Education Endowment Foundation

Headsprout Early Reading® in Special Schools (HERiSS) A randomised controlled trial

Evaluation report

October 2023

Dr Louise Denne, Dr Samantha Flynn, Dr Paul Thompson,
Dr Nikita Hayden, Dr Caitlin Murray, Dr Rebecca Morris, and
Prof Richard Hastings

WARWICK
THE UNIVERSITY OF WARWICK

Education Endowment Foundation

The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries and colleges to improve teaching and learning for 2 – 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were re-endowed with an additional £137m, allowing us to continue our work until at least 2032.

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

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

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Department for Education

What Works Network

Contents

About the evaluator

Dr Samantha Flynn, Joint Principal Investigator, Centre for Educational Development, Appraisal and Research (CEDAR),[1] The University of Warwick.

Dr Louise Denne, Joint Principal Investigator, CEDAR, The University of Warwick.

Professor Richard P. Hastings, Co-investigator, Head of Department, mentor for Dr Flynn and Dr Denne, CEDAR, The University of Warwick.

Dr Paul Thompson (replacing Dr Tom Bailey), Lead Statistician, CEDAR, The University of Warwick.

Dr Rebecca Morris, Co-investigator and Lead on Education, Centre for Education Studies, The University of Warwick.

Contact details for correspondence:

Dr Louise Denne

Centre for Educational Development, Appraisal and Research (CEDAR),

The University of Warwick,

Coventry, CV4 7AL,

United Kingdom

Email: L.Denne@warwick.ac.uk

Acknowledgements

The delivery and evaluation teams would like to thank all schools, staff, and pupils who have taken part in Headsprout Early Reading® in Special Schools (HERiSS).

The evaluation team would also like to thank Karen Talty and her team and pupils at Park House School in London, for piloting the online versions of the baseline and post-intervention assessments including the reading self-concept measure.

There are several members of staff from both teams that we would like to acknowledge:

Implementation Support Officers: Beverley Jones; Esther Thomas; Kate Geary; Laura Jones; and Esther Owen.

Research Assistants and Administrative Assistants: Anam Iqbal; Rajvir Bachu; Belinda Moyo; Emma Scripps; Olugbemi Moronfolu; Vicky Hill; Sophia Abdi; Nicholas Burdett; Diana Smith; Nathan Keates; Jane Pringle; Melanie Howell; Krysia Waldock; Caitlin Williams; Nicholas Manktelow; Josephine Stanley; Alison Baker; Suzi Sapiets; Chioma Kpogho; Atiyya Nisar; Sebastian Croft; Louise Carter; Olivia Hewlett; Sarah Chung; and Wardah Hussain.

Finally, we would like to thank Dr Julia Badger for independently reviewing the interviews for teacher/teaching assistants and implementation support officers.

[1] CEDAR was the name of the centre at the time of the study, but is now currently known as the Centre for Research in Intellectual and Developmental Disabilities (CIDD).

Executive summary

The project

Headsprout Early Reading® in Special Schools (HERiSS) is a computer-based, targeted reading intervention for pupils in special schools in England. The intervention is based on Headsprout Early Reading® (HER®), a programme originally developed in the United States by Headsprout® for children aged four to seven years in mainstream schools. HER® aimed to improve early reading through 100 online episodes, which targeted phonemic awareness, phonics, and reading fluency.

While the original online episodes continued to be used in HERiSS, a team from Bangor University adapted the intervention to support implementation in special schools. The target age was changed to pupils between five and eleven years (Key Stage 1 and Key Stage 2) who were at a reading level suitable for the HER® programme, who could follow instruction, engage with, and respond as necessary to the programme. Special schoolteachers and teaching assistants received initial training, an support manual, and fortnightly implementation support to deliver the programme.

The intervention was scheduled to start in September 2020, but the preparation and recruitment of schools was disrupted due to the Covid-19 pandemic. It eventually started in September 2021 and ran for nine months, finishing in May 2022. Each pupil in the intervention group was scheduled to receive the programme for 10–30 minutes three times per week until the end of the programme.

The University of Warwick evaluated the two-arm, cluster randomised controlled efficacy trial, in which 382 children in 55 schools took part after being randomised at the school level to receive HERiSS (treatment) or business as usual (control). Further evaluation of the programme was conducted through surveys and interviews.

Table 1: Key conclusions

Key conclusions
1. Children in special schools receiving Headsprout Early Reading® in Special Schools (HERiSS) made no additional progress in reading, on average, compared to children in other special schools. This result has a low security rating
2. There is some evidence of HERiSS having a positive impact on oral reading fluency. However, this finding should be interpreted with caution, as there is considerable uncertainty around the result, with data from 45% of the pupils missing from this analysis
3. There was the equivalent of one month's additional progress found in reading for pupils eligible for free school meals (FSM) who received HERiSS compared to FSM eligible pupils who did not receive the intervention. For these pupils, there was also evidence of a positive impact on phonemic segmentation fluency (the ability to break down a word into its phonetic components). However, due to there being a limited number of FSM pupils in the evaluation, these results should be interpreted with caution
4. A total of 55 schools and 382 pupils were recruited to this trial, which is one of the largest randomised controlled trials to be funded in special schools in the United Kingdom. Difficulties of implementation in special schools including pressures on staffing, time constraints, meeting individual pupil needs, pupil and staff absence, and logistics were exacerbated by Covid-19, meaning the intervention could not be delivered as intended
5. Factors that facilitated implementation included Senior Leadership Team buy-in and support, pupils' positive response to the intervention, ease of use of the online programme, and good logistical organisation in schools

EEF security rating

These findings have a low security rating. This was an efficacy trial, which tested whether the intervention worked under develop-led conditions in a number of schools. The trial was a well-designed two-armed, cluster randomised controlled efficacy trial. The trial was not as well-powered as originally intended because the number of schools recruited was significantly lower than expected. 27% of pupils who started the trial were not included in the primary analysis, with more pupils withdrawing from the intervention than the control arm, making it more difficult to securely evaluate the impact. Furthermore, the lack of engagement with all training components, as well as the intervention not being delivered as intended makes it harder to accurately estimate the impact of the trial.

Additional findings

Pupils in HERiSS schools made, on average, zero months' additional progress in reading compared to those in the control group equivalent. This is our best estimate of impact, which has a low security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes negative effects of up to four months less progress and positive effects of up to three months of additional progress.

The intervention had positive effects on oral reading fluency, as well as for phonemic segmentation for free school meals eligible children. However, given that no corrections for multiple comparisons were made for secondary outcome analyses and the FSM subgroup analysis included a reduced sample, there is greater uncertainty about these findings on secondary and subgroup analyses compared to the primary outcome analyses of the whole sample.

Schools interviewed were enthusiastic about taking part in research focusing on special educational needs and disability. All teachers and most pupils interviewed stated that the intervention was positively experienced by pupils and teachers would continue with HERiSS.

However, the diverse needs of pupils, frequent absences, staff turnover, and accompanying pressures on staff in special schools were barriers to implementation and further aggravated by pressures on staff due to Covid-19. These pressures led to only 23% of schools fully completing the training, implementation, and supervision requirements of the programme, with 38% of schools meeting none of the requirements and only ten pupils in total reaching episode 40. This likely influenced the results, as the impact evaluation suggests that the effects of the intervention increased with better compliance from schools.

Covid-19 also exacerbated difficulties in recruitment of schools to take part in the evaluation, with low recruitment putting the trial at risk. The trial progressed despite lower-than-expected recruitment numbers as this was the first intervention in special schools funded by the EEF and one of the few randomised controlled trials that has been conducted in this context. Much useful information was obtained regarding conducting interventions in special schools, including the need for ease of implementation, planning for student absences, staff workload, and staff turnover. This evaluation also shows that conducting randomised controlled trials in special schools is feasible, despite difficulties.

Cost

The average cost of the HERiSS intervention for one special school was £2,067, or £29 per pupil per year when averaged over three years. This is assuming all pupils in one class follow HERiSS, and schools cover staff costs to attend training and implementation support.

Impact

Table 2: Summary of impact on primary outcome(s)

Outcome / group	Effect size (95% CI)	Estimated months' progress	EEF security rating	No. of pupils	*P*-value	EEF cost rating
Reading	-0.01 (-0.28, 0.26)	0	🔒🔒🔒🔒🔒	382	0.951	£ £ £ £ £
Reading – FSM eligible pupils	0.09 (-0.3, 0.48)	1	N/A	201	0.657	£ £ £ £ £

Introduction

Background

Much research that aims to teach children with special educational needs and disability (SEND) to read has focused on how to recognise words (sight word instruction) without teaching the individual letter/sound components and decoding skills (phonics instruction). Although some approaches to sight word instruction can effectively teach individual target words, research using these approaches demonstrates issues with generalising these skills to be able to develop functional conventional reading skills (Katims, 2000; Browder, *et al.,* 2006; Joseph and Seery, 2004; Sermier Dessemontet, *et al.,* 2019; Reichow, *et al.,* 2019). This, coupled with a lack of training in an effective and systematic way of teaching reading to children in special schools (Roberts-Tyler, Hughes, and Hastings, 2019) has led to insufficient teaching effectiveness to enable pupils to access and benefit from phonics-based early reading intervention, and a lack of adaptation of programmes to provide sufficient support for delivering programmes effectively for children with SEND. These potentially surmountable learning challenges can be perceived as indicating that pupils with SEND cannot learn conventional early reading skills.

The ability to read is a critical life skill (Marchand-Matella, Slocum, and Martella, 2004). The considerable attainment gap in reading for children with SEND increases across the key stages (DfE, 2018) and translates to poorer post-secondary outcomes related to employment, housing, and social engagement for people with intellectual disabilities (IDs) (Parkin, *et al.,* 2018). As such there is a compelling need for research into the development and evaluation of evidence-based methods that target literacy skills, that can be delivered in schools effectively, and result in improved outcomes.

Recent evidence indicates that explicit, systematic, and intensive instruction in the early primary years for children considered ‘at risk’ of reading difficulties, can have significant and sustained effects on reading skills (Coyne, *et al.,* 2004). Further, increasing evidence indicates many children with an ID can benefit from phonics-based instruction (Allor, *et al.,* 2010; Bradford, *et al.,* 2006; Sermier Dessemontet, *et al.,* 2019). The ability to decode (learning correspondences between written and spoken sounds and being able to read these fluently at both a word and sentence level) is an essential component to becoming a proficient reader. As outlined in the widely supported conceptual framework proposed as the Simple View of Reading (Gough and Tunmer, 1986), decoding alone is not sufficient for proficient reading (i.e. understanding what is being read), but decoding skills are an essential element of early reading instruction to enable future proficient reading. Considerable evidence for early phonics instruction also supports this notion (Foorman, *et al.,* 2016).

Headsprout Early Reading® (HER®) is a computer-based, targeted reading programme developed in Seattle, United States by Headsprout®, an applied learning sciences company, which brought together on a non-commercial basis, an understanding of effective instructional processes along with the use of technology to ensure reading success for all learners. HER® was based on four years of research and development (Layng, Twyman, and Strikeleather, 2003) and has been designed to teach beginning readers skills and strategies using phonics instruction. It includes instruction in phonemic awareness, print awareness, phonics, sounding out, segmenting, blending, and reading with comprehension. This is consistent with recommendations from the Rose reviews (Rose and DCSF, 2009) and the EEF) SEND guidance report (EEF, 2021) delivered in 100 episodes. HER® utilises highly effective instructional principles (employed in Direct Instruction; Schieffer, *et al.,* 2002; Kinder, Kubina, and 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 >85% of the words in which they appear (e.g. s, ee, an, ish, out, ing, old). Fluency practice then allows for mastery of decoding strategies before introducing less stable elements (e.g. k, oo, ay, ow). HER® also employs sophisticated adaptive learning technology—instruction adapts to individual responses, providing additional instruction or practice, and high levels of response and feedback. It does not, however, feature in the list of validated systematic synthetic phonics (SSP) programmes published by the Department for Education (DfE) as part of their phonics validation process in April 2021. There is no statutory requirement for schools to choose one of the programmes from this list, but validation status demonstrates that a programme meets the DfE criteria for an effective SSP programme.

HER® was developed, and has mostly been tested, in groups of children in mainstream schools (aged four to seven) (Tyler, *et al.,* 2015a). In addition to the formative evaluations conducted by the programme developers, randomised studies evaluating HER® with children aged four to seven in mainstream schools (Tyler, *et al.,* 2015a, Huffstetter, *et al.,* 2010; Twyman, Layng, and Layng, 2011) found those receiving HER® demonstrated greater improvements in reading than those receiving business as usual. Where available, effect sizes indicated small to large effects across most measures of reading accuracy and word recognition in these studies. For example, in one study, Cohen's d effect sizes ranged from 0.34 to 1.67, with the majority of the effect sizes being close to or above 1 (Tyler, *et al.,* 2015a).

