phoneme awareness, reading accuracy, reading rate, and reading comprehension (all secondary outcomes) were compared using two-level linear regression models, with learners nested within schools. All models were adjusted for the corresponding score pre-randomisation. Additional sensitivity analyses also adjusted for learner age and sex.

Four measures of fidelity to the delivery of the online intervention were considered (percentage of episodes completed at >90% accuracy, percentage of episodes that should have been repeated that were repeated, and mean number of episodes completed per week during the intervention period). Due to the skewed nature of the three ‘percentages’ measures, they were analysed using two-level ordinal regression (up to 50% of sessions/51 to 75% of sessions/>75% of sessions) and logistic regression (0 to 25% of sessions, 26-50% of sessions, 51-65% of sessions, 76-99% of sessions, and 100% of sessions for episode repetitions, and 0%, 1-99%, 100% and Not applicable for benchmark completion data) respectively. Between-group mean differences in the number of episodes completed per week during the intervention period were compared using two-level linear regression.

Results are presented as adjusted mean differences (or odds ratios and relative risk ratios for categorical data) with 95% confidence intervals and p-values.

For participants providing both pre- and post-randomisation YARC standard scores, the overall change in mean score is described, with effect sizes calculated using Cohen's d (with pre-randomisation standard deviation used as the divisor).

Analyses were conducted using Stata version 13.0.

Baseline data

Learners were recruited from across four year groups (Years 1 to 4), with learners from Year 2 comprising the largest year group (106/269 learners, or 39.4% of all randomised learners) and learners from Year 4 the smallest (31/269, or 11.5%). Compared to Schools randomised to SSG, those randomised to ISG included a higher percentage of learners from Year 1 (40.0% compared to 25.6%) and a lower percentage of learners from Year 3 (7.1% compared to 25.6%). Learners from Year 2 and Year 4 were well balanced by Trial arm. Trial arms were also well balanced according to the sex of the learners, with 104/269 female learners included in the study (38.7%), age at baseline (median age of 6 years, IQR from 5 to 7 years), and baseline YARC standard scores (Table 3).

Numbers analysed

As described previously, primary outcome data were available for 253/269 learners (94.1%) from all 22 randomised schools. Post-randomisation letter sound knowledge data were also available for 253 learners. However, post-randomisation sound isolation data were available for 209 learners (77.7% of all learners), sound deletion data for 252 (93.7%), phoneme awareness for 209, reading accuracy for 251 (93.3%), reading rate for 120 (44.6%), and reading comprehension for 223 (82.9%). Outcome data were available for at least one learner from all 22 schools.

Outcomes and estimation

Baseline data indicated the mean early word recognition score was 90.7 in learners from schools randomised to ISG (standard error (SE) = 1.04) and 89.3 in learners from schools randomised to SSG (SE = 1.16). However, there was no evidence of a between-group difference for the early word recognition score (adjusted mean difference = 0.55, 95% CI: -3.05 to 4.15, p = 0.764).

Similarly, there was no evidence of any differences in any of the mean YARC standard scores at baseline (Table 3). The conclusions drawn from these analyses were not altered when also adjusting for learner age pre-randomisation and learner sex (Table 3). The school-level ICCs in the original analyses in Table 3 ranged from 0.03 (sound deletion) to 0.36 (letter sound knowledge). Adjusting for learner age and sex produced lower ICCs (ranging from 0.00 for sound isolation and phoneme awareness to 0.22 to letter sound knowledge), indicating that while these variables explained some of the variation that was attributable to between-school differences, a substantial amount of variation remained in some instances.

For participants in schools allocated to the ISG, the mean number of episodes completed per week during the intervention was 1.4 (SE = 0.05), and while this was higher than for those in schools allocated to the SSG (mean = 1.1, SE = 0.06), there was no statistical discernible difference between groups (mean difference 0.23, 95% CI: -0.18 to 0.64, p = 0.278). As shown in Table 5, the percentage of sessions that were completed with >90% accuracy and the percentage of sessions that were repeated (from those that should have been repeated) were similar between groups, with no evidence of any discernible differences. However, there was a discernible difference between Groups in terms of the percentage of benchmark assessments completed that should have been completed. Seven pupils in the ISG completed 0% compared to 54 pupils in the SSG (5% compared to 42%). All required benchmark assessments were completed by 95 pupils in the ISG compared to 36 pupils in the SSG (68% compared to 28%, relative risk ratio 20.36, 95% CI: 3.47 to 119.59, p = 0.001). There was a higher percentage of pupils in the SSG where the child had not reached the point where they would be eligible to receive any benchmark assessments (16% compared to 9%).

Overall, mean YARC standard scores were consistently higher post-intervention than they were pre-randomisation. Effect sizes ranged from 0.21 for reading accuracy to 0.59 for letter sound knowledge (Table 6).

Exploratory Post-Hoc Analysis

Exploratory post-hoc analyses were conducted to provide additional insights into associations between process measures and our primary outcome (YARC Early Word Recognition standardised scores). Further figures relating to these can be found in the accompanying Supplementary File; Table S1 provides a summary of these results.

A positive relationship was indicated between episode completion and the primary outcome (see Figure S1 and S2), and between episode reached and the primary outcome (see Figures S3 and S4). We regressed the number of episodes completed onto YARC early word recognition standard score post-randomisation, adjusting for pre-randomisation YARC early word recognition standard score and trial arm. We found that, for every additional 10 episodes completed, the mean YARC Early Word Recognition, standardised score post-randomisation (adjusted for baseline score and trial arm) increased by 1.9 points (95% CI: 0.8 to 3.0, p = 0.001).

Although there was no association between the proportion of episodes repeated that should have been repeated and the primary outcome (95% CI: -0.05 to 0.03, p = 0.512; see Table S1), a positive relationship was indicated between the percentage of episodes completed above 90% and the primary outcome (95% CI: 0.18 to 0.29, p = <0.001; see Figure S5 and S6), and the average number of episodes per week and the primary outcome (95% CI: 1.85 to 6.90, p = 0.001; See Figures S7 and S8).

There was no evidence of an overall association between benchmark assessment completion and the primary outcome. However, there was a statistically significant difference between the 100% category (all required benchmarks completed) and the 0% category (no required benchmarks completed; See Table S1).

Staff confidence following training

Following the initial training, staff all reported that they felt somewhat confident to very confident delivering HER episodes, using the stories, and accessing the programme resources. Staff reported they were less confident conducting and responding to benchmark scores and delivering the fluency activities (Figure 2).

Discussion

The extent to which implementation support influences outcomes for programmes in which the core element of the intervention is delivered directly via an app or computer-based programme is not yet clear from the literature. This study investigated the effects of providing ongoing implementation support for schools implementing Headsprout Early Reading as supplementary reading instruction with beginning readers on learner reading outcomes, with the Implementation Support Group schools receiving support throughout the intervention period, and the Standard Support Group schools only receiving initial training and technical support. The results indicate that there was no significant difference in reading outcomes between the groups. These findings deviate from what is typically found in relation to implementation support, with previous research suggesting that this support typically improves outcomes (e.g., Kretlow & Bartholomew, 2010; Elish-Piper & L'Allier, 2011; Owen et al., 2021).

One hypothesized mechanism for the action of implementation support on outcomes is that the quality of fidelity/implementation may be improved with such support. We may have observed a lack of effect because our implementation support model did not improve implementation fidelity sufficiently to have an impact on reading outcomes. Although resources for the trial did not allow for extensive measurement of implementation fidelity, some relevant data were available via the monitoring of implementation that occurred during the study (see Table 5).

There was no significant difference between the Standard Support and Implementation Support Group on the implementation of HER episodes, with both groups averaging fewer than half the number of HER episodes suggested each week. However, there were other differences in implementation indicating some effects of the ongoing support on implementation. There was a higher percentage of learners in the SSG where the child had not reached the point where they would be eligible to receive any benchmark assessments (16% compared to 9%), and there was a significant difference in the implementation of the benchmark assessments, with 68% of learners in the ISG having completed the appropriate benchmark assessments as compared to only 28% of the learners in the SSG. Further, it was also found that only 1 out of 11 schools in the SSG conducted any fluency activities, in comparison with 9 out of 11 ISG schools.