Recently, the programme has been delivered with children with SEND (aged 5 to 19) and shown similar results—confirming that it is also possible to teach early reading skills to children with SEND (Grindle, *et al.,* 2013; Tyler, *et al.,* 2015b). In a recent feasibility and pilot research study employing a pre-test, post-test group design with 26 participants, half were randomised to receive HER® and half did not (Roberts-Tyler, Hughes, and Hastings, 2019). Researchers explored and trialled important aspects of a randomised controlled trial (RCT) evaluation to inform a full-scale RCT with children with ID in special schools in the United Kingdom (UK). In addition to informing the design of a future study (e.g. 85% retention, providing a longer intervention period to map onto school timetables, and providing specific training and support for use with pupils with SEND), there was also a significant difference between reading outcomes for those in the HER® group compared with 'education as usual', with large effect sizes for reading accuracy and increases in reading age in favour of the HER® group. Although this is a positive indication of potential effects of the programme, the main focus of this small-scale feasibility and pilot research was to inform the feasibility of a future RCT.

In Grindle *et al.* (2021), 55 pupils attending a single school for children with severe intellectual and developmental disability were randomly assigned to either an HER® group as supplementary instruction, or a waitlist control. Phoneme segmentation fluency (PSF) scores at follow-up were 1.82 times higher in the intervention group (p=.012) in comparison to the reading-as-usual group. Nonsense word fluency (NWF) scores at follow-up were 2.27 times higher in the intervention group (p=.006) when compared to the reading-as-usual group.

In the above studies with children with SEND, HER® was either delivered or supported by a team of trained researchers and included additional support strategies specifically targeting the additional needs of children with SEND. These strategies have been manualised into an accompanying support manual for teaching staff supporting children using HER®. Having shown that it is possible to teach children with SEND early reading skills in small pilot studies, Headsprout Early Reading® in Special Schools (HERiSS) aims to evaluate whether it is possible to deliver HER® at a larger scale by special schoolteachers and teaching assistants (TAs).

Using a mixed methods approach (collecting quantitative and qualitative data as outlined below) this study aimed to test the effectiveness of HER® in special schools for pupils aged between 5 and 11, Key Stage 1 and Key Stage 2, who are lacking in prior reading skills, delivered by teachers/TAs. Bangor University was responsible for recruiting participants and delivering the programme and the University of Warwick was responsible for its evaluation. Bangor University originally planned to recruit 110 special schools across England. This was ambitious and the recruitment rate was below target even before the nationwide Covid-19 lockdown and school closures, largely because of the length of time taken (up to 91 days) between initially contacting a school to sending out a Memorandum of Understanding (MOU), and up to a further 56 days to receive the signed MOU (see section 'Recruitment and Retention' below). In April 2020, following discussions with the EEF, the study was postponed for one academic year, with recruitment paused during lockdown. When recruitment resumed in late 2020, the target was reduced to 60 schools in recognition of the ongoing pressures on schools caused by the pandemic and in particular the increased difficulties of recruitment given staff absences. In total, 61 schools signed the MOU, 56 schools went on to obtain parental consent for all pupils put forward for the study and were therefore fully recruited, one school withdrew prior to baseline data collection and one school post-randomisation, leading to 54 schools remaining. Schools were randomly assigned to either an intervention or teaching-as-usual group. The evaluation team collected pre- and post-test data from all pupils put forward for the study, which measured their reading skills and their reading self-concept. The process evaluation included getting teachers' and pupils' views about their experiences of using HER® as well as asking pupils about their feelings about reading both before and after HER®.

As noted above, this study was designed before the Covid-19 pandemic. In addition to the pausing of recruitment in April 2020 and postponement of the intervention for one academic year, the pre- and post-data collection procedures from pupils as well as teacher and pupil interviews were adapted to be able to be conducted online. The process evaluation was augmented by adding questions to the teacher interviews and teaching-as-usual surveys and also adding interviews with the implementation support officers (ISOs) to understand any implementation challenges in an environment that was still affected by the global pandemic. No other adaptations to the study were needed for delivery of the intervention in a Covid environment.

This was one of the first large-scale cluster RCTs to explore the effects of phonics instruction on the reading skills of pupils with SEND. It was also the first to ask Key Stage 1 and Key Stage 2 pupils with SEND about their experiences of using HER® and has helped us to understand how pupils feel about using the programme.

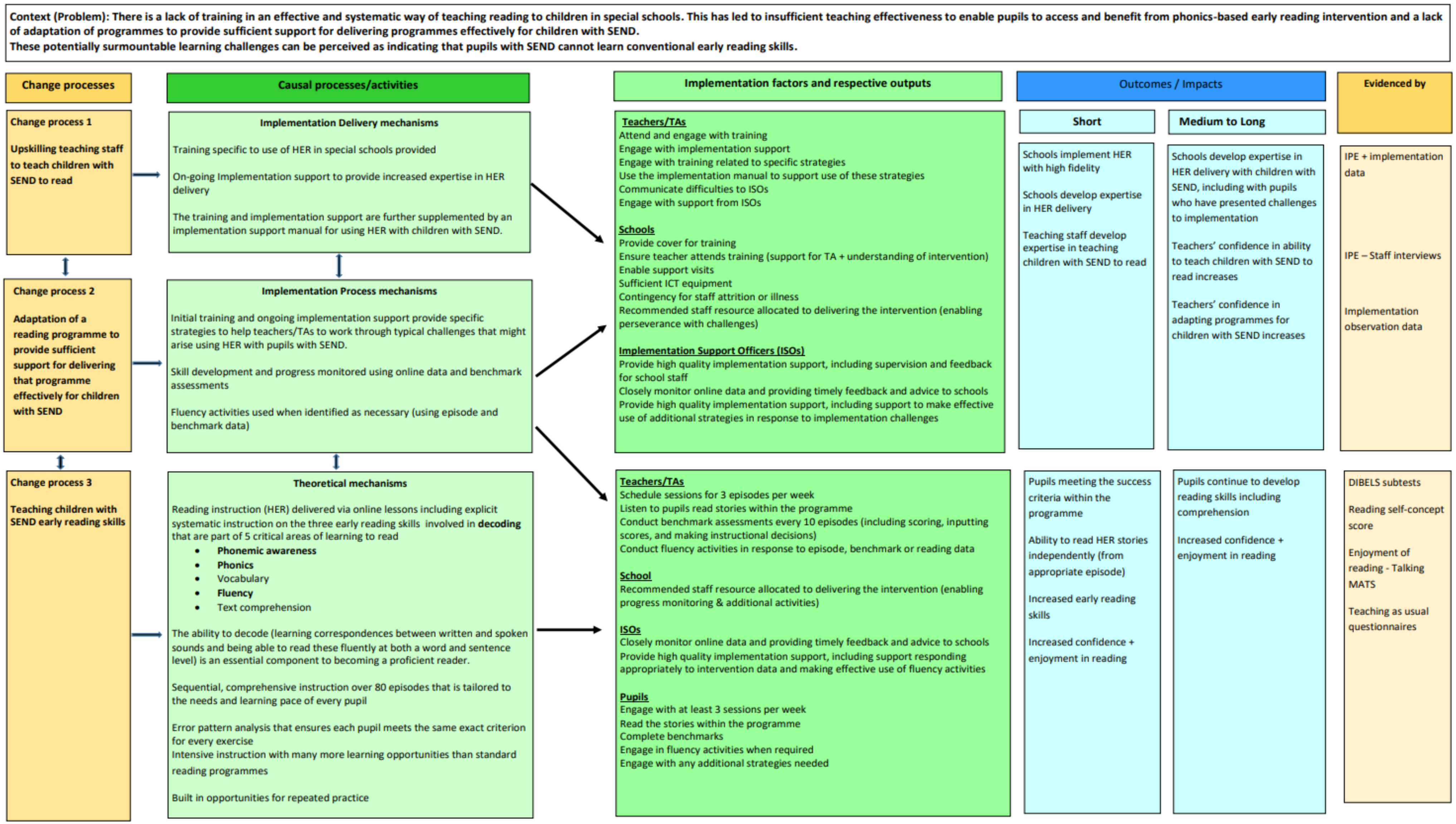
Intervention

HERiSS involved evaluating the effectiveness of an intervention using HER® delivered by staff in special schools (for more information about HER® see https://www.headsprout.com/main/ViewPage/name/headsprout-early-reading). The intervention began following training in September 2021 and continued through until the end of May 2022. The intervention was aimed at pupils aged between 5 and 11 years in Key Stage 1 and Key Stage 2 who were identified by school teaching staff as lacking early reading skills.

The intervention involved three change processes (see Figure 1):

- Change process 1: Upskilling teaching staff to support children with SEND to read;
- Change process 2: Adaptation of a reading programme to provide sufficient support for delivering that programme effectively for children with SEND; and
- Change process 3: Teaching children with SEND early reading skills.

Figure 1: Logic model



DIBELS®=Dynamic Indicators of Basic Early Literacy Skills®; HER=Headsprout Early Reading®; IPE=implementation and process evaluation; ISO=implementation support officer; SEND=special educational needs and disability; TAs=teaching assistants.

There were four key components to the intervention: training; implementation; additional support; and ongoing support. These include the causal mechanisms (delivery, process, and theory) needed to affect change.

Training

This component involved training staff (teachers/TAs) to oversee the delivery of HER® to children in their school. Digital training resources developed by the delivery team (seven core videos and five additional support videos recorded on YouTube) were provided to all schools in the intervention group, which took staff approximately three to four hours to complete. Following this, schools were offered a choice of dates and were invited to each attend two live online and interactive workshops to follow-up on key aspects of programme delivery. Each of these were approximately two hours. The main objective of the first workshop was to introduce schools to the programme, the delivery team, and ISOs, to run through the digital training resources and how to use them, and to take participants step by step through a typical HER® session. The second workshop provided a summary of frequently asked questions/reminders from the previous workshop and included a section on troubleshooting and working through challenges. Both workshops gave participants opportunities to ask questions and raise any concerns schools may have had. Workshops were led by the delivery team and attended by at least two members of staff from each intervention school. Four ISOs attended at least one of each workshop and one ISO (who joined the study in January 2022) watched recordings of two workshops. A member of the evaluation team attended one of each of the two workshops as an observer. Ad hoc training was provided to schools where, for example, identified staff were ill and missed the main training delivery. All schools in the intervention group were offered a choice of dates for the training completed in September 2021. All members of staff involved in the delivery of HER® were expected to engage with the digital training resources and at least one member of staff per school was expected to attend the two webinars. Any staff turnover directly involving the individuals allocated to deliver the programme would receive access to the digital training resources and be given a training session during a school visit. Schools were not able to begin to use HER® until they confirmed that all staff involved in HERiSS had accessed the digital training resources and at least one member of staff had attended the two webinars. One member of the evaluation team attended one session of each of the two webinars provided to help inform the implementation and process evaluation (IPE).

Implementation

This component involved the delivery of HER® between 5 and 15 pupils per school. Based on a combination of small group episode delivery and 1:1 work as required, implementation of HER® with the maximum 15 pupils required between 10 and 20 hours per week. It was therefore necessary for it to be timetabled accordingly and for two staff members to be allocated to HER® delivery for this study, to enable reliable implementation each week.

HER® involves sequential, comprehensive instruction that is tailored to the needs and learning pace of every pupil. Pupils work through activities in an online programme, which adapts instruction in response to their answers. Activities are designed to be engaging and resemble computer games. There are 100 computer-delivered lessons and, depending on individual children and their needs, sessions typically take between 10 and 30 minutes. Stories are available throughout the programme and can be read within the online episodes as well as outside of the Headsprout® sessions (these are available as printable books). Benchmark assessments take place after every ten episodes and take the form of a story. School staff listen to each child read the story individually and rate the reading. These ratings and other monitoring activities are used to decide if the child needs to work on the additional 1:1 fluency activities.

Reading instruction includes explicit systematic instruction on the three early reading skills involved in decoding that are part of five critical areas of learning to read:

- phonemic awareness;
- phonics;
- vocabulary;
- fluency; and
- text comprehension.

Following benchmark assessments and based on previous research findings (Grindle, *et al.,* 2013; Tyler, *et al.,* 2015b) some pupils may require additional 1:1 fluency activities. These fluency activities include specific aims (e.g. 30 words of passage read correctly in a minute), which increase as the programme progresses to ensure true mastery of the reading skills being taught.

The extent to which this is different to usual practice varies across schools. Many special schools use reading programmes that include some form of phonics instruction, although very few have had training specific to reading instruction with children with SEND. However, they are unlikely to be using programmes in which the core instruction is delivered via the computer, employing explicit fluency-based instruction in phonics, and decoding strategies. They are also unlikely to be using a programme that has been piloted specifically with children with SEND, and which includes support strategies relevant to their pupils.

Additional support

This component involves the provision, to all schools in the intervention group, of an implementation support manual specifically designed for using HER® with children with SEND; this provided additional support for high-quality implementation such as suggestions for additional activities where pupils are having difficulty with attending, motivation, or specific concepts (e.g. negation), and included suggestions for additional 1:1 fluency exercises.

Ongoing support

This component involves ongoing implementation support provided by Implementation Support Officers (ISOs) via a combination of in-person (when possible) and online/telephone support. There were five ISOs in total working on this project. ISOs received training in both HER® delivery and effective implementation support for HER®. They were in close contact with the delivery team lead and received fortnightly supervision meetings to discuss implementation challenges in their allocated schools. Support provided by ISOs included supervision (including *in situ* observation, and where appropriate working with a pupil to demonstrate suggested supplementary activities) and feedback for school staff, effective use of additional strategies in response to implementation challenges, and the close monitoring of pupil progress using online software data (captured by HER® on a pupil-by-pupil basis) providing timely feedback to schools in respect of this.

The support model involved schools receiving fortnightly supervision, including one school visit (when possible) every four weeks.

Those schools allocated to the HER® arm implemented HER® across one school year, post-randomisation, and at least three members of staff attended the training (see above).

The intervention assumed the ability of schools to provide the infrastructure and resources necessary to support implementation including timetabling HER®/TA time, information technology (IT) support and availability, the motivation of teachers/TAs to deliver the intervention, and the motivation and ability of the pupils to participate in the intervention. The implementation support was designed to monitor and address issues in relation to these assumptions.

Those schools in the control group (business as usual) continued to offer reading instruction as usual. Control group schools received two payments totalling £1,000: £250 on completion of pre-test assessments; and £750 on completion of post-test assessments.

All schools were asked to sign an MOU (Appendix G). This outlined the responsibilities of each school and their staff members, whether in the intervention or control group. This aimed to ensure that schools in the control group did not attempt to buy HER® and implement it themselves during the school year, and that schools in the intervention group allocated the resources and commitment needed for implementation with fidelity. This was monitored via the teaching-as-usual questionnaire administered both before and after the intervention period.

Table 3: Template for Intervention Description and Replication (TIDieR)

Aspect of TIDieR	Exemplification relating to the evaluation
Brief name	Headsprout Early Reading® in Special Schools (HERiSS)
Why: Rationale, theory, and / or goal of essential elements of the intervention	Children with special educational needs and disability (SEND) have been taught early reading skills in small pilot studies using an online reading programme Headsprout Early Reading® (HER®) HERiSS aims to evaluate whether it is possible to deliver HER® at a larger scale by teachers/teaching assistants (TAs)
Who: Recipients of the intervention	Recipients of HERiSS: Special school staff who have been identified to be trained to support pupils with early reading skills Recipients of HER®: Pupils in special schools aged between 5 and 11 years in Key Stage 1 and Key Stage 2 who have been identified by school teaching staff as lacking early reading skills (see 'Participant selection' below for details)
What: Physical or informational materials used in the intervention	HER®: a computer-based, targeted reading programme Implementation support manual: specifically designed for using HER® with children with SEND
What: Procedures, activities, and / or processes used in the intervention	Teacher/TA training in the use of HER® The delivery of HER® three times per week per pupil Additional activities where needed as set out in the implementation support manual Teacher/TA participation in bi-weekly support sessions (by telephone/video call and *in situ* where possible once per month) delivered by implementation support officers (ISOs)
Who: Intervention providers / implementers	Trained teachers/TAs were responsible for implementation assisted by ISOs provided by the delivery team
How: Mode of delivery	Online computer programme with support provided by teachers/TAs
Where: Location of the intervention	Special schools in classrooms/computer labs depending on school resources (this has not been specified)
When and how much: Duration and dosage of the intervention	Duration: The intervention began following training in September 2021 and ISO support continued throughout until the end of May 2022 Dosage: Defined as the delivery of HER® three times per week per pupil for the duration of the intervention
Tailoring: Adaptation of the intervention	No adaptations have been made to HER® but the implementation support manual provides additional support for high-quality implementation and includes suggestions for additional 1:1 fluency exercises
How well (planned): Strategies to maximise effective implementation	Bi-weekly sessions with ISOs included supervision and feedback for school staff, effective use of additional strategies in response to implementation challenges, and the close monitoring of online data providing timely feedback to schools in respect of this

Evaluation objectives

Primary research question

1) What is the impact of Headsprout Early Reading® (HER®) on the reading skills of Key Stage 1 and Key Stage 2 pupils in special schools?

Secondary research questions

2) What is the impact of HER® on the reading self-concept of pupils (including reading for pleasure/spontaneous reading)?

3) What is the impact of HER® on the different components of reading fluency (e.g. letter naming fluency [LNF], phonemic segmentation fluency [PSF], NWF, oral reading fluency [ORF], and word reading fluency [WRF])?

Additional/exploratory research questions

4) Does the impact of HER® differ with pupils' prior reading skill, age (school year), receipt of free school meals (FSM), type of primary need, and whether English is their first language?

Protocol and statistical analysis plan (SAP) links

A link to the published protocol and SAP can be found on the EEF website (**link here**).

Ethics and trial registration

The British Educational Research Association (BERA) ethical guidelines (2018) and the University of Warwick's strict research Code of Practice were adhered to at all times. Ethical approval was obtained from the Humanities and Social Sciences Research Ethics Committee (HSSREC) at the University of Warwick (dated Wednesday 11 December, 2019 ref: HSSREC 37/19-20).

This trial is registered at the International Standard Randomised Controlled Trial Number (ISRCTN), registration number ISRCTN46208295.

The delivery and evaluation teams abided by the data protection principles set out in the General Data Protection Regulation (GDPR) (GDPR, 2018). Our legal basis for processing personal data was public task in Article 6(1)(e), special category data relying on Article 9(2)(j), and our ethical basis was informed consent in Article 6(1)(a). Before randomisation, an information sheet with an opt-in consent form was sent to parents of pupils selected by schools to participate (see Appendix F). After receiving the information sheet, parents were able to consent to their child participating in testing for the evaluation and having their data included in the analyses. The delivery team worked with schools to encourage parents to return consent forms in a timely manner. In the information sheet, parents were encouraged to contact the evaluation team with any questions about testing. Parents had the option to withdraw their child's data at any point during the trial, and this was made clear in the information sheet.

Both the delivery and evaluation teams had access to the participant database (schools and pupil data), which is encrypted and held securely on departmental servers. Only the evaluation team have access to the trial database (pre- and post-intervention data collected). Participant information was treated confidentially, and all participants were informed of their right to withdraw from testing and/or inclusion of their data in the evaluation at any stage. No person or school was identified in the reporting of this trial.

Data protection

For the purposes of conducting the evaluation to assess the impact of HER®, Bangor University and the University of Warwick were joint data controllers and processors of the personal data of pupils (e.g. pupil names, data from the HER® programme) obtained from schools. The University of Warwick is a data controller for all data collected as part of the evaluation. For more details on the type of personal data and special category data collected for this trial, please see 'Privacy Notice' (Appendix G).

At the end of the trial, the University of Warwick will share the data with the EEF data archive processor through secure data portals, where the data are encrypted and saved on the secure servers for further analysis. At that point EEF will become the data controller and the University of Warwick will no longer have any responsibility for the data.

The legal basis for processing personal data for this project is informed consent under GDPR Article 6(1)(a), 'the data subject has given consent to the processing of his or her personal data for one or more specific purposes'. The delivery and evaluation teams will securely delete all personal data within six months of the project finishing. The University of Warwick will retain the anonymised data from this project until spring 2031 to permit further analysis including special category data on SEND status.

All schools' data is treated with the strictest confidence and will be transferred securely and saved in secure locations only accessible by the delivery and evaluation teams in line with GDPR and the Data Protection Act 2018 (Legislation.gov.uk, 2018). Neither individual participant names nor the names of participating schools will appear in any report arising from this project.

The data will be shared with the DfE, the EEF, EEF archive manager, the Office for National Statistics (ONS), and potentially other research teams, subject to the appropriate approvals. Data will be matched with the National Pupil Database (NPD) for analysis after the trial. Further matching to the NPD, other administrative data may take place during subsequent research to better understand the impact of the project.

EEF will act as the data controller for the archive, which is managed on their behalf by the FFT and held in the ONS Secure Research Service. The archive does not contain any information that can be used to directly identify an individual pupil. For example, the archive does not include names, addresses, or dates of birth. The archive does contain the Pupil Matching Reference (PMR), which is an identifier used by the DfE to enable the linking of archive data to the NPD.

Project team

Delivery team

- Dr Emily Roberts-Tyler, Principal Investigator, School of Educational Sciences, Bangor University, leading school and staff recruitment, training of schools, and provision of implementation support and supervision of ISOs during the trial.
- Professor Carl Hughes, Co-investigator, School of Educational Sciences, Bangor University, part of the advisory group supporting school and staff recruitment, and implementation support during the trial.
- Dr Corinna Grindle, Co-investigator, part of the advisory group supporting school and staff recruitment, and implementation support during the trial, and part of the team training Headsprout® schools, The Centre for Behaviour Solutions.
- Dr Claire McDowell, Co-investigator, part of the advisory group supporting school and staff recruitment, and implementation support during the trial, and part of the team training Headsprout® schools, Ulster University.
- Nationwide team of five ISOs. The ISOs were employed by Bangor University and trained and supervised by the delivery team lead and advisory group. ISOs received training in both HER® delivery and effective implementation support for HER®. They were in close contact with the delivery team lead and received fortnightly supervision meetings to discuss implementation challenges in their allocated schools.

Evaluation team

- Dr Samantha Flynn, Joint Principal Investigator, Centre for Educational Development, Appraisal and Research (CEDAR), The University of Warwick.
- Dr Louise Denne, Joint Principal Investigator, CEDAR, The University of Warwick.
- Professor Richard P. Hastings, Co-investigator, mentor for Dr Flynn and Dr Denne, CEDAR, The University of Warwick.
- Dr Paul Thompson, Lead Statistician, CEDAR, The University of Warwick.

- Dr Rebecca Morris, Co-investigator and Lead on Education, Centre for Education Studies, The University of Warwick.
- Nationwide team of postgraduate ad hoc researchers.

Methods

Trial design

Table 4: Trial design

Trial design, including number of arms		Two-arm parallel cluster randomised controlled efficacy trial
Unit of randomisation		School level
Stratification variable (s) **(if applicable)**		School size (>=70 pupils vs <70 pupils) Based on average school sizes DfE (DfE, 2018)
Primary outcome	**Variable**	Pupils' reading skills (pupils' ability to sound out words—not comprehension)
	Measure **(instrument, scale, source)**	A composite score derived from Dynamic Indicators of Basic Early Literacy Skills®, Eighth Edition (DIBELS®) (pupil testing)
Secondary outcome(s)	**Variable(s)**	Components of reading fluency (e.g. letter naming fluency [LNF], phonemic segmentation fluency [PSF], nonsense word fluency [NWF], word reading fluency [WRF], and oral reading fluency [ORF]) Pupils' reading self-concept (including reading for pleasure / spontaneous reading)
	Measure(s) **(instrument, scale, source)**	Reading Self-Concept Scale (RSCS) (pupil questionnaire) DIBELS® component analysis (pupil testing)
Baseline for primary outcome	**Variable**	Pupils' reading skills (pupils' ability to sound out words—not comprehension)
	Measure **(instrument, scale, source)**	DIBELS® composite score
Baseline for secondary outcome(s)	**Variable**	Pupils' reading self-concept (including reading for pleasure / spontaneous reading) Components of reading fluency (e.g. letter naming fluency, phonemic segmentation fluency, nonsense word fluency, oral reading fluency, and word reading fluency)
	Measure **(instrument, scale, source)**	Reading self-concept scale (Pupil questionnaire) DIBELS® component analysis (Pupil testing)

This study was an efficacy trial that ran over one academic year involving Key Stage 1 and Key Stage 2 pupils in special schools across England. This efficacy trial was a two-armed cluster RCT. Schools were randomly assigned to either intervention or 'business as usual' (control) groups. For the control group, literacy lessons were conducted as normal. Schools receiving the intervention and associated training implemented HER® with a group of eligible pupils (see section 'Sample for eligibility') including access to the delivery team's 'implementation manual' and ongoing support throughout the intervention. For the schools who received the intervention, the intention was that this replace any phonics instruction that would otherwise be implemented with the participating pupils.