Episode frequency was similar, indicating that the dosage and quality of the core instruction received was similar across the groups. In this respect, the absence of a difference in outcomes may be understandable; if the core instruction was implemented similarly in both groups, we might expect to see similar outcomes. However, there were some other differences in implementation indicating that ongoing support may have affected implementation in some ways (improved use of benchmark assessments and fluency activities).

Regardless of these differences in implementation between groups, there was no difference in outcomes between groups in this study. However, these implementation differences still have potentially significant implications for learners, particularly those at-risk of reading difficulties. The use of benchmark assessments to monitor progress and make instructional decisions is an important aspect of interventions that allow educators to respond early to difficulties. The use of the programme seen in SSG schools indicates that there is a risk those learners most at-risk might not receive the support the programme can provide to enable greater benefit to struggling learners. Further, within this study we investigated the short-term effects of our ISG model, but it is conceivable that differences in reading outcomes might become apparent over a longer period of time. One reason for this could be that completing the programme without using the benchmark assessments may have an impact on the reading skills developed that is more apparent when the programme has been completed. This would be consistent with the notion of cumulative dysfluency – that deficits in lower level component skills can impose a barrier to learning more complex, composite skills that require the components (McDowell & Keenan, 2001; Gallagher, Bones & Lombe, 2006). Missing the benchmark assessments and the impact this could have on identifying learner difficulties and making effective instructional decisions could lead to cumulative dysfluency at a later stage. It is also possible that the individualised implementation support, coaching, and troubleshooting schools received in the ISG during the first year of implementation may have enhanced how well the programme was embedded within those schools, including the establishment of a high quality learning environment, with consistent implementation practices and routines (Outhwaite, Gulliford and Pitchford, 2019; Gulliford et al., 2021), and the way in which learner support was facilitated beyond the educational technology (Gulliford et al., 2021). Future research should therefore also examine implementation as well as outcomes over a longer period of time.

It is also interesting to note that there was variability in reading outcomes and programme progress across schools, regardless of study group. In the ISG, the average episode completion for schools ranged from 20 to 46 (with an overall mean of 32), and in the SSG this ranged from 10 to 56 (with an overall mean of 27). Average episode number reached provides a very similar picture, with only 6 schools reaching the second half of the programme and no learners completing all 80 episodes of the programme. The exploratory post-hoc analysis conducted indicates that episode reached, percentage of episodes above 90%, and episode frequency, are all positively associated with our primary outcome. Further, some effects of the completion of benchmark assessments were also found. These analyses indicate that the strongest predictor of outcome is the *quality* of intervention receipt; the effects were more prominent for indicators of high quality (e.g., episode scores and episode frequency) compared to simply completion of the intervention (e.g., episodes completed). This provides useful information for the development of improved support models and may help schools make decisions regarding their implementation practices.

Limitations and future considerations

Model of training and intervention delivery

All schools received initial training (which was based on our previous experiences using the programme across many schools) and technical support – both of which could be considered “enhanced” support in comparison with schools simply purchasing and implementing the programme independently. Similarly, all schools knew they were participating in a trial which would be comparing implementation across schools. It is conceivable that implementation in some of the SSG schools was enhanced by perceived competition, especially when considering the findings of Schechter et al., (2017) that a contest element to implementation across teachers was related to better implementation. Further, it is arguable the funding and recruitment model for the study resulted in the recruitment of particularly motivated schools, which may have led to better implementation regardless of group. It is also conceivable that the quality of ‘standard literacy instruction’ flooded the effects of the technology-based intervention. However, the extent to which the HER sessions were timetabled outside of the standard literacy sessions varied across schools; it was not always the case that HER sessions translated to additional time engaged in literacy activities, though all learners did continue to receive the main phonics provision provided by the school. Further measures of what standard literacy instruction looked like for each learner in future research would allow for a clearer understanding of whether this might explain the results of the current study.

Although the initial training was designed to include sufficient coverage of the core elements of the programme along with implementation issues we had encountered in previous projects, there were still some aspects staff reported being unsure about following the training session. ISG schools requested the first support sessions to take the form of reminders of key aspects of the training in situ, further supporting the notion that transferring to practice can be challenging (Klingner, Vaughn, Hughes & Arguelles, 1999). Further, there is a clear relationship between aspects rated as lower in the post-training questionnaire and the subsequent implementation, indicating some training improvements that could be made to enhance confidence and good delivery of these components.

Staff delivering the intervention also varied between groups. For example, in two SSG schools, HER was delivered solely by class teachers, whereas this was not the case in any of the ISG schools. Conversely, in seven of the ISG schools, HER was delivered solely by teaching assistants, whereas this was only the case for four of the SSG schools. For all other schools in both groups, HER was delivered by both class teachers and teaching assistants. Although the core intervention is delivered via the computer, the involvement and support of class teachers may facilitate implementation through cooperative scheduling, and access to appropriate resources and space to conduct an intervention. It is therefore possible that the appropriate frequency of HER sessions was more difficult to achieve for more of the ISG schools than SSG schools.

Model of implementation support

Although all support visits involved observing a session and providing feedback on core aspects of HER delivery, the support visits did evolve over the course of the intervention period. The initial 2-3 visits were focused on getting schools up and running and ensuring they were confident in running the core intervention sessions. As such, these support visits were often largely steered by the particular issues or queries staff had, and there was no systematic feedback on implementation provided to staff. Following these initial visits, the Implementation Checklist was introduced to formalise the structure of the support visit sessions and ensure staff implementing the intervention had clear feedback on all aspects of HER session delivery. Although this was used in all subsequent support visits, it was only used to generate a formal report that was shared with the headteacher once during the intervention period, and once at the end of the intervention period. Sharing this information highlighting areas for development with senior leadership sooner and more often could have led to a greater impact on implementation. For example, if headteachers were presented with the data on episode frequency in relation to what was recommended on a biweekly basis, efforts might have been made to increase either staff or IT resources for the intervention.

Our implementation support model led to a relatively low frequency of performance feedback compared to some other models described in the literature. For example, Mortenson and Witt (1998) found that the effects of performance feedback were reduced when feedback sessions were provided weekly rather than daily. In the current study, even the intended fortnightly visits were not possible to deliver. This was largely due to scheduling challenges; with the distance between schools being up to 90 miles, any requests for rescheduling due to staff absence, school inspections, or other events in the school diary presented significant challenges. Such challenges highlight the importance of creating capacity within schools to enable greater frequency of feedback and support following training, either via providing ‘programme champions’ with additional training, or through exploring peer coaching.

Considering the relatively straightforward delivery of HER in comparison with other, teacher-delivered reading programmes, such models of implementation support could have significant potential in supporting effective delivery and helping to embed good implementation practices following training.

Despite the similarities in the support model and frequency of performance feedback, our findings contrasted with Owen et al., (2021) who found significant positive effects of ongoing support on outcomes for a maths fluency intervention. One explanation for this difference relates to the measures used. Fluency measures (the outcome measures used in Owen et al., 2021) are typically more sensitive to change. In the current study, reading rate data were not available for many participants. A standalone reading fluency measures (such as the Word Reading Fluency or Oral Reading Fluency subtests of the Dynamic Indicator of Basic Early Literacy Skills; DIBELS) could be utilised in future research to ensure a measure of reading fluency is available for most participants. Another explanation for this could relate to the technology and feedback mechanisms built into HER; the very nature of this intervention is that the core element of the programme is delivered via the app or computer, with session feedback available for teachers to view following an episode for each learner. Our data demonstrate this aspect of the intervention was implemented similarly across both groups, and that reading outcomes were similar. This indicates that the implementation support did not impact the delivery of the core element, and as such we could conclude that the initial training was sufficient in equipping teachers with the skills to deliver this component effectively to yield positive outcomes. However, as previously discussed, there were notable differences in implementation of other aspects of the programme, particularly in relation to the role of the teacher or educational practitioner in managing the interplay between the app-based learning and teacher intervention for learners who require further support. This has important implications for the development of educational apps more broadly, arguably highlighting the “double-edged sword” of educational technology. With the core component of an intervention being delivered via an app or computer-based programme, there is a risk of an over-reliance on the technology to facilitate the learning of all learners accessing the programme. Well-designed educational technology will have sophisticated instruction to adapt to learners who require additional support, but teachers and other educational practitioners still have a crucial role in ensuring progress and identifying learners requiring additional support, particularly when being delivered within a school context. For example, in the context of reading specifically, it is only when engaging with a reading activity outside of the programme that issues with generalisation are uncovered for some learners, especially those with special educational needs and therefore at greater risk of reading failure. With this in mind, exploring how technology could further support the interplay between app-based learning and teacher intervention seems an important area for development in educational apps.