The following outcome measures were used in the trial (detail in ‘Outcome measures’ section below):

- The Dynamic Indicators of Basic Early Literacy Skills®, Eighth Edition (DIBELS®); and
- a reading self-concept questionnaire (adapted by the evaluation team from The Reading Self-Concept Scale (RSCS); Chapman and Tunmer, 1995).

Participant selection

Special schools (between 15 and 301 pupils in size) in England were invited to take part. Recruitment was the responsibility of the delivery team. It focused on a geographical spread around the Midlands, and Central and North England (see ‘Recruitment and Retention’ section, below). Schools had to have at least five, and up to 15 pupils at Key Stage 1 and/or Key Stage 2 who met clearly described pupil eligibility criteria. These criteria were provided to schools in the information sheet as well as in discussions between the delivery team and their school contact:

Pupils who did not have a reading ability beyond the level of HER®[2] and could:

- sit at a computer for up to ten minutes;
- understand and follow one- or two-step instructions;
- imitate spoken sounds/words;
- respond to feedback (praise or correction); and
- use some self-initiated speech (in English) (single words to short sentences).

Outcome measures

Baseline measures

Baseline measures were:

- The Dynamic Indicators of Basic Early Literacy Skills® 8th edition (DIBELS®; and
- A reading self-concept questionnaire (adapted by the Evaluation team from The Reading Self-concept Scale; Chapman and Tunmer, 1995).

All baseline assessments were collected during the period April 2021 – July 2021 online by a team of research assistants trained to use each measure and who had also received training in working with children with SEND. Pupils’ characteristics, such as school year, first language, receipt of school meals, and primary need were also collected at baseline from participating schools to inform the exploratory subgroup analyses.

Reading skills

The DIBELS® is a short assessment, taking approximately 20 to 30 minutes to administer per pupil by a trained individual. The Year 2 version of the DIBELS® with the Oral Reading Fluency Component was used as this best corresponds with expected reading ability in Key Stage 2. It assesses fluency in five core components: Letter Naming; Phonemic Segmentation; Nonsense Word; Word Reading; and Oral Reading. The DIBELS® provides data on the number of correct responses per minute to assess the above skills, and as such it can be used to measure improvement in each of the core components. The Early Childhood Research Institute on Measuring Growth and Development (ECRI-MGD) examined reliability and validity of the DIBELS® in a four-year, longitudinal research study and found that all DIBELS® measures displayed adequate reliability (Good, *et al.,* 2004). The DIBELS® testing and scoring materials are

[2] Schools were provided with a short paragraph of text taken from the HER® programme to determine whether pupils had a sufficiently low level of decoding skills.

widely available and free to use, have been used in HER® pilot studies in UK special schools (Tyler, *et al.,* 2015a; Tyler, *et al.,* 2015b) and therefore did not require further piloting for this study (although see below comments regarding piloting of the online data collection process developed for the study).

The DIBELS® is an appropriate assessment as it tests the outcomes (increased early reading skills and functional reading skills) highlighted in the HERiSS logic model (Phonemic awareness, Phonics, and Fluency) and avoids likely limitations of standardised score reading tests with this population.

The evaluation team have used an adapted version of the DIBELS® to enable the assessment to be conducted by a research assistant online with a pupil at school (supported by a teacher or TA). This was necessary because of UK government guidance in respect of the Covid-19 pandemic during the data collection period. In the adapted version, the DIBELS® pupil assessment materials were put onto a PowerPoint presentation replicating the order of letters/words and the fonts of the original materials. The key difference is that letters/words were presented individually (i.e. slide by slide) rather than on a whole page as is the case with materials used face-to-face. The reason for this is that during face-to-face delivery, the assessor pointed to the letters/words one by one as the pupils progressed through the testing materials. This was not possible remotely. The materials for on-line delivery were developed and piloted by a team at the University of Warwick for another HER® study. To check the suitability for delivery of the materials in a school setting the materials were further piloted with four pupils who met the eligibility criteria for HERiSS at a special school in London. To the best of our knowledge this is the first study to deliver an adapted version of the DIBELS® online.

Reading self-concept

A brief (less than five minutes) survey of reading self-concept was undertaken with pupils pre- and post-intervention. As no such measure suitable for children in special schools existed, an adapted version of an existing survey (The Reading Self-concept Scale; Chapman and Tunmer, 1995) was developed by the evaluation team to include scales with adjusted response formats (e.g. thumbs up/thumbs down) and suitable for delivery online was used (Appendix D). The self-concept subscale includes six-items across three categories: competence; attitude; and difficulty (range 0–12). This adapted version was piloted by the evaluation team in a special school in London.

Primary outcome

The composite score of DIBELS® was used as the primary outcome to gather data on pupils' reading skills. It was used to answer the primary research question 1 and partly answer secondary research questions 3 and 4. The DIBELS® includes several versions of each of the component measures to be used in case of any interruptions during initial testing (e.g. fire alarms) and for post-data collection purposes.

The composite score is calculated following the DIBELS® composite score calculation guide supplement. Given the ability of children within special schools does not translate to an equivalent age-adjusted score for their peers in mainstream schools, all children were scored at the 'First Grade' level, which is the UK equivalent of Year 2 primary school (ages six to seven). The scoring can be calculated as follows:

Step 1. Multiply each subtest raw score by the weight listed.

Subtest	Raw score	Weight	Weighted score
Letter naming fluency (LNF)		x 10.72	=
Phonemic Segmentation fluency (PSF)		x 2.13	=
Nonsense Word Fluency – Correct Letter Sounds (NWF-CLS)		x 23.13	=
Nonsense Word Fluency – Words Read Correct (NWF-WRC)		x 7.79	=
Word Reading Fluency (WRF)		x 13.51	=
Oral Reading Fluency – Words Read Correct (ORF-WRC)		x 25.36	=
Oral Reading Fluency – Accuracy (ORF-ACC)		x 0.25	=

Step 2. Sum the weighted scores from Step 1.

Step 3. Subtract the mean of the weighted score from the sum of the weighted scores.

$\_\_\_\_\_\_\_\_ -3{,}371 = \_\_\_\_\_\_\_\_$
(value from Step 3)

Step 4. Divide value from Step 3 by standard deviation (SD).

________________ $\div 2{,}251 =$ ____________________
(value from Step 3)

Step 5. Multiply value from Step 4 by 40 and round to the ones place.

________________ $\times 40 =$ ____________________ (rounds to ones place).
(value from Step 4)

Step 6. Add the scaling constant for the season in which the student was tested to obtain the final composite score.

Constants: Fall/Beginning = 360, Winter/Middle = 400, Spring/End = 440 (note all participants were collected in Spring/End)

________________ $+$ ______________= ____________________
(value from Step 5) (constant) (final composite score)

In the case of discontinuation, a zero score will be associated with any component after discontinuation. It is still possible that missing values are recorded in the dataset. For the primary analysis and to permit calculation of composite scores, any missing values in the component (subtest) scores were recoded as zero, effectively making no contribution to the total score. If all component scores were missing, the composite score at that timepoint was missing. When running the multiple imputations as sensitivity analysis, any missing scores were imputed (as detailed later). This allows us to compare the influence of true zeros due to no points scored or a discontinuation from genuinely missing data at random. This process as detailed in the results has the effect of causing a shift in the distribution away from zero as missing data may have real values other than zero. This decision was discussed with both the research and delivery teams as the most appropriate course of action given the process taken for data collection and when a true missing was reported.

Secondary outcomes

The reading self-concept measure was repeated post-intervention and used as a secondary outcome measure to in part answer secondary research question 2. Baseline and post-intervention data collection was approximately one year apart. The measure is short and asks questions about a child's feelings about and attitudes towards reading. There are no right answers, and the risk of testing effects is low.

All DIBELS® components were repeated post-intervention and used as secondary outcome measures to address secondary research question 3 and part answer secondary research question 4.

All baseline and post-intervention assessments were conducted online by a team of research assistants, blind to the group allocation, trained to use each measure, and who had also received training in working with children with SEND. Data were collected by research assistants at the time of delivery of each measure. Scores were recorded on a paper-based scoresheet and research assistants also inputted the scores to a Qualtrics survey, which were downloaded into an excel spreadsheet. At the end of the data collection all paper-based score sheets were checked by an independent research assistant against the excel spreadsheet. There were no recorded instances of unblinding.

Sample size

Sample size calculations were conducted using R version 4.1.2 (2021-11-01) and PowerUpR version 1.1.0 (1,2,3).

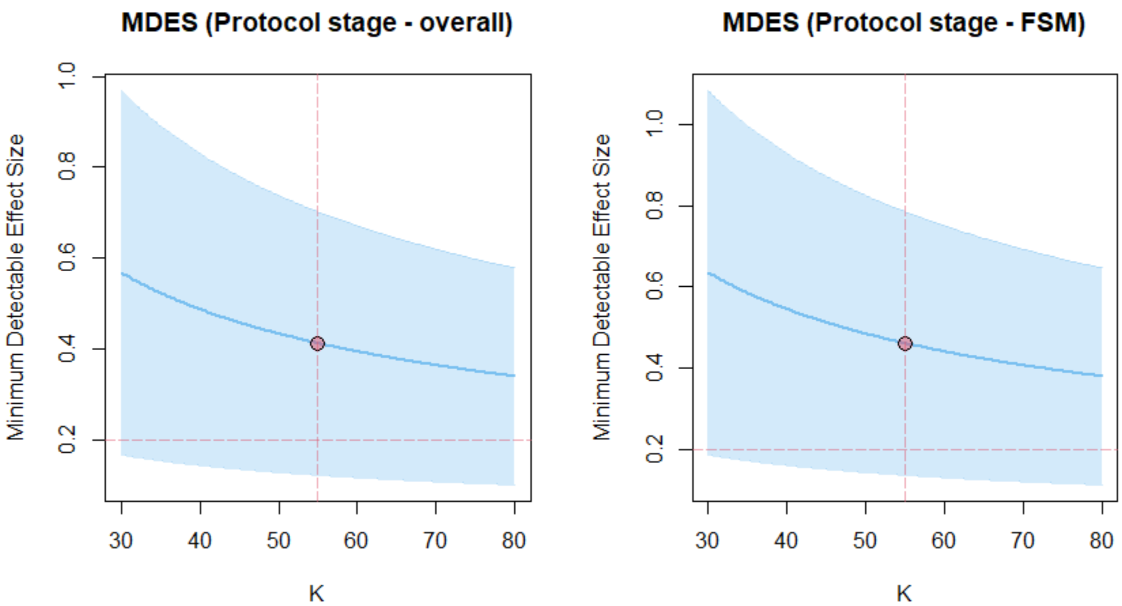
Working with the delivery team and previous experience from testing in special schools, we conservatively estimated that 15% of all pupils in a special school may be eligible for HER®, and so we estimated that we would be able to recruit between 5 and 15 primary age pupils from most special schools in selected regions in England (North West and West Midlands). Prior to assessments moving online due to the Covid-19 pandemic, these regions were initially selected for logistical reasons as it was deemed feasible for the delivery team to support schools across these regions, and for the evaluation team to conduct assessments across these regions. We anticipated that this number of pupils will be manageable for two staff (a trained teacher and TA) per school delivering HER®.

We aimed to randomise 55 schools (28 schools in the control arm and 27 schools in the business as usual arm of the study). Calculations are based on 80% power, a (two-sided) significance level of 5%, with (following pupil-level attrition, see below) an average of 2.33 pupils within each class, and an average of three classes within each school at follow-up.[3] Although this was relatively unknown in the case of special schools, we cautiously accounted for an intracluster correlation coefficient (ICC) as high as .40 at the class level and .10 at the school level, in accordance with guidance (Morena, 2011; Storey, McDowell, and Leslie, 2020). Demack (2019) provides a theoretical and empirical review of three-level cluster randomised trials in education with the purpose of exploring the implications of ignoring classroom-level clustering. There are few special school education trials, and none with sufficient sample size to obtain accurate estimates of appropriate ICCs. The current study is the largest education trial in a special school context to date. Therefore, we assumed estimates of mean ICCs (primary) recommended from the Demack (2019) review's discussion as conservative moderated by expert knowledge of the special school environment by both delivery and evaluation teams. We allowed for an estimated correlation of pre- and post-intervention scores (both on the same DIBELS® composite test) on the outcome of .5 for level 1, and 0 for levels 2 and 3, representing a modest or no correlation, respectively between the two timepoints (Dong and Maynard, 2013). These figures represent a conservative position on the effect size able to be detected. The power calculation resulted in a minimum detectable effect size (MDES) of .412 (Figure 2, left plot).

For the free school meals (FSM) sub-group analysis, we estimated an MDES of .465 (Figure 2, right plot); based upon the above assumptions, and official statistics showing 43% of FSM pupils in state-funded special schools.[4]

We allowed for a 10% attrition rate for schools (for both groups), and one pupil per school lost to attrition (11.1% pupil-level attrition from average of eight pupils initially [slightly lower than the originally expected ten] recruited per retained school) in retained schools across the trial. Our dropout rate was considered as a somewhat conservative estimate, as while both lower and higher rates of attrition have been observed in extant studies of the HER® intervention (Grindle*, et al.,* 2021; Moerbeek and Schie, 2019), this was a relative unknown with respect to large-scale studies in a special school setting. We prepared the sample size calculations under the assumption that 60 schools would be recruited (with 54 schools retained) and eight pupils recruited within each school (with an average of seven pupils retained) to achieve a sample of 378 pupils (after attrition at both the school and pupil level).

Figure 2: Plot showing the minimum detectable effect size (MDES) as a function of number of schools (K)



[3] There are no official published figures on the number of pupils from special schools that schools can put forward for studies. The delivery and evaluation teams made the estimate that schools would have at least five eligible pupils to participate and that they could put forward and up to 15 pupils. We then conservatively reduced this number at randomisation stage to seven pupils based on recruitment estimates.
[4] see https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics

Randomisation

Prior to randomisation, schools were asked to sign an MOU (as suggested in the EEF Recruitment and Retention pack). This included permission to access data gathered by HER®, consent to be randomised, and commit to the outcome (intervention or control), allowing time and space for testing, and ensuring three staff could attend training and deliver HER® (where appropriate).

Following baseline testing (May 2021 – July 2021), schools were allocated to the intervention or control groups (1:1) through a process of minimisation (adaptive stratified sampling; Altman and Bland, 2005). This is an effective procedure for small sample sizes that also ensures that intervention arms are balanced with respect to predefined factors as well as for the number of participants in each group. Additionally, this allowed researchers to immediately notify participating schools of their allocation, facilitating planning for the start of the intervention in September 2021 (i.e. the start of the new academic year) and minimising waiting times. The Minim software was used to individually allocate schools via the minimisation process to one of the groups (Bland, 2017). One key balancing factor was used—school size, which was dichotomised using a cut-off based on the median value from the DfE (2018) data. This assumed that larger special schools had at least 70 pupils and smaller schools fewer than 70 pupils.

Minimisation was completed by a member of the evaluation team who remained independent from delivery, testing, and statistical analysis, as recommended in the EEF Randomised Trials in Education document.

Statistical analysis

Primary analysis

The primary analysis examined mean follow-up DIBELS® composite scores, adjusting for the respective baseline measure, school size, and with the covariate of interest specified as a dichotomous intervention/control variable (Kahan and Morris, 2012).

The analysis for the primary outcome measure used a linear mixed model (LMM), given that the measure was continuously distributed. Some skew was observed in the raw outcome measure, so the models were re-run with robust estimators to ensure standard errors (SEs) were estimated as accurately as possible.

Nested model comparison was based on likelihood ratio tests (Chi square) and both Bayesian and Akaike's Information Criteria (BIC and AIC, respectively), with lower indices indicating the preferred model. All analyses were conducted in R (version 4.0.3 - 2020-10-10), using the R packages: Tidyverse (Wickham *et al.,* 2019), lme4 (Bates *et al.,* 2015), glmmTMB (Brooks *et al.,* 2017), and lmerTest (Kuznetsova *et al.,* 2017).

Following the plan outlined in the SAP, due to the number of class/school year clusters within school being relatively small (Gelman and Hill, 2006; Hayes and Moulton, 2017; Raudenbush and Bryk, 2002), we adopted a two-step process to explore issues with over- or under-fitting using a sensitivity analysis. In the first step, we fitted a two-level design (pupil – level 1; school – level 2), then in a second analysis, we included an additional level (pupil – level 1; class/school year – level 2; and school – level 3).

Step 1 model (two-level design):

$$\boldsymbol{L1}: Y_{ik} = \beta_{0k} + \beta_{1k} X1_{ik} + r_{ik}, \ r_{ik} \sim N\left(0, \sigma^2{}_{|X}\right)$$

$$\boldsymbol{L2}: \beta_{0k} = \gamma_{00} + \gamma_{01} X2_k + \gamma_{02} INT_k + \mu_{0k}, \qquad \mu_{0k} \sim N(0, \tau_2^2)$$

$$\beta_{1k} = \gamma_{10}$$

where, Y are the DIBELS® composite scores; $X1$ are the baseline DIBELS® composite scores; INT is the intervention/control variable; $X2$ are indicator of school size with dichotomous split at n=70 pupils' threshold (strata variable). μ_{0k} is the random intercept term for school.

Step 2 model (three-level design):

$$\boldsymbol{L1}: Y_{ijk} = \beta_{0jk} + \beta_{1jk} X1_{ijk} + r_{ijk}, \ r_{ijk} \sim N\left(0, \sigma^2{}_{|X}\right)$$

$$\boldsymbol{L2}: \beta_{0jk} = \gamma_{00k} + \mu_{0jk}, \qquad \mu_{0jk} \sim N(0, \tau_2^2)$$

$$\beta_{1jk} = \gamma_{100}$$

$$\boldsymbol{L3}: \gamma_{00k} = \delta_{000} + \delta_{001} X2_k + \delta_{002} INT_k + \varsigma_{0k}, \ \varsigma_{0k} \sim N(0, \tau_3^2)$$

where, Y are the DIBELS® composite scores; $X1$ are the baseline DIBELS® composite scores; INT is the intervention/control variable; $X2$ are indicator of school size (strata variable). μ_{0jk} and ς_{00k} are the random intercept terms for class and school respectively.

Secondary analysis

Components of reading fluency

Given that the distribution of most of the secondary outcome measures were positively scored, integers, and skewed, secondary outcomes have been analysed using generalised linear mixed models (GLMMs; Poisson or negative binomial family depending on overdispersion) for each of the five components of reading fluency (e.g. letter naming fluency, phonemic segmentation fluency, nonsense word fluency, oral reading fluency, and word reading fluency) (Barnett and Dobson, 2008; Gelman and Hill, 2006).

GLMM two-level design

$$\boldsymbol{L1}: g(Y_{ik}) = \beta_{0k} + \beta_{1k} X1_{ik} + r_{ik}, \ \ r_{ik} \sim N\left(0, \sigma^2{}_{|X}\right)$$

$$\boldsymbol{L2}: \beta_{0k} = \gamma_{00} + \gamma_{01} X2_k + \gamma_{02} INT_k + \mu_{0k}, \qquad \mu_{0k} \sim N(0, \tau_2^2)$$

$$\beta_{1k} = \gamma_{10}$$

Note: $g(.) = log_e(.)$, where $g(.)$ is the log link function for the Poisson GLMM

where, Y are the DIBELS® composite scores; $X1$ are the baseline DIBELS® composite scores; INT is the intervention/control variable; $X2$ are indicator of school size (strata variable). μ_{0k} is the random intercept term for school.

Models for these five subscales report incident rate ratios (IRRs) (exponentiating the parameter estimates from the model). The IRR is an effect size measure reported when the outcome is a count or if the score is positive and has some positive skew with integer responses. This effect size is interpreted as the comparison of incidence of events happening at different times. For example, when considering the secondary outcomes in this study, we compare the difference in counts pre- and post-intervention. For each outcome, these models include the covariate of interest specified as a dichotomous variable based on whether the pupil was assigned to the intervention or control group. In addition, covariates for the baseline scores for the respective outcomes and the school size covariate are also included.

Reading self-concept

Similarly, to the primary outcome measure, a general linear mixed model (GLMM) has been used to predict the secondary outcome of pupil reading self-concept (including reading for pleasure/spontaneous reading). The reading self-concept measure is a bespoke measure adapted by the evaluation team based on an existing measure. The measure contains six-items with each item scoring either ‘yes’ or ‘no’ (coded 1 or 2 according to whether reverse coded or not) therefore the total score use in the analysis will range between 0 and 12. For the reading self-concept measure, Hedges’ g will be reported with 95% confidence intervals (CIs).

The reading self-concept measure has used an linear mixed model rather than an ordinal mixed model as specified in the SAP. This was due to a misunderstanding of the scoring criteria for the reading self-concept. Within the SAP, scoring was considered to provide a total score between 0 and 6, so an ordinal model seemed most appropriate. However, after review with the evaluation team scoring was corrected to be between 0 and 12; therefore, an ordinal model was not suitable for this outcome. Instead, we have followed the same analysis as the primary outcome and reported accordingly. The distribution of the outcome is presented in Appendix C.

Analysis in the presence of non-compliance

To ascertain the influence of non-compliance on the predictions made by the intention-to-treat (ITT), an instrumental variable approach was used. Two binary compliance indicators, full compliance, and partial compliance, defined as exceeding a proportion of the six binary response questions for compliance were specified. A separate instrumental variables analysis was conducted for each derived compliance measure to investigate the influence of full compliance and partial compliance (Table 5).

Compliance binary response questions were as follows:

All teachers/TA's supporting pupils through HERiSS had been through the digital training (this item was scored dichotomously 'Yes/No').

1. At least one member of staff attended the two webinars (this item was scored dichotomously 'Yes/No').

2. At least two trained members of staff oversaw implementation to the end of the academic year (this item was scored dichotomously 'Yes/No').

3. Schools engaged with ISO provided fortnightly supervision (attending at least 70% of scheduled sessions) and with monthly ISO visits in schools (attending 70% scheduled sessions) (this item was dichotomously scored, 'Yes' if 70% of supervision were held and 70% of ISO visits in schools were attended, and 'No' otherwise).

4. Teachers followed recommendations from ISO sessions including using the activities outlined in the HER® manual (measured using a 4-point rating completed by ISOs following each visit). This item was turned into a dichotomous measure with a 4 (used most of the time) as 'Yes' and 1–3 (no evidence, rarely used, sometime used) being 'No'.

5. Timetabling three HER® sessions per pupil per week (this item was scored dichotomously scored 'Yes'/'No').

Table 5: Compliance definitions for compliance analyses

<table>
<tr><th>Compliance items</th><th>Full</th><th>Partial (scenario A)</th><th>Partial (scenario B)</th></tr>
<tr><td>1</td><td rowspan="6">6 out of 6</td><td rowspan="3">2 out of 3</td><td rowspan="5">4 out of 5</td></tr>
<tr><td>2</td></tr>
<tr><td>3</td></tr>
<tr><td>4</td><td rowspan="2">1 out of 2</td></tr>
<tr><td>5</td></tr>
<tr><td>6</td><td>ü</td><td>O</td></tr>
</table>

Compliance item 6 is the only measure that is HER®-specific and recommended by the Headsprout® team. Compliance items 1 to 3 relate to training and 4 and 5 relate to ISO support and the additional strategies (HERiSS-specific) that are a part of the HERiSS intervention.

Partial compliance for scenario A includes item 6 and at least one from each of the training and additional support measures (to a total of 4).

Partial compliance for scenario B includes four out of five of items 1–5. This scenario is included as compliance item 6 is not entirely within the control of schools—three sessions per week per pupil may be timetabled but schools cannot guarantee that pupils will be present/engage in the timetabled session on a particular day.

As recommended by the EEF, a two-stage least squares approach was used to estimate the model and Huber-White SEs are reported, which are robust to clustering (EEF, 2018). The R packages 'ivpack' and 'ivreg' were used to implement the two-stage instrumental variable analysis (Jiang and Small, 2014; Fox, Klieber, and Zeileis, 2021).