Conclusion

The results of this study indicate that, for Headsprout Early Reading (a programme in which the core instructional components are delivered directly via an app or computer-based programme), the implementation support provided had no effect on reading outcomes, and no effect of implementation support on delivery of the core elements of the programme. These findings may have implications for providing access to high quality instruction in early reading skills at scale, with minimal training. However, there were some effects of implementation support on the implementation of other programme elements relating to the responsiveness of educators to the individual’s learning within the programme. Further investigation of the impact implementation of these elements might have on outcomes upon completion of the programme, and further development of effective training and support structures and resources to enhance delivery of these components, would help ensure the programme can be implemented to allow for optimal outcomes for all learners. Further, the ways in which educational technology applications can be designed to acknowledge and better support the interplay between app-based learning and accessing additional support external to the technology requires further investigation.

References

Abrami, P. C,, Lysenko, L., & Borokhovski, E. (2020) The effects of ABRACADABRA on reading outcomes: An updated meta-analysis and landscape review of applied field research. *Journal of Computer Assisted Learning,* 1– 20. https://doi.org/10.1111/jcal.12417

Bishop, M. and Santoro, L.E. (2006), Evaluating beginning reading software for at-risk learners. *Psychology in the Schools*, *43,* 57-70. https://doi.org/10.1002/pits.20129

Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-Assisted Instruction in Support of Beginning Reading Instruction: A Review. *Review of Educational Research, 72, 1*, 101–130. https://doi.org/10.3102/00346543072001101

Buzhardt, J., Greenwood, C.R., Abbott, M. & Tapia, Y. (2006). Research on Scaling Up Evidence-Based Instructional Practice: Developing a Sensitive Measure of the Rate of Implementation. *Education Technology Research and Development, 54,* 467–492 https://doi.org/10.1007/s11423-006-0129-5

Cheung, A. C. K. & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis, *Educational Research Review, 7, 3*, 198-215

Clarfield, J., & Stoner, G. (2005). The Effects of Computerized Reading Instruction on the Academic Performance of Students Identified with ADHD, *School Psychology Review, 34, 2*, 246-254, https://doi.org/ 10.1080/02796015.2005.12086286

Coyne, M. D., Kame'enui, E. J., & Simmons, D. C. (2004). Improving Beginning Reading Instruction and Intervention for Students with LD: Reconciling “All” with “Each”. *Journal of Learning Disabilities, 37, 3,* 231-239. https://doi.org/10.1177/00222194040370030801

Durlak, J.A. and DuPre, E.P. (2008), Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. American Journal of Community Psychology, 41: 327-350 327. https://doi.org/10.1007/s10464-008-9165-0

Elish-Piper, L., & L’Allier, S. K. (2011). Examining the Relationship between Literacy Coaching and Student Reading Gains in Grades K–3, *The Elementary School Journal, 112, 1* 83-106. https://doi.org/10.1086/660685

Gallagher, E., Bones, R., & Lombe, J. (2006). Precision teaching and education: Is fluency the missing link between success and failure?, *Irish Educational Studies, 25:1, 93-105,* DOI:10.1080/03323310600597642

Gulliford, A., Walton, J., Allison, K., & Pitchford, N.J. (2021). A qualitative investigation of implementation of app-based maths instruction for young learners. Educational and Child Psychology, 38, 90-108.

Greenwood, C. R. (2009). Treatment integrity: revisiting some big ideas. *School Psychology Review, 38, 4,* 547

Grindle, C.F., Hughes, J. C., Saville, M., Huxley, K. and Hastings, R.P. (2013), Teaching early reading skills to children with autism using mimiosprout early reading. *Behavioral. Interventions, 28,* 203-224. https://doi.org/10.1002/bin.1364

Grindle, C. F., Murray, C., Hastings, R. P., Bailey, T., Forster, H., Taj, S., Paris, A., Lovell, M., Brown, F. J., Hughes, J. C. (2021). Headsprout Early Reading for children with severe intellectual disabilities: A single blind randomised controlled trial. *Journal of Research in Special Educational Needs*, *21,* 334-344.

Higgins, S.E., Xiao, Z., & Katsipataki, M. (2012). The Impact of Digital Technology on Learning : A Summary for the Education Endowment Foundation. Retrieved May 2020 from: https://educationendowmentfoundation.org.uk/evidence-summaries/evidence-reviews/digital-technology/

Huffstetter, M., King, J. R., Onwuegbuzie, A. J., Schneider, J. J., & Powell-Smith, K. A. (2010). Effects of a computer-based reading program on the early reading and oral language skills of at-risk preschool children. *Journal of education for students placed at risk*, *15,* 279-298

Kerres, M., & Buchner, J. (2022). Education after the pandemic: What have we (not) learned about learning. *Education Sciences, 12,* 315-323 https://doi.org/10.3390/educsci12050315

Kim, M. K., McKenna, J. W., & Park, Y. (2017). The Use of Computer-Assisted Instruction to Improve the Reading Comprehension of Students With Learning Disabilities: An Evaluation of the Evidence Base According to the What Works Clearinghouse Standards. *Remedial and Special Education, 38, 4,* 233–245. https://doi.org/10.1177/0741932517693396

Kinder, D., Kubina, R., & Marchand-Martella, N. E. (2005). Special Education and Direct Instruction: An Effective Combination, *Journal of Direct Instruction, 5, 1,* 1-36

Klingner, J. K., Vaughn, S., Hughes, M.T., & Argüelles, M.E. (1999). Sustaining research-based practices in reading: A 3-year follow-up. *Remedial and special education, 20, 5,* 263—74.

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education, 33, 4,* 279-299

Kretlow, A. G., Cooke, N. L., & Wood, C. L. (2012). Using in-service and coaching to increase teachers' accurate use of research-based strategies. *Remedial and Special Education, 33, 6,* 348-361

Kretlow, A. G., Wood, C. L., & Cooke, N. L. (2011). Using in-service and coaching to increase kindergarten teachers' accurate delivery of group instructional units. *The Journal of Special Education, 44, 4,* 234-246

Kulik, C. C., & Kulik, J. A. (1991). Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior, 7,* 75-94.

Layng, T. V. J., Twyman, J. S., & Stikeleather, G. (2003). Headsprout Early Reading: Reliably teaching children to read. *Behavioral Technology Today, 3,* 7-20

McCollum, J. A., Hemmeter, M. L., & Hsieh, W.-Y. (2013). Coaching Teachers for Emergent Literacy Instruction Using Performance-Based Feedback. *Topics in Early Childhood Special Education, 33(1), 28–37,* https://doi.org/10.1177/0271121411431003

McDowell, C., & Keenan, M (2001). Cumulative Dysfluency: Still evident in our classrooms, despite what we know. *Journal of Precision Teaching and Celeration, 17, 1-6*

National Reading Panel (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC: National Institute for child health and development.

Neufeld, B., & Roper, D. (2003). Coaching: A strategy for developing instructional capacity: Promises and practicalities. *The Aspen Institute Program on Education.* Retrieved May 2020, from: https://www.aspeninstitute.org/publications/coaching-strategy-developing-instructional-capacity-promises-and-practicalities/

Noell, G. H, DuHon, G. J, Gatti, S .L, Connell, J. E. (2002). Consultation, follow-up, and implementation of behavior management interventions in general education. *School Psychology Review, 31,* 217–234

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L. et al. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34,* 87–106

Outhwaite, L.A., Gulliford, A., & Pitchford, N.J. (2019). A new methodological approach for evaluating the impact of educational intervention implementation on learning outcomes. *International Journal of Research and Method in Education.* DOI: https://doi.org/10.1080/1743727X.2019.1657081

Owen, K. L., Hunter, S. H., Watkins, R. C., Payne, J. S., Bailey, T., Gray, C., Hastings, R. P., & Hughes, J. C. (2021) Implementation Support Improves Outcomes of a Fluency-Based Mathematics Strategy: A Cluster-Randomized Controlled Trial, *Journal of Research on Educational Effectiveness*, *14:3, 523-542,* https://doi.org/10.1080/19345747.2021.1875526

Pennington, R. C. (2010). Computer-Assisted Instruction for Teaching Academic Skills to Students With Autism Spectrum Disorders: A Review of Literature. *Focus on Autism and Other Developmental Disabilities, 25, 4,* 239–248 https://doi.org/10.1177/1088357610378291

Roberts-Tyler, E.J., Hughes, J.C. and Hastings, R.P. (2020), Evaluating a computer-based reading programme with children with Intellectual Disabilities: feasibility and pilot research. *Journal of Research in Special Educational Needs, 20,* 14-26. https://doi.org/10.1111/1471-3802.12458

Rose, J. (2006). Independent review of the teaching of early reading. Department for education and skills.