Compliance will be instrumented by the intervention allocation (Angrist and Imbens, 1995). The stage 1 model is defined as follows:

$$Compliance_k = \beta_0 + \beta_1 Treatment_k + \varepsilon_{jk}$$

Predicted values for, $Compliance_k$, from the stage 1 model will be included in the stage 2 model, as follows:

$$Y_{ik} = \beta_0 + \beta_1 \widehat{compliance}_k + \beta_2 baseline_{ik} + \beta_3 schoolsize_k + r_{ik}$$

Missing data analysis

In the SAP, we stated that if over 5% of cases were missing, a generalised linear mixed model would be fitted, specifying a binary dependent measure of missingness. This model's purpose was to give some indication whether particular covariates were correlated to missingness and whether their inclusion into the imputation model would be beneficial. It was plausible that data could have been missing at random (MAR) based on: type of primary need; English as a first language (EFL); and baseline pupil reading ability; and therefore, in that circumstance these variables would be used as covariates in the imputation model. Given our assumption that our data was MAR, a multiple imputation model was run using both primary analysis covariates (van Buuren, 2018; Austin *et al.,* 2021; Plumpton *et al.,* 2016) and the secondary outcomes as components of the composite are likely correlated. A sensitivity analysis compared the imputed model to the ITT analysis (Gelman and Hill, 2006). The amount of missing data in the baseline was 7.98% and at follow-up, 26.68%. Therefore, we implemented a multiple imputation approach using chained equations.

We found that the imputation models needed to be simplified by removal of the random intercept (school-level clustering) as the variance was zero to close-to-zero. Imputations were based on zero-inflated negative binomial regressions for the DIBELS® components. Once the components were imputed, the composite score for the primary outcome was calculated for each imputed dataset and analysed. Twenty imputed datasets were generated (following recommendations by van Buuren, 2018), and analysed results were pooled following Rubin's rules (Rubin, 1996). Trace plots were generated to assess convergence and have been include in Appendix E.

Subgroup analyses

Subgroup analyses evaluated the impact of the intervention on the FSM subgroup (subsample), for both primary and secondary outcomes, to ascertain the size of treatment effects for FSM pupils. A further subgroup analysis evaluated the interaction of FSM and the treatment effect on the primary and secondary outcomes; to examine whether the treatment effect was conditional on FSM eligibility. We present the two-level model given that the primary analysis found no evidence for variation explained by the class level.

$$\boldsymbol{L1}: g(Y_{ijk}) = \beta_{0jk} + \beta_{1jk} X1_{ijk} + \beta_{2jk} FSM_{ijk} + r_{ijk}, \quad r_{ijk} \sim N\left(0, {\sigma^2}_{|X}\right)$$

$$\boldsymbol{L2}: \beta_{0k} = \gamma_{00} + \gamma_{01} X2_k + \gamma_{02} INT_k + \mu_{0k}, \qquad \mu_{0k} \sim N(0, \tau_2^2)$$

$$\beta_{1k} = \gamma_{10}$$

Note: $g(.) = log_e(.)$, where $g(.)$ is the log link function for the secondary outcome measures, whereas the primary outcome, $g\left(Y_{ijk}\right) = Y_{ijk}$.

Additional analyses and robustness checks

To assess the robustness of the findings, a number of additional pupil-level covariates have been introduced: pupils' prior reading skill; school size; and type of primary need (dummy coded to indicate whether individuals have this SEND condition or not). The following are covered: Autism Spectrum Disorder (ASD); Specific Learning Difficulty (SpLD); Moderate Learning Difficulty (MLD); Severe Learning Difficulty (SLD); Profound and Multiple Learning Disability (PMLD); Speech, Language, and Communication Needs (SLCN); Physical Disability (PD; or other), English as a first language (EFL), and free school meal status (FSM). These covariates extend the linear mixed model analyses used for the primary outcome using the same model structure. In addition, the interactions between the specified covariates and intervention/control group variable were included into the model to assess whether the treatment effect is conditional on additional covariates.

Estimation of effect sizes

Effect size for the primary outcome measure, DIBELS® composite score is reported as Hedges' g (adjusted mean difference) (Hedges, 2007). Initially, the SAP detailed the effect size for the three-level model but following sensitivity analysis of the two- and three-level models, we defer to the two-level LMM for primary outcome. Therefore, a sample estimate of the effect size equivalent to Hedges' g with 95% confidence interval assuming unequal cluster sizes and using the pooled within-groups standard deviation, is reported. The calculation of the effect size and associated confidence intervals will be calculated using the EEF Analytics R packages for cluster randomised trials (Kasim, Xiao, Higgins and Troyer, 2017):

$$d_t = \frac{\widehat{\beta_1}}{\sigma_T}\sqrt{1-\rho\left(\frac{(N-n_U^T m^T - n_U^C m^C)+n_U^T+n_U^C-2}{N-2}\right)},$$

where $\widehat{\beta_1}$ is the adjusted mean difference in DIBELS® composite score between trial arms; σ_T is the unconditional variance.

$$\sigma_T = \sqrt{\frac{\sum_{i=1}^{m^T}\sum_{j=1}^{n_i^T}(Y_{ij}^T - Y_{..}^T)+\sum_{i=1}^{m^C}\sum_{j=1}^{n_i^C}(Y_{ij}^C - Y_{..}^C)}{N-2}},$$

and,

$$n_U^T = \frac{(N^T)^2-\sum_{i=1}^{m^T}(n_i^T)^2}{N^T(m^T-1)},\ n_U^C = \frac{(N^C)^2-\sum_{i=1}^{m^C}(n_i^C)^2}{N^C(m^C-1)},\ N = N^T + N^C$$

The variance estimator is given by,

$$V(d_t) = \left(\frac{N^T+N^C}{N^T N^C}\right)(1+(\tilde{n}-1)\rho) + \frac{[(N-2)(1-\rho)^2+A\rho^2+2B\rho(1-\rho)]\delta^2}{2(N-2)[(N-2)-\rho(N-2-B)]},$$

Where the auxiliary constants A and B are defined by $A = A^T + A^C$,

$$A^T = \frac{(N^T)^2\sum_{i=1}^{m^T}(n_i^T)^2+(\sum_{i=1}^{m^T}(n_i^T)^2)^2-2N^T\sum_{i=1}^{m^T}(n_i^T)^3}{(N^T)^2},$$

$$A^C = \frac{(N^C)^2\sum_{i=1}^{m^C}(n_i^C)^2+(\sum_{i=1}^{m^C}(n_i^C)^2)^2-2N^C\sum_{i=1}^{m^C}(n_i^C)^3}{(N^C)^2},$$

$$B = n_U^T(m^T-1)+n_U^C(m^C-1).$$

Confidence intervals are based on the model calculated standard errors and using $d_t \pm 1.96 * SE$ (Kasim et al, 2017).

Calculation of the rate ratio and odds ratios are directly calculated by exponentiating the parameter estimates from the count and logistic models, respectively. Confidence intervals are directly calculated via glmmTMB R package (Brooks *et al.,* 2017) using the 'confint.glmmTMB' function. This uses the Wald method by default, using the standard error and t-distribution with Satterthwaite degrees of freedom.

Estimation of ICC

In the SAP, we stated that ICCs (at both the classroom and school level), would be calculated for the null model (without covariates predicting the DIBELS® composite score); and then for the primary model (i.e. the model including the baseline DIBELS® composite score, with baseline DIBELS® composite score and school size strata as covariates). However, given that the three-level model did not converge, we report the ICC from the two-level model accordingly. The ICC for the two-level model is defined as:

$$ICC = \frac{\sigma^2{}_{sch}}{\sigma^2+\sigma^2{}_{sch}},$$

where σ^2 is the residual variance, and $\sigma^2{}_{sch}$ is the random intercept variance according to schools (school-level clustering).

Implementation and process evaluation

Research methods

Implementation (both delivery and process mechanisms) is the critical aspect of this study. Evidence suggests that it is possible to teach pupils with SEND the early reading skills necessary to become proficient at reading. Whether that can be delivered at scale and via school staff and within the weekly curriculum is the underlying question in the HERiSS logic model.

Using a mixed method approach the implementation and process evaluation (IPE) includes data about the fidelity to the intervention, the experience and perceptions of teachers/TAs, pupils, and ISOs, and an understanding of business as usual (see Table 6). Implications and considerations for a future effectiveness trial are discussed.

Study research questions

5) How well was HER® implemented? Did schools comply with the intervention as defined above?

6) What are the barriers and facilitators for good implementation? (e.g. school/class size, setting of HER® within schools, timetabling of HER®/TA time, IT support and availability, ease of intervention manual use, selection of TAs).

7) Can HER® be delivered to Key Stage 1 and Key Stage 2 pupils in special schools with high fidelity, and what is the usual dosage?

   a) How many HER® episodes do pupils access/complete per week, and throughout the intervention?
   b) How are the adaptations in the implementation manual used for pupils undertaking HER®?
   c) What is the relationship between fidelity to HER® and the impact of HER® on pupils' reading skills?

8) To inform future research in special schools:

   a) What are the most effective approaches for recruiting special schools and pupils to participate in an RCT?
   b) What are the retention rates of schools and pupils? What are the reasons for dropout of schools/pupils?

9) How does HER® differ from reading business as usual?

10) What are pupils' and teachers' experience of, attitudes toward, and perceptions of the HER® programme, as well as its impact?

Data collection methods

Fidelity and dosage

Fidelity and dosage data were gathered directly from HER®; HER® software allows for an analysis at pupil level including episode accuracy scores, frequency of episodes, use of benchmarks, and benchmark scores.

High fidelity is defined as evidence of at least 80% fidelity for the data collected about both:

- repeating episodes when required, and
- completing and acting appropriately with benchmark assessments (with either a continuation, repetition of an episode, or a fluency activity).

Target dosage is defined as completing three sessions of HER® per week per pupil.

Fidelity and dosage data were used to answer study research question 7.

Compliance data

Compliance data (defined above) were collected by the delivery team (data in respect of attending training) and by ISOs as part of their scheduled visits. All ISO data collection was cross-checked by the delivery team. Compliance data were used to answer study research question 5.

Interviews with pupils

Interviews with pupils who received HER® were conducted to establish what they liked and disliked about HER® as part of answering study research question 10. A sampling framework had been proposed to sample up to 15 pupils from ten schools in the intervention group with a range of ages, primary needs, and initial reading scores but the experience of engaging with schools to complete the post-intervention data collection was such that all intervention schools were invited to offer pupils the opportunity to discuss their experiences. Informed consent for pupils to take part in the interviews was provided by pupils' primary carers, with information and consent forms being sent out and collected by the pupils' schools. Nine pupils took part in interviews to share their views on Headsprout®. Pupils were recruited from six different schools. All pupils interviewed were male, and pupils were in Year 3 to Year 6 (i.e. ages 7–11). Pupils had a range of conditions and needs including: autism; speech, language, and communication need; severe learning disability; mental health need; and a physical disability. Schools scheduled times for the pupil interviews with the study team. Interviews were conducted online via Zoom with individual pupils, plus a member of school staff to support the pupil and to set up the call.

Interviews were conducted using Talking Mats, a symbol-based communication tool used to help people to reflect and express their views on a given topic. It has been used successfully with children with a range of communication support needs, including disabled children (Bradshaw *et al.,* 2018; Germain, 2004). A set of topics and associated symbols for the HERiSS Talking Mat was agreed by Dr Flynn and Dr Hayden, a member of the CEDAR team trained in the use of Talking Mats, with consultant support from Dr Bradshaw, a Talking Mats trainer and expert. The interview asked pupils about whether they liked/disliked elements of HER® and how they feel about reading including both confidence and enjoyment. Interviews were conducted by Dr Hayden.

Interviews with teachers/TAs

Semi-structured interviews with teachers/TAs from schools in the intervention group were conducted to understand the mechanisms that contribute to/explain the outcomes of the efficacy trial. Interviews were conducted using the online platform Microsoft Teams. The interview schedule was developed by the evaluation team in discussion with the delivery team and the EEF and included topics such as attitudes/perceptions of delivering HER®, barriers and facilitators to delivering HER®, any unexpected consequences for pupils taking part in HERiSS, perceptions of how pupils experienced HER®, perceptions of the training and subsequent support provided (by ISOs and by school leadership), the usefulness of the HER® manual, and confidence in teaching children with SEND to read and more generally. In addition, the interviews explored Covid-19-related challenges with an attempt to fully describe how these challenges differ from pre-pandemic times. Teachers were also asked about whether the government phonics policy changes, in effect from the start of the 2021/2022 school year, have had any impact on how phonics were delivered in schools as well as their involvement in the trial, and how this would have been different before this policy was introduced.

All intervention schools were invited to take part in the interviews. Information sheets and consent forms were sent to and collected by schools. Informed consent to take part in the interviews was provided by teachers/TAs. The invitations to participate asked schools to prioritise where possible the most senior, or experienced teacher involved in making decisions, and those involved in the direct implementation of HERiSS. Decision makers were, in addition, asked about the decision-making process to take part in HERiSS including what motivated them to take part, challenges that arose, and reflections of taking part in an RCT.