Rose, J. (2009). Independent review of the primary curriculum: Final report. Department for education.

Savage, R. S., Erten, O., Abrami, P. C., Hipps, G., Comaskey, E., van Lierop, D. (2010). ABRACADABRA in the hands of teachers: The effectiveness of a web-based literacy intervention in grade 1 language arts programs, *Computers and Education, 55, 2,* 911-922, https://doi.org/10.1016/j.compedu.2010.04.002

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,* 310–328. https://doi.org/10.1037/a0031025

Schechter, R. L., Elizabeth R., Kazakoff, E. R., Bundschuh, K., Prescott, J. E., & Paul

Schieffer, C., Marchand-Martella, N. E., Martella, R. C., Simonsen, F. L., Waldron-Soler, K. M. (2002). An analysis of the Reading Mastery program: Effective components and research review. *Journal of Direct Instruction, 2,* 87–119

Storey, C., McDowell, C., Leslie, J.C. (2017). Evaluating the efficacy of the Headsprout© reading program with children who have spent time in care. *Behavioral Interventions, 32,* 285- 293. https://doi.org/10.1002/bin.1476

Twist, L., Jones, E. and Treleaven, O. (2022). The Impact of Covid-19 on pupil attainment. NFER: Slough.

Twyman, J. S., Layng, T. V .J., Layng, Z. R. (2011). The likelihood of instructionally beneficial, trivial, or negative results for kindergarten and first grade learners who complete at least half of Headsprout Early Reading, *Behavioral technology today, 6,* 1-19,

Tyler, E., Hughes, J., Beverley, M., & Hastings, R. (2015). Improving early reading skills for beginning readers using an online programme as supplementary instruction. *European Journal of Psychology of Education, 30, 3,* 281-294, https://doi.org/10.1007/s10212-014-0240-7

Tyler, E. J., Hughes, J. C., Wilson, M. M., Beverley, M., Hastings, R. P., Williams, B. M. (2015). Teaching Early Reading Skills to Children with Intellectual and Developmental Disabilities Using Computer-Delivered Instruction: A Pilot Study. *Journal of International Special Needs Education, 18, 1,* 1–11, https://doi.org/10.9782/2159-4341-18.1.1

Tyler, E. J., Watkins, R. C., Roberts, S. E., Hoerger, M., Hastings, R. P., Hulson-Jones, A. L., & Hughes, J. C. (2019). The Collaborative Institute for Education Research, Evidence and

Impact: A Case Study in developing regional research capacity in Wales. *Wales Journal of Education, 21, 1,* 89-108 https://doi.org/10.16922/wje.21.1.6

Watkins, R., Hulson-Jones, A. L., Tyler, E. J., Beverley, M., Hughes, J. C., & Hastings, R. P. (2016). Evaluation of an online reading programme to improve pupils' reading skills in primary schools: Outcomes from two implementation studies. *Wales Journal of Education, 18, 2,* 81-104, https://doi.org/10.16922/wje.18.2.7

Welsh Government (2013). *National Literacy and Numeracy Framework*. Cardiff: Welsh Government

Welsh Government (2014). *Qualified for Life*. Cardiff: Welsh Government

Welsh Government (2015). Effective use of data and research evidence. Cardiff: Welsh Government

Whitcomb, S. A., Bass, J. D., & Luiselli, J. K. (2011). Effects of a computer-based early reading program (Headsprout) on word list and text reading skills in a student with autism. *Journal of physical and developmental disabilities, 23,* 491-499, https://doi.org/10.1007/s10882-011-9240-6

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs

Tables and Figures

Table 1: Number of learners with scores below the range for which standardised scores were provided at pre and post-test in both arms of the trial

YARC subtest	Time	ISG	SSG
Early Word Recognition	Pre	4	7
	Post	3	4
Letter Sound Knowledge	Pre	15	27
	Post	14	10
Sound Insertion	Pre	11	2
	Post	3	5
Sound Deletion	Pre	4	7
	Post	4	3
Phoneme Awareness	Pre	6	7
	Post	4	2
Reading Accuracy	Pre	4	7
	Post	5	5
Reading Rate*	Pre	0	1
	Post	0	0
Reading Comprehension*	Pre	7	3
	Post	4	7

*See Table 2 for information on missing cases for these measures

Table 2: Baseline characteristics of learners from randomised schools

		ISG	**SSG**	**Overall**
Number of schools		11	11	22
Number of learners		140	129	269
Year group*	**Year 1**	56 (40.0)	33 (25.6)	89 (33.1)
	Year 2	57 (40.7)	49 (38.0)	106 (39.4)
	Year 3	10 (7.1)	33 (25.6)	43 (16.0)
	Year 4	17 (12.1)	14 (10.9)	31 (11.5)
Sex of learner*	**Female**	55 (39.3)	49 (38.0)	104 (38.7)
	Male	85 (60.7)	80 (62.0)	165 (61.3)
Whether child is eligible for free school meals*	**No**	68 (50.4)	83 (64.8)	151 (57.4)
	Yes	67 (49.6)	45 (35.2)	112 (42.6)
Age at baseline†		6 (5 to 6)	6 (6 to 7)	6 (5 to 7)
Baseline YARC standard scores	**Early word recognition‡**	85.5 (9.49)	84.5 (11.32)	85.0 (10.40)
	n missing	*0*	*0*	*0*
	Letter sound knowledge‡	82.9 (12.65)	81.5 (14.64)	82.2 (13.63)
	n missing	*0*	*0*	*0*
	Sound isolation‡	87.4 (15.06)	88.6 (12.81)	88.0 (14.01)

	n missing	*1*	*0*	*1*
	Sound deletion‡	90.2 (12.63)	88.1 (13.31)	89.2 (12.98)
	n missing	*2*	*0*	*2*
	Phoneme awareness‡	87.8 (12.44)	87.0 (11.47)	87.4 (11.96)
	n missing	*2*	*0*	*2*
	Reading accuracy‡	84.7 (7.94)	83.9 (9.47)	84.3 (8.70)
	n missing	*3*	*3*	*6*
	Reading rate‡	81.2 (4.97)	75.8 (21.98)	77.8 (17.60)
	n missing	*131*	*114*	*245*
	Reading comprehension‡	79.7 (10.22)	81.1 (8.91)	80.5 (9.52)
	n missing	*71*	*46*	*117*

* Frequency (percentage). †Median (interquartile range (IQR)). ‡ Mean (standard deviation (SD))

Table 3: Between-group comparison of post-randomisation YARC standard scores

Outcome	ISG Mean (SE)	SSG Mean (SE)	N in model (learners)	N in model (schools)	Adjusted Mean Difference (ISG – SSG) *	95% Confidence Interval		p-value	Intra-cluster Correlation Coefficient
						Lower Limit	Upper Limit		
Early word recognition	90.7 (1.04)	89.3 (1.16)	253	22	0.55	-3.05	4.15	0.764	0.08
Letter sound knowledge	92.3 (1.52)	91.3 (1.56)	253	22	-1.28	-9.39	6.84	0.758	0.36
Sound isolation	94.5 (1.56)	94.5 (1.42)	208	22	0.22	-4.37	4.81	0.925	0.04
Sound deletion	92.6 (1.14)	92.8 (1.08)	250	22	-0.80	-4.25	2.66	0.651	0.03
Phoneme awareness	91.9 (1.22)	92.3 (1.05)	207	22	-0.81	-4.13	2.51	0.634	0.04
Reading accuracy	87.2 (1.03)	86.0 (1.12)	246	22	0.32	-3.60	4.25	0.871	0.13
Reading rate†	88.1 (1.24)	87.5 (1.27)	120	22	-0.12	-5.85	5.61	0.967	0.32
Reading comprehension†	85.6 (1.04)	83.3 (1.17)	223	22	2.00	-2.53	6.52	0.387	0.12

*Adjusted for the corresponding score prior to randomisation. †Due to the low response rate for the pre-randomisation reading rate and comprehension scores, these variables were not included in the final analyses.

Table 4: Between-group comparison of post-randomisation YARC standard scores, adjusting for learner age pre-randomisation and learner sex

Outcome	**Adjusted Mean Difference (ISG – SSG) ***	**95% Confidence Interval**		**p-value**	**Intra-cluster Correlation Coefficient**
		Lower Limit	**Upper Limit**		
Early word recognition	1.24	-1.64	4.13	0.398	0.02
Letter sound knowledge	-1.67	-7.78	4.45	0.593	0.22
Sound isolation	-0.61	-4.62	3.40	0.766	0.00
Sound deletion	-1.05	-4.61	2.51	0.564	0.04
Phoneme awareness	-1.49	-4.60	1.61	0.346	0.02
Reading accuracy	0.14	-3.50	3.78	0.940	0.10
Reading rate†	-0.56	-4.31	3.19	0.769	0.11
Reading comprehension†	0.89	-3.29	5.07	0.676	0.09

*Adjusted for the corresponding score prior to randomisation, learner age pre-randomisation, and learner sex. †Due to the low response rate for the pre-randomisation reading rate and comprehension scores, these variables were not included in the final analyses.

Table 5: Between-group comparisons of categorical fidelity measures

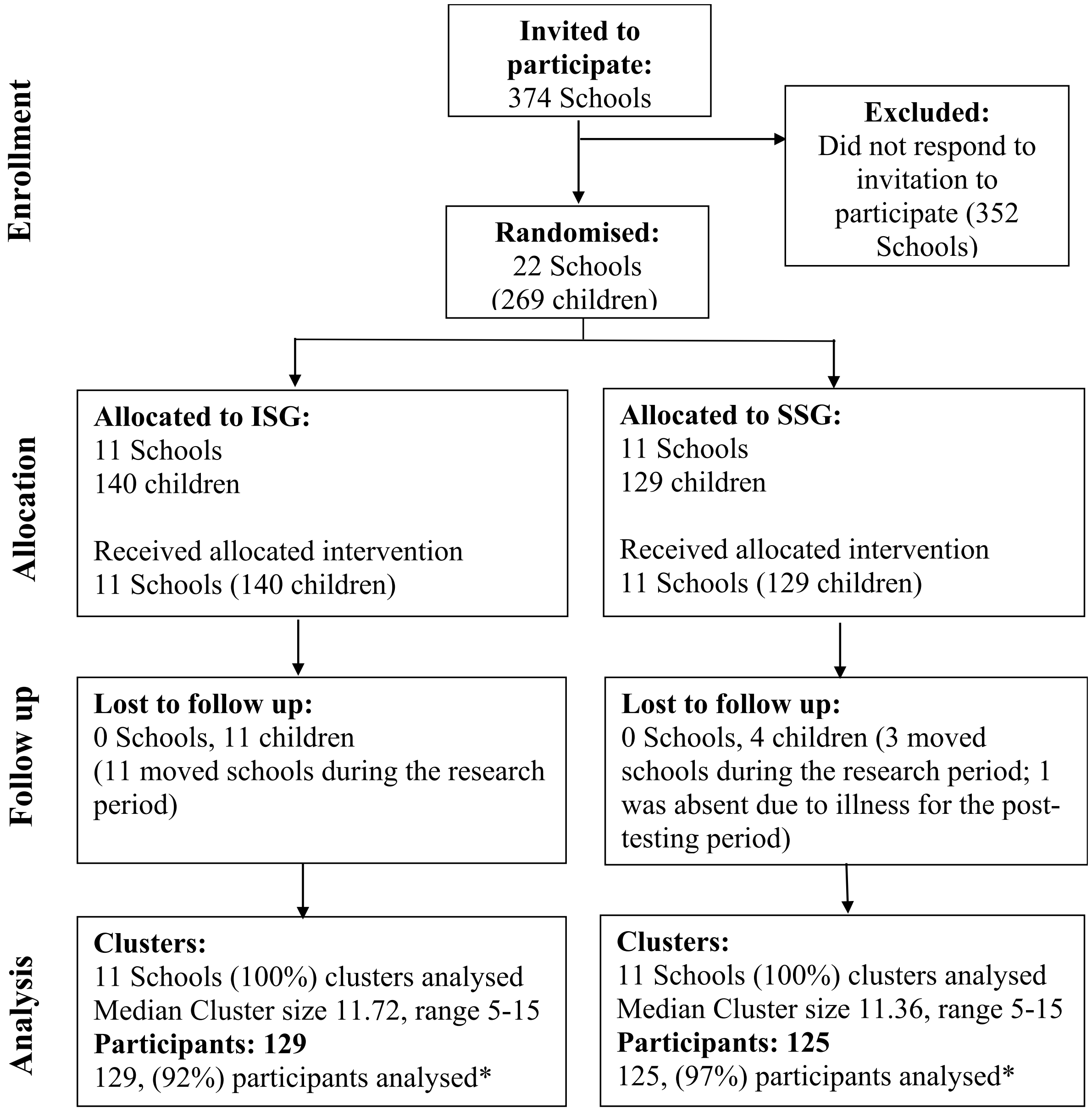
Outcome	**Responses**	**ISG [n (%)]**	**SSG [n (%)]**
Percentage of sessions that were completed at >90% accuracy	Up to 50% of sessions	7 (5.4)	17 (13.7)
	51 to 75% of sessions	41 (31.8)	35 (28.2)
	>75% of sessions	81 (62.8)	72 (58.1)
Percentage of sessions that should have been repeated that were repeated	0 to 25%	9 (6.6)	32 (24.8)
	26 to 50%	22 (16.2)	11 (8.5)
	51 to 75%	23 (16.9)	15 (11.6)
	76 to 99%	30 (22.1)	19 (14.7)
	100%	52 (38.2)	52 (40.3)
Percentage of benchmark assessments completed that should have been completed	0%	7 (5.0)	54 (41.9)
	1 to 99%	26 (18.6)	19 (14.7)
	100%	95 (67.9)	36 (27.9)
	Not applicable	12 (8.6)	20 (15.5)

Table 6: Effect sizes for change in YARC standard scores pre- and post-randomisation

Outcome	N (learners)*	Pre-intervention Mean (SD)	Post-intervention Mean (SD)	Effect size
Early word recognition	253	85.1 (10.29)	90.0 (12.37)	0.42
Letter sound knowledge	253	82.3 (13.72)	91.8 (17.29)	0.59
Sound isolation	208	88.2 (14.25)	94.4 (15.13)	0.42
Sound deletion	250	89.4 (13.03)	92.8 (12.43)	0.27
Phoneme awareness	207	87.7 (12.52)	92.1 (11.49)	0.37
Reading accuracy	246	84.5 (8.73)	86.8 (12.01)	0.21
Reading rate	23	77.7 (17.97)	83.8 (5.83)	0.43
Reading comprehension	144	80.4 (9.64)	85.9 (12.80)	0.48