Thirteen participants, five teachers and eight TAs, from 12 schools took part in a semi-structured interview asking about their experiences of delivering HER®. The length of time in their current school averaged seven years and ranged from 14 to just over one year. All but one of the participants had a career background in SEND and all participants had worked for at least one year in the setting that they were currently in. All participants were involved in supporting literacy, seven across the whole school and six focused on the class they support. One participant was the literacy lead for the school. The number of pupils supported through HERiSS ranged from one to ten.

Interviews were conducted by Dr Denne and used to directly answer study research questions 6 and 10 but also to contribute to an understanding of study research questions 5, 7, and 9.

Interviews with ISOs

ISO input is a key component of implementation compliance. All five ISOs were sent information sheets and consent forms and were invited to take part in a semi-structed interview conducted using the online platform Microsoft Teams, asking about their experiences of supporting schools to deliver HERiSS. All five took part. The total number of schools allocated for support was 26. One participant had an overarching role supporting the other ISOs as well as being responsible for two schools. The other ISOs supported an average of six schools each ranging from nine to two. Three reported that not all schools allocated to them engaged in the intervention. The total number of schools that engaged was 22.

Four of the five ISOs were involved in HERiSS at the start of the academic year 2021/2022. One ISO joined the team in January 2022 to focus on two schools that had not at that stage engaged with the delivery team.

The interview schedule was developed by the evaluation team in discussion with the delivery team and the EEF and was designed to gather in-depth data about topics including ISO perceptions of the facilitators and barriers to implementation in special school settings, attitudes/perceptions of schools' engagement with additional support provided (the HER® manual and ISO recommendations), and support provided by school leadership. The interviews were conducted by Dr Denne and used to partly answer study research questions 5, 6, and 7.

Pre- and post-intervention teaching-as-usual survey

A pre- and post-intervention teaching-as-usual survey was sent to teachers in all schools (intervention and control) to establish what business as usual is for literacy before and during the intervention period (including standard literacy activities, school IT facilities and expertise, usual use of IT for teaching pupils). The survey was distributed via Qualtrics and was put together by both the evaluation and delivery teams in discussion with the EEF. These data have been used to partly answer study research question 9.

School and pupil recruitment and retention data

The delivery team was responsible for recruiting schools to the trial and kept a database (recruitment and retention log) of how many schools were approached, how many were recruited, school and pupil retention figures at all stages, and reasons for attrition (if given). The evaluation team worked closely with the delivery team to maintain this log as information regarding school and pupil retention arose during baseline and post-intervention IE data collection. Recruitment information was also shared with the EEF throughout the recruitment period.

The recruitment and retention log has been used to partly answer study research question 8.

Table 6. IPE methods overview

Research methods	Data collection methods	Participants / data sources	Data analysis methods	Research questions addressed	Implementation / logic model relevance
Quantitative (Data gathered from software)	Direct from HER®	Data at pupil level (up to 480)	Descriptive	7	Dosage and fidelity critical to success of underpinning theoretical mechanisms
Quantitative (Direct observation)	Direct observation (ISO visits in schools where possible)	HER® sessions (all pupils)	Descriptive	5, 6, and 7	Dosage and fidelity critical to success of underpinning theoretical mechanisms
Qualitative (survey)	Survey	All schools (to be completed by one teacher at each school)	Descriptive	9	Key to understanding if the intervention has resulted in the change processes identified

Qualitative (Interview)	Interview (Talking Mats)	Up to 15 pupils across 10 schools	Mixed method	10	Pupils' perceptions are a key component of Change process 3 and associated outputs
Qualitative (Interview)	Interview (semi-structured)	12–15 teachers across 10 schools	Structured qualitative approach	6 and 10	Teachers' perceptions are a key component of Change processes 1 and 2
Qualitative (Interview)	Interview (semi-structured)	Four ISOs	Structured qualitative approach	5, 6, and 7	ISO perceptions and experiences are a key component of Change processes 1 and 2
Quantitative (Direct observation)	Direct observation (where possible)	Data collected by delivery team responsible for recruitment and retention	Descriptive	5, 7, and 8	Useful information for future studies on a larger scale

Analysis

The IPE data collection employed a mixed method approach as outlined above. For the analysis a triangulation protocol as outlined by Tonkin-Crine *et al.* (2015) was used to integrate the respective data sources. This method of analysis was chosen because of the interrelated nature of the data sources—with many study research questions best answered using more than one data source, and because the triangulation process facilitates the validity of findings by assessing the extent to which different data sources agree with or contradict findings.

Individual data source analysis

Individual data sources were first analysed as follows:

1) Fidelity and dosage data gathered directly from HER® were analysed descriptively.

2) Compliance data gathered by the delivery team were analysed descriptively.

3) Pupil interview data were analysed by Dr Hayden. The Talking Mats interview audios were transcribed, videos were reviewed, and observation descriptions were added to the transcripts. The Talking Mats Effectiveness Framework of Functional Communication (EFFC) was administered to each pupils' interview. The EFFC is used to determine whether each Talking Mats is an example of 'effective communication' based on several observable indicators (developed by Talking Mats Limited, 2014; Murphy *et al.,* 2010). These observable indicators include seven items including: 'Engagement'; 'Participant's understanding of issue for discussion'; 'Interviewer's understanding of participant's views'; 'Participant – on track'; 'Symmetry'; 'Real time'; and 'Interviewer's satisfaction'. These items were measured on a 5-point scale from 4 to 0: 'Always' (4); 'Often' (3), '50/50' (2), 'Occasionally' (1), and 'Never/none' (0). These items were then summed to produce a score out of 28, with a score of at least 75%, or 21, indicating effective communication. Pupils whose Talking Mats scored under 21 on the EFFC were reviewed by a second researcher, and if agreed between both reviewers, removed from the analysis.

   Pupils' Talking Mats data were first analysed quantitatively, with the most popular symbols identified by calculating mean scores for each symbol. Mean symbol scores were calculated by allocating the following scores to each top-scale item: 'Like' = 2; 'Not sure' = 1; 'Don't like' = 0. The responses from participants for each option symbol were summed and then divided by the number of participants, with possible scores ranging from 0–2. Higher scores indicated more positive feelings toward that specific option symbol. Top-scale responses were also summed to quantitatively summarise how pupils felt about Headsprout® overall. The transcriptions with observation descriptions were then analysed using content analysis (Hsieh and Shannon, 2005). As the amount of verbal data provided by pupils was minimal, this analysis remained descriptive and so was presented symbol-by-symbol.

4) and 5) Teacher/TA interview data and ISO interview data were analysed by Dr Denne. Both sets of interview audios were transcribed verbatim. A structured qualitative approach was used to analyse these data. Analysis was predominantly deductive with transcripts examined for pre-specified themes based on the study research

questions, but with the flexibility of adding additional themes that were apparent across the data set and added depth to the understanding of schools' experience of HERiSS. A master theme table was produced. A second reviewer independently analysed 20% of the transcripts (three teacher and TA interviews and one ISO interview) and the findings were discussed collaboratively. There was agreement across all the pre-specified themes. All additional themes that were identified by the second reviewer had been captured in the master theme table developed by Dr Denne. Once the independent review was complete the master theme table was synthesised, and a thematic map was produced for both teacher/TA interview data and ISO interview data. The synthesis and thematic maps were once again reviewed and validated by the second reviewer.

6) Data from the pre- and post-intervention teaching-as-usual survey were downloaded from Qualtrics into IBM® SPSS® Statistics Version 28.0.1.1 (15) (IBM®, Armonk, NY, USA). Programmes that schools were using were coded as follows to reflect the key characteristics of HER® so that differences between HER® (intervention arm) and business as usual (control arm) could be analysed: adapted for SEND; phonics; sequential learning; one to one instruction; increased learning opportunities; and repetition. Data were analysed descriptively.

7) Data from the recruitment and retention log were analysed descriptively.

Triangulation

A triangulation matrix was developed by Dr Denne tabulating key findings across all data sources. All instances of convergence (complete agreement), complementarity (where data sources complemented each other), and dissonance (contradiction) were highlighted. Key findings were mapped onto the logic model where appropriate. The matrix was reviewed and agreed by the Evaluation Team and used to interpret the overall findings of the IPE.

Costs

A full economic cost-effectiveness analysis of the intervention has not been included in this study. However, the following cost data were collected to enable schools to decide whether to invest in the programme.

- Personnel for the implementation of the programme: These data were collected by the delivery team, primarily through ISO visits—the number of staff members involved at each school, the numbers of HER® sessions delivered per school and the staff time involved, and the time the number of person days per school delivered by ISOs.
- Personnel during training for the implementation of the programme: These data were collected by the delivery team at the point of delivery and include the length of time spent per school in training, the number of school staff involved in training, the administrative time spent to set up each training session, and the time involved per school on the part of each trainer.
- Programme costs: These data were collected by the delivery team and include the costs of travel for each ISO visit per school and the costs of telephone support for each school over the duration of the intervention.
- Facilities, equipment, and materials: These data were collected by the delivery team and include the cost of the HER® licence per school, the cost of the provision of the HER® support manual per school, and the costs to schools of reproducing support materials when needed (collected during ISO visits).
- Other programme inputs: The ISOs kept note of any other costs arising as a result of the delivery of the intervention.

Timeline

Table 7: Timeline

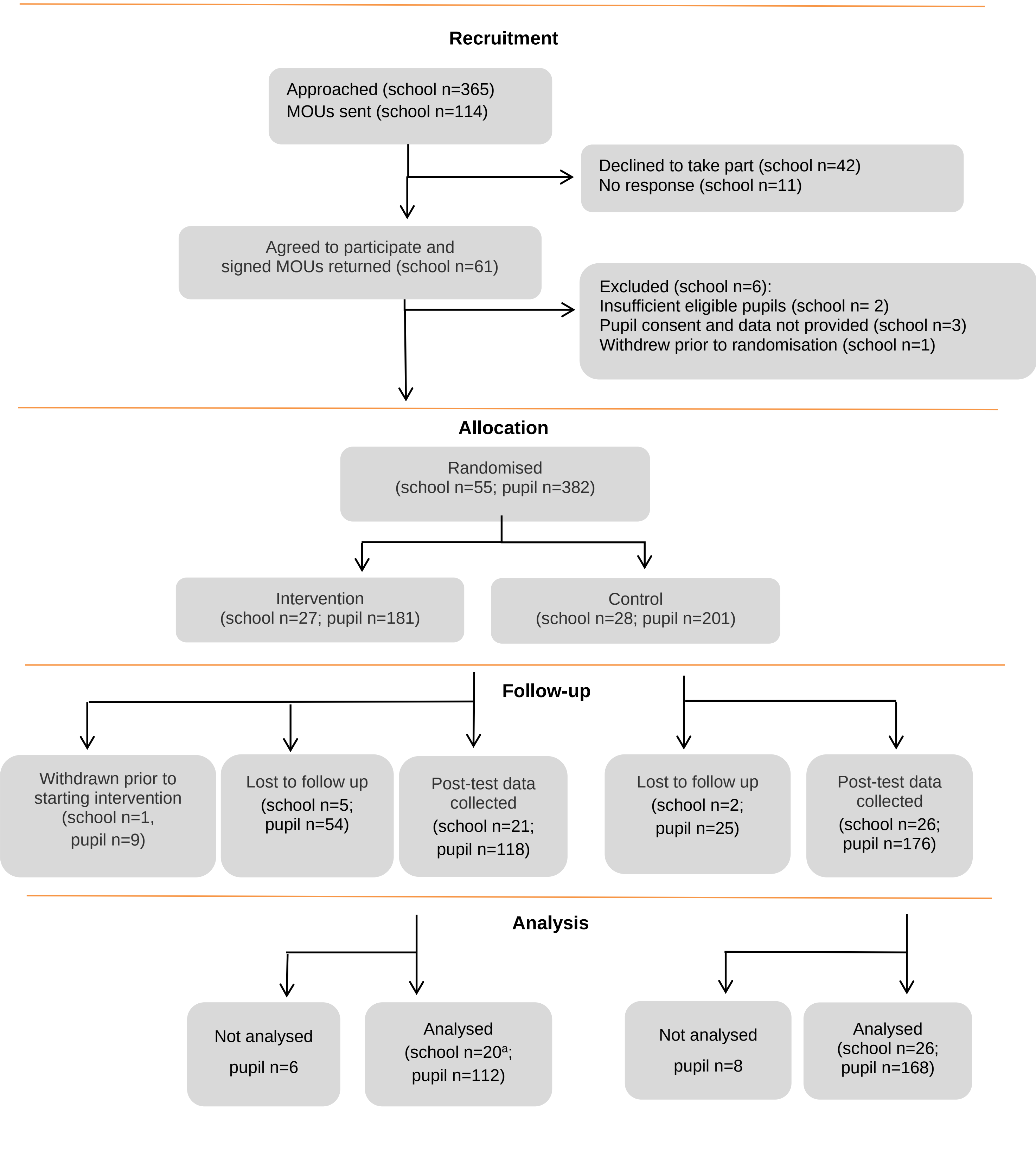
Dates	Activity	Staff responsible / leading
September 2019	Updated proposal and budget	Warwick
August 2020 – March 2021	Develop study materials	Warwick (Bangor)
December 2019 – May 2021[a]	Recruitment	Bangor
April 2021 – July 2021	Parental consent obtained	Bangor
April 2021 – July 2021	Pupil personal data collected	Bangor
April 2021 – July 2021	School teaching-as-usual data collected	Warwick
April 2021 – July 2021	Baseline pupil data collection	Warwick
April 2021 – July 2021	Randomisation	Warwick
September 2021	Training resources released and live session	Bangor
September 2021 – mid-May 2022	Intervention delivery	Bangor
End of May 2022 – mid-July 2022	Post-intervention pupil data collection	Warwick
End of May 2022 – mid-July 2022	Post-intervention qualitative data collection	Warwick
Mid-July 2022 – Dec 2022	Data analysis and write up	Warwick
November 2022	EEF report first draft due	Warwick
July 2023	Final report due	Warwick

[a] *Recruitment paused during lockdown and study postponed for one full academic year because of the Covid-19 pandemic*

Impact evaluation results

Participant flow including losses and exclusions

Figure 3: Participant flow diagram (two arms)



[a] *Please note that among those not analysed in the intervention arm were all responses from one school, so the number of actually analysed schools in the intervention arm was n=20. MOUs=Memorandum of Understandings.*

A flow of participants through the evaluation is presented in Figure 3.