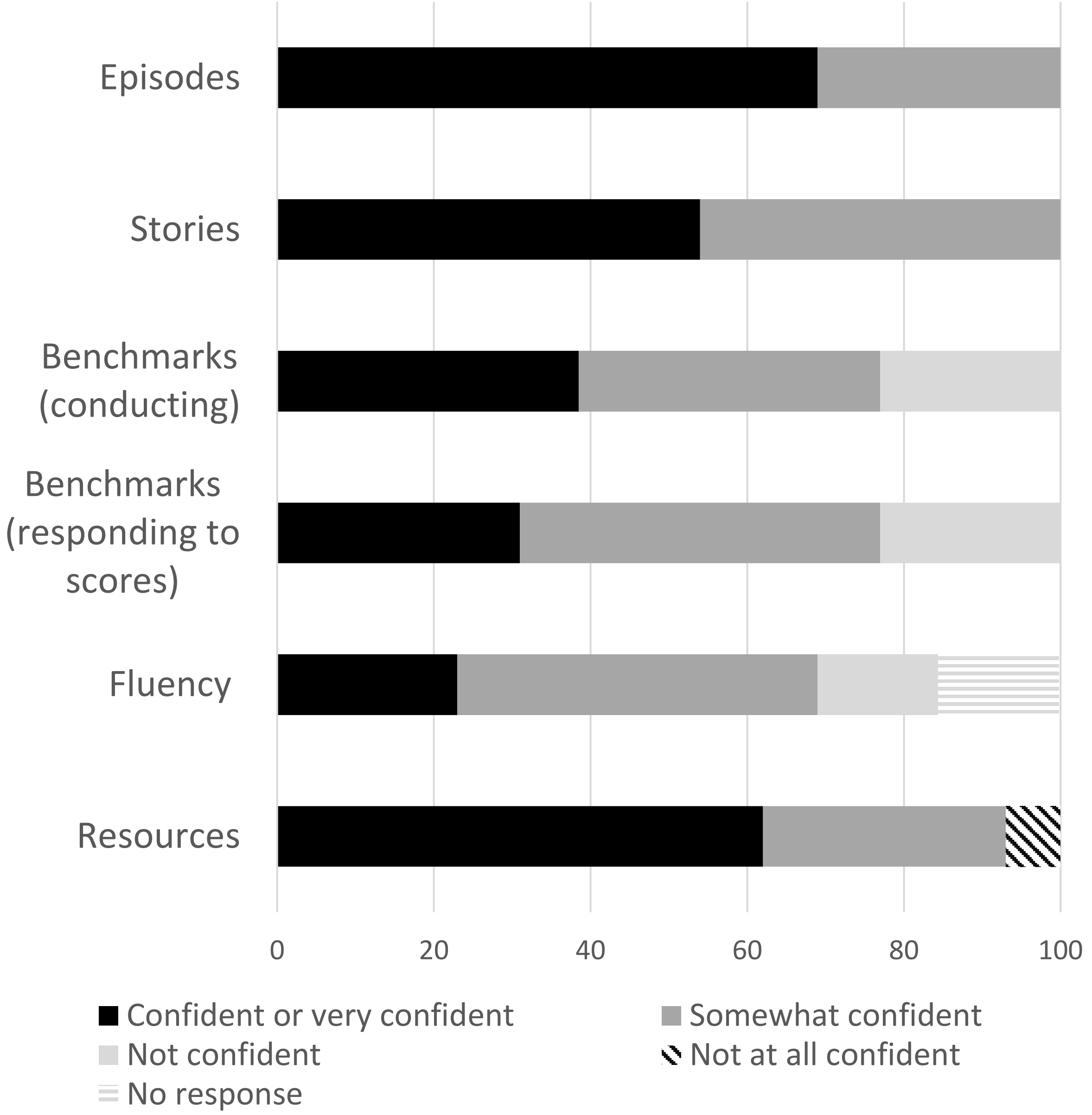
*Only calculated where both pre- and post-randomisation scores were available

Figure 1: CONSORT school and participant flow diagram


Enrollment
Allocation
Follow up
Analysis
Invited to participate: 374 Schools
Excluded: Did not respond to invitation to participate (352 Schools)
Randomised: 22 Schools (269 children)
Allocated to ISG: 11 Schools 140 children
Received allocated intervention 11 Schools (140 children)
Allocated to SSG: 11 Schools 129 children
Received allocated intervention 11 Schools (129 children)
Lost to follow up: 0 Schools, 11 children (11 moved schools during the research period)
Lost to follow up: 0 Schools, 4 children (3 moved schools during the research period; 1 was absent due to illness for the post-testing period)
Clusters: 11 Schools (100%) clusters analysed
Median Cluster size 11.72, range 5-15
Participants: 129
129, (92%) participants analysed*
Clusters: 11 Schools (100%) clusters analysed
Median Cluster size 11.36, range 5-15
Participants: 125
125, (97%) participants analysed*

*some participants in both groups were excluded from the analysis of some YARC subtests due to assessment error. See results for further details.

Figure 2. Implementation confidence for each aspect of the programme following training



Supplementary File: Exploratory Post-Hoc Analysis for "Effects of implementation support on children's reading outcomes following an online early reading programme: a cluster-randomised controlled trial"

Table S1: Associations between process measures and our primary outcome

Process measures		**Adjusted mean difference***	**Lower 95% CI**	**Upper 95% CI**	**p-value**
Number of episodes completed		0.19	0.08	0.30	0.001
% of episodes above 90%		0.23	0.18	0.29	<0.001
% of episodes repeated that should have been repeated		-0.01	-0.05	0.03	0.512
Episode reached		0.29	0.20	0.38	<0.001
% of benchmark assessments completed that should have been completed	0%	Reference category			0.073
	1-99%	1.24	-2.70	5.17	
	100%	4.25	0.19	8.31	
Average number of episodes per week		4.37	1.85	6.90	0.001

*Based on a two-level linear regression model of YARC early word reading standard score post-randomisation, accounting for clustering of pupils within schools. Model adjusts for pre-randomisation YARC early word recognition standard score and trial arm. Coefficients are per unit increase.

Relationship between number of episodes completed and outcome

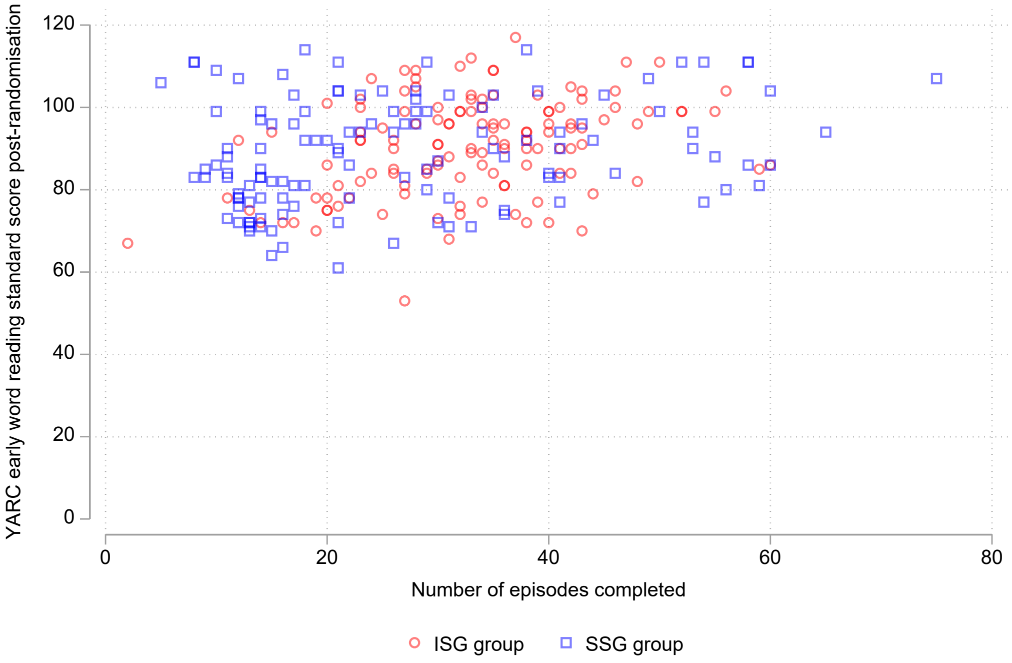


Figure S1: Scatter plot of the relationship between number of episodes completed and YARC early word recognition standard score post-randomisation

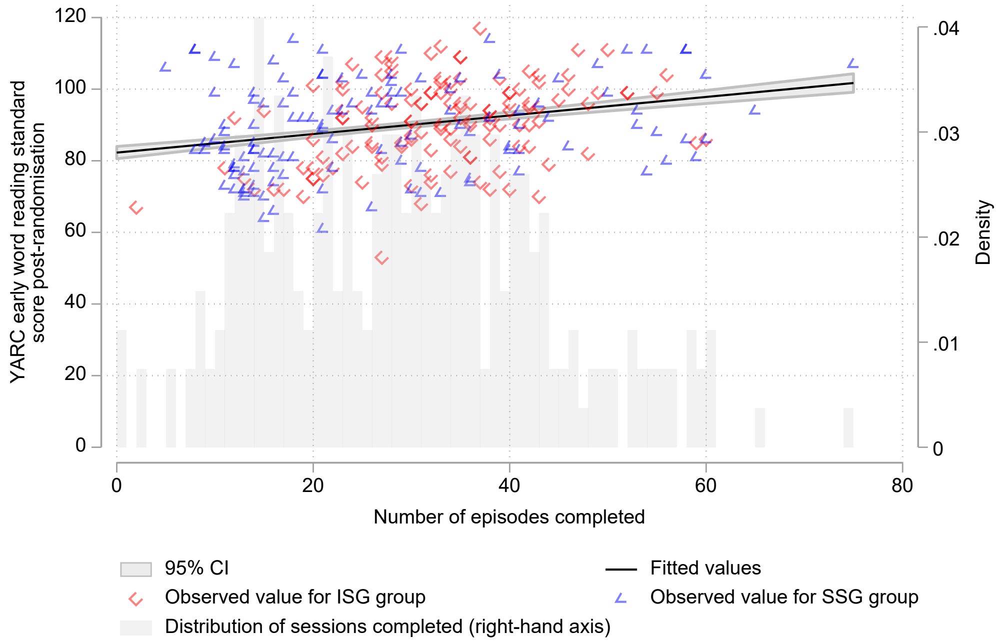


Figure S2: Relationship between episode completion and YARC early word recognition standard score post-randomisation, and distribution of sessions completed

Relationship between episode reached and outcome

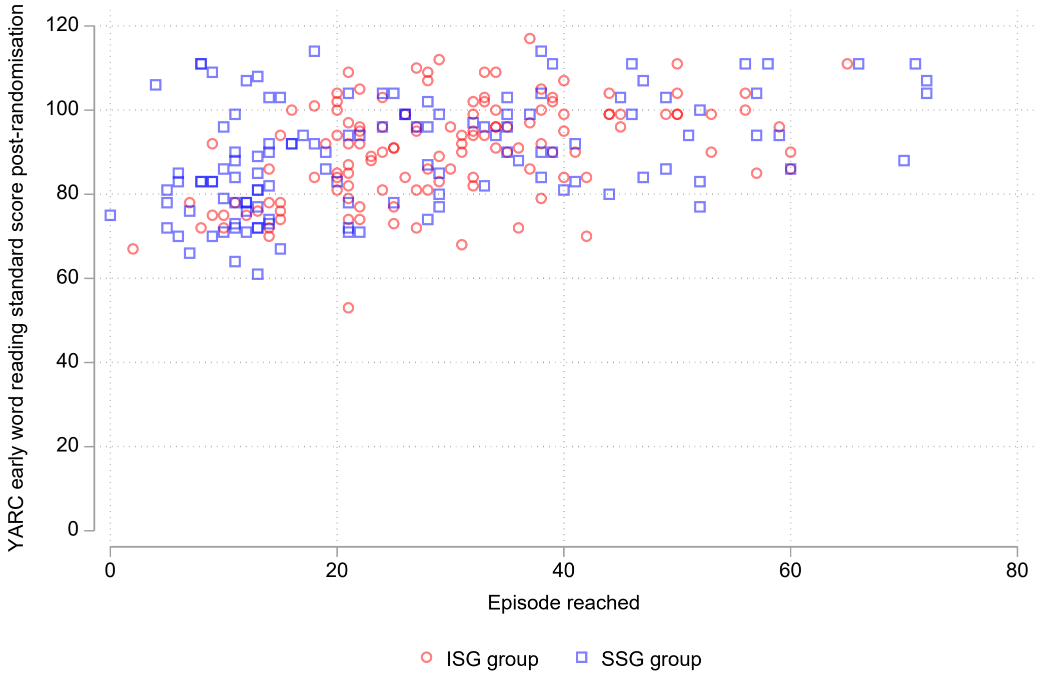


Figure S3: Scatter plot of the relationship between episode reached and YARC early word recognition standard score post-randomisation

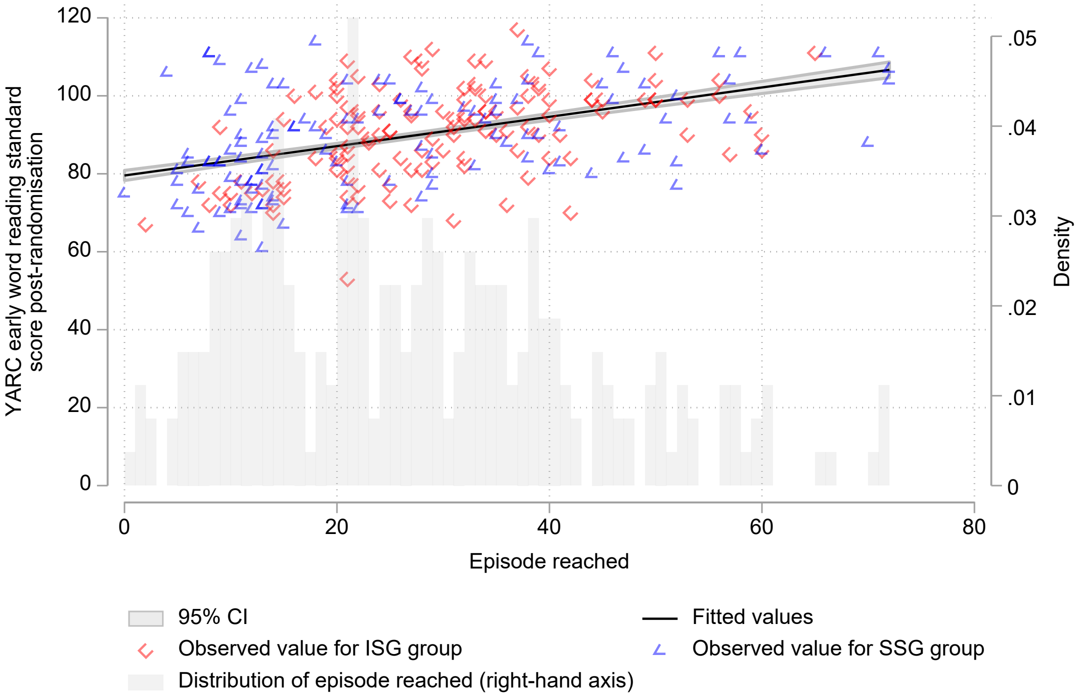


Figure S4: Scatter plot of the relationship between episode reached and YARC early word recognition standard score post-randomisation, and distribution of episode reached

Relationship between % of episodes completed above 90% and outcome

The scatterplot below suggests a positive relationship between the % of episodes completed above 90% and the primary outcome.

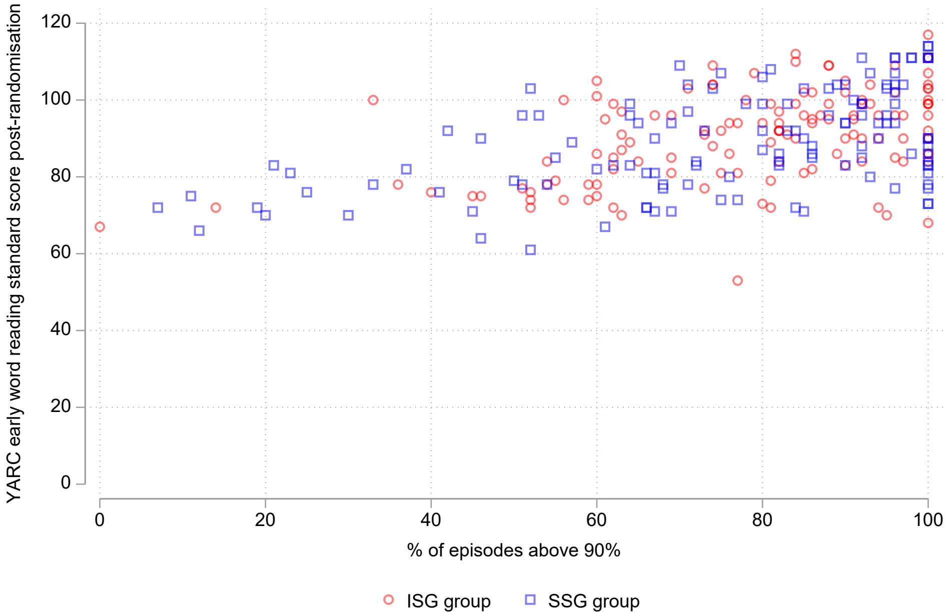


Figure S5: Scatter plot of the relationship between % of episodes above 90% and YARC early word recognition standard score post-randomisation