The sample of interest consists of pupils from special schools in England and Wales at Key Stage 1 and Key Stage 2 for the academic years 2020/2021 and 2021/2022. These pupils were assessed at the end of the school year (Summer Term – July) at both baseline and post-test.

A total of 382 pupils were involved in the trial. Pupils were given a unique pupil identifier as well as a unique school identifier. No pupils withdrew consent from the trial, but one school did pull out prior to randomisation citing a lack of capacity post-pandemic. From these 382 pupils, 102 were lost to withdrawal, follow-up, or could not be analysed and therefore, 280 pupils (47 schools) were analysed.

Table 8: Minimum detectable effect size (MDES) at different stages

		Design		Protocol / randomisation		Analysis (actual values from two-level model)	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.298	0.327	0.413	0.461	0.32	0.402
Pre- / post-test correlations	Level 1 (pupil)	0.5	0.5	0.5	0.5	0.64	0.68
	Level 2[a] (class)	0	0	0	0	-	-
	Level 3[a] (school)	0	0	0	0	0.004	0.012
Intracluster correlations (ICCs)	Level 2 (class)	0.4	0.4	0.4	0.4	-	-
	Level 3 (school)	0.1	0.1	0.1	0.1	0.04	0.04
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two	Two	Two
Average cluster size		9	3.89	7	3.02	6.95	3.87
Number of schools	Intervention	49	49	27	27	20	20
	Control	50	50	28	28	26	26
	Total:	99	99	55	55	46	46
Number of pupils	Intervention	441	191	181	98	112	64
	Control	450	194	201	103	168	88

Total:	891	385	382	201	280	152

[a] *Sufficient information to estimate the pre- and post-test correlations was not available, so this was conservatively set at zero for both levels 2 and 3.*

Table 8 shows the MDES for the overall and FSM samples at the different stages of the trial. At the time of preparing the trial protocol the MDES was 0.298 for the overall sample and 0.327 for the FSM sample. At randomisation this was larger for both the overall sample (0.413) and for the FSM sample (0.461) reflecting a reduced number of schools in the trial than had been assumed when writing the protocol. For the final analysis sample, the MDES was slightly different as the three-level design did not fit sufficiently (models were run with different optimisers but model convergence in the three-level model was still not achieved), so a two-level model was used instead (i.e. no class-level clustering, school level only). Therefore, an actual MDES of 0.32 (overall sample) and 0.402 (FSM sample) were estimated, which were marginally larger than our protocol assumed figures given the change in model and our ICC was much smaller at the school level (ICC=0.04) compared to the conservative assumed values (assumed to be 0.4 school level and 0.1 class level—for both samples but turned out to be 0.04 for school level). We found higher pre- and post-test correlations at level 1 (0.64), which is considerably higher than the conservative estimate made in the sample size calculation.

Attrition

The trial had a relatively high rate of attrition, as presented in Table 9. From the 55 schools that were randomised and allocated to one of two arms, intervention (school n=27, pupil n=181) and control (school n=28, pupil n=201), 63 pupils from six schools in the intervention group and 25 pupils from two schools in the control group were lost to follow-up.

Of the intervention group, one school withdrew post-randomisation (school n=1, pupil n=9) citing an inability to commit because of staffing/resources. This school had not begun intervention. At post-testing, a further five schools and 45 pupils were lost for the following reasons:

- did not respond to the evaluation team (school n=2; pupil n= 15);
- could not fit the sessions in before the end of term (school n=1; pupil n = 11);
- not officially withdrawn but felt it conflicted with current literacy programme (school n=1; pupil n= 13); and
- not officially withdrawn but 'staffing crisis' (school n=1; pupil n= 6).

A further nine pupils were lost across intervention schools:

- in Key Stage 3;
- mute/non-verbal;
- withdrawn by school, no reason given; and
- long-term absence.

Of the control group at post-testing, two schools and 25 pupils were lost for the following reasons:

- did not respond to the evaluation team (school n=1; pupil n= 6); and
- not officially withdrawn but felt post-tests too difficult for pupils (school n=1; pupil n= 8).

A further 11 pupils were lost across control schools:

- in Key Stage 3;
- withdrawn by school, no reason given; and
- left school/long-term absence.

In total, this led to the number of schools lost at post-testing for schools n=8 and pupils n=88, with post-test data collected from 294 pupils and analysed from 280 pupils across 47 schools.

Attrition was higher in the intervention group because of the relative number of schools lost to the study. As noted above one school cited the capacity to commit to the intervention post-randomisation, two schools did not officially withdraw but cited reasons for not participating, and two schools never engaged with either the delivery team or the evaluation team.

Table 9 presents the total rate of attrition for the pre- and post-test primary outcome included in the analysis. The total rate of attrition for primary outcome is 27% for DIBELS® composite. As this rate of attrition is greater than the 5% threshold specified in the SAP, we investigated further whether the missing data is at random (MAR) or whether there is a pattern of missingness that can be predicted by some available variables by using a logistic analysis (see section ‘Missing data analysis’ for further details).

Table 9: Pupil-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	181	201	382
	Analysed	112	168	280
Pupil attrition	Number	69	33	102
(from randomisation to analysis)	Percentage	38	16	27

Pupil and school characteristics

Table 10 reports the main sample characteristics for the schools and pupils as randomised. The most relevant difference at pupil level is the proportions in each trial arm relating to special educational needs and disabilities (SEND) status: Autism Spectrum Disorder (ASD) was 50% in the intervention group and 19% in the control group. Similarly, pupils with MLD, 17% in the intervention group and 4% in the control group. Another difference was seen in pupils with SLCN, 24% in the intervention group and 11% in the control group. Most other categorical variables showed no other relevant differences between intervention and control schools in school-level characteristics.

Continuous variables showed some small differences between intervention and control groups, but this difference is exaggerated given the level of skew in most variables (see Appendix E for distribution plots). We see Hedges’ g effect sizes of between 0.07 and 0.402 among these variables, with the largest differences appearing when there is significant missingness present. All analyses have been controlled for baseline values, so relative differences are accounted for in the final results.

Table 10: Baseline characteristics of groups as randomised

School level (categorical)	National level (DfE 2019) No. of pupils (%)	Intervention group n/N (missing)	%	Control group n/N (missing)	%	
School size (large)	141+ pupils 31.1%	Pupils: 170/181 (0) Schools: 19/21	94	Pupils: 185/201 (0) Schools: 24/26	92	
School size (small)	Up to 75 pupils 26.3%	Pupils: 11/181 (0)	6	Pupils 16/201 (0)	8	

		Schools: 2/21		Schools: 2/26		
Pupil level (categorical)	**National level (DfE 2019) %**	**n/N (missing)**	**%**	**n/N (missing)**	**%**	
FSM	37.9%	98/181 (7)	54	103/201 (9)	51	
EAL	14.6%	23/181 (1)	13	28/201 (14)	14	
SEND (ASD)	30.0%	90/181 (3)	50	38/201 (18)	19	
SEND (MLD)	12.9%	30/181 (3)	17	8/201 (18)	4	
SEND (SLD)	21.6%	45/181 (3)	25	73/201 (18)	36	
SEND (PMLD)	6.9%	0/181 (3)	0	4/201 (18)	2	
SEND (SLCN)	7.3%	43/181 (3)	24	22/201 (18)	11	
SEND (SpLD)	1.6%	5/181 (3)	3	8/201 (18)	4	
SEND (PD)	3.3%	11/181 (3)	6	11/201 (18)	6	
SEND (SEMH)	12.7%	4/181 (3)	2	5/201 (18)	3	
SEND (Other)	6.2%	9/181 (3)	5	11/201 (18)	6	
Pupil level (continuous)		**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	**Effect size (Hedges' g)**
BASELINE DIBELS® Composite	N/A	175/181 (6)	394 (15.0)	195/201 (6)	390 (9.80)	0.319

BASELINE LNF	N/A	175/181 (6)	7.07 (7.92)	195/201 (6)	5.43 (6.97)	0.221
BASELINE PSF	N/A	171/181 (10)	3.93 (7.77)	187/201 (14)	4.46 (7.37)	-0.07
BASELINE NWF-CLS	N/A	168/181 (13)	13.7 (10.0)	184/201 (17)	13.5 (11.1)	0.019
BASELINE NWF-WRC	N/A	168/181 (13)	2.93 (4.77)	184/201 (17)	1.92 (3.53)	0.242
BASELINE WRF	N/A	166/181 (15)	6.23 (7.32)	187/201 (14)	4.16 (4.90)	0.336
BASELINE ORF – correct	N/A	106/181 (75)	19.7 (28.4)	106/201 (95)	10.6 (14.7)	0.402
BASELINE ORF – errors	N/A	106/181 (75)	6.25 (3.64)	106/201 (95)	7.12 (4.45)	-0.214
BASELINE ORF-Accuracy)	N/A	105/181 (76)	49.1 (35.1)	102/201 (99)	42.7 (31.9)	0.191
Reading self-concept (baseline)	N/A	162/181 (19)	9.02 (1.63)	179/201 (22)	8.82 (2.04)	0.108

ASD=Autism Spectrum Disorder; DfE=Department for Education; DIBELS®=Dynamic Indicators of Basic Early Literacy Skills®, Eighth Edition; EAL=English as an Additional Language; FSM=free school meals; LNF=letter naming fluency; MLD=Moderate Learning Difficulty; N/A=not applicable; NWF-CLS=nonsense word fluency – correct letter sounds; NWF-WRC=nonsense word fluency – words read correct; ORF-ACC=oral reading fluency – accuracy; ORF – correct=oral reading fluency – correct words; ORF – errors=oral reading fluency – errors made; PD=Physical Disability; PMLD=Profound and Multiple Learning Disability; PSF=phonemic segmentation fluency; SD=standard deviation; SEMH=Social, Emotional, and Mental Health; SEND=special educational needs and disability; SLCN=Speech, Language, and Communication Needs; SLD=Severe Learning Difficulty; SpLD=Specific Learning Difficulty; WRF=word reading fluency.

Outcomes and analysis

Primary analysis

Afte running both two- and three-level models, we found that the three-level model did not converge and produced a singularity error message. This was a clear indication of model overfitting due to the variance for class/school year random slope being zero or near zero (class-level random effect). Therefore, following our SAP, we concluded that the two-level structure (school-level clustering only) is the most appropriate for this analysis and all subsequent analyses have been reported on the basis of a two-level model. As per the statistical analysis plan (SAP), all secondary outcomes, subgroup, and any additional analyses default to the appropriate two-level model accordingly.

Figure 4 shows the histogram of the DIBELS® composite at 12 months post-test. The distribution appears skewed and so we also report the model estimates using a robust LMM. The skew is potentially likely to provide unbiased point estimates but with overly optimistic (too narrow) confidence intervals and smaller *p*-values.

Figure 4: Histogram of the DIBELS® composite at 12 months (post-test), and