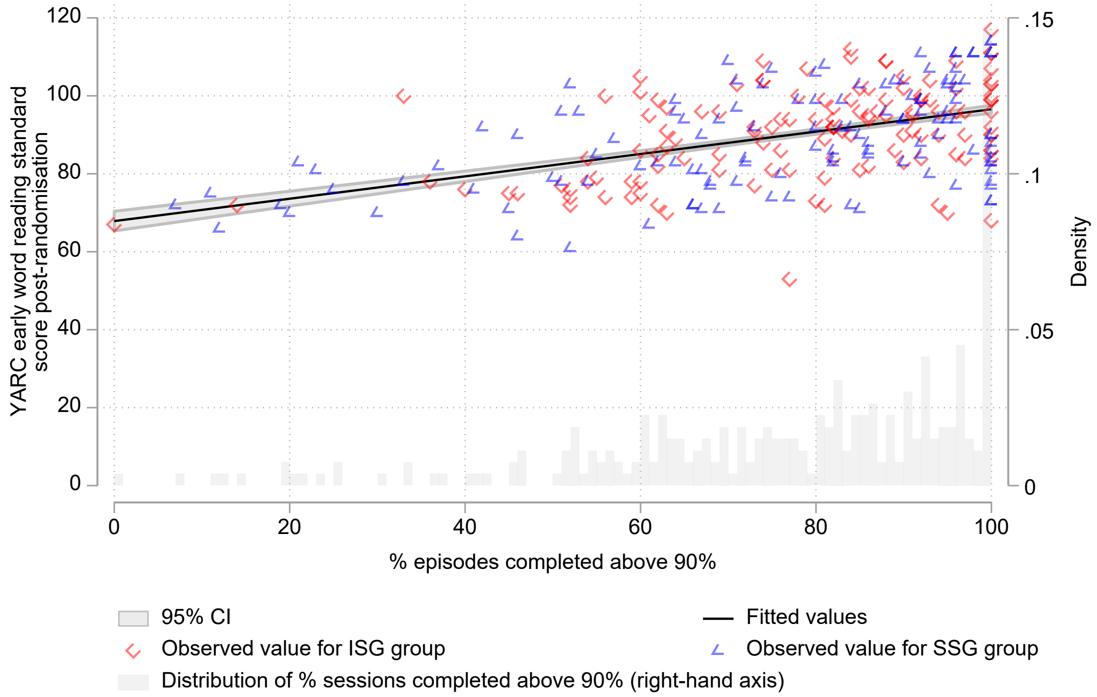


Figure S6: Scatter plot of the relationship between % of episodes above 90% and YARC early word recognition standard score post-randomisation, and distribution of % of sessions completed above 90%

Relationship between average number of episodes per week and outcome

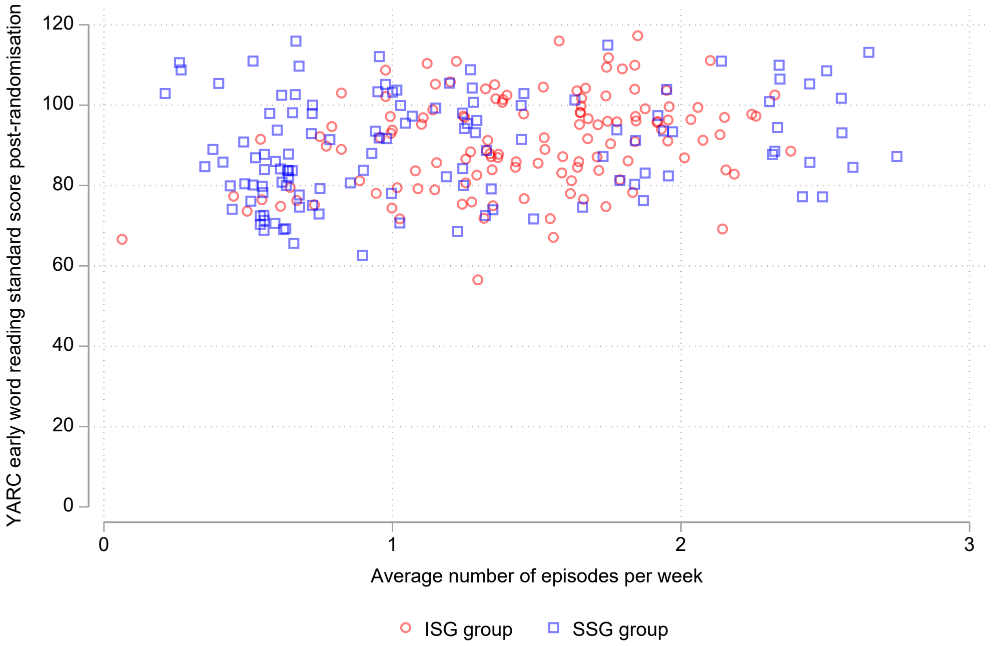


Figure S7: Scatter plot of the relationship between average number of episodes per week and YARC early word recognition standard score post-randomisation

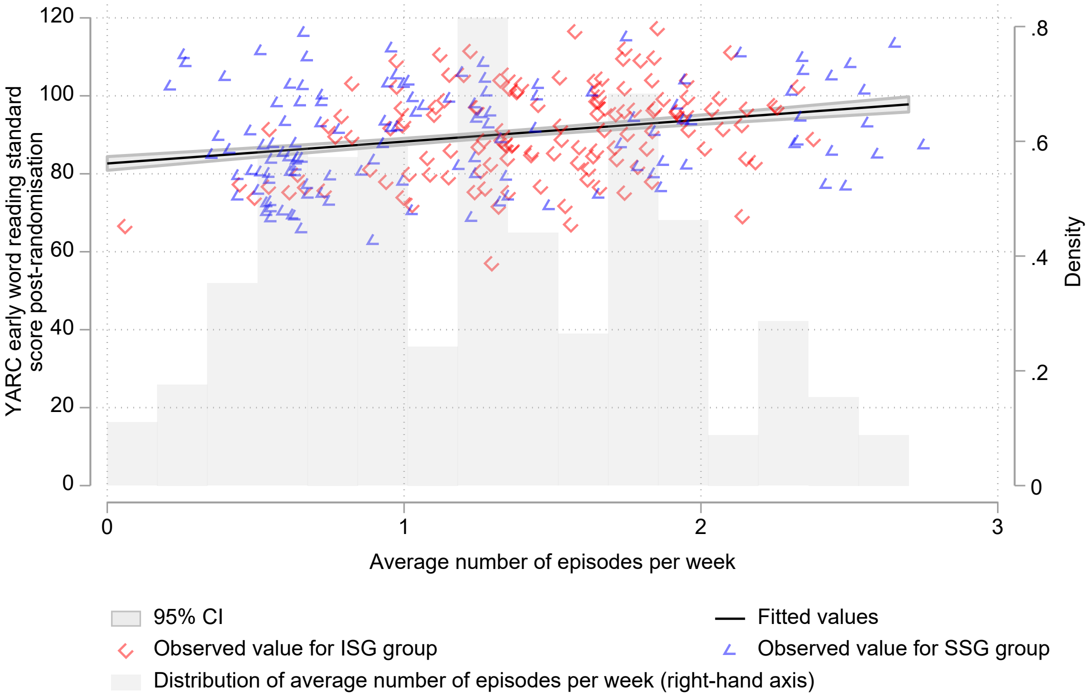


Figure S8: Scatter plot of the relationship between average number of episodes per week and YARC early word recognition standard score post-randomisation, and distribution of average number of episodes per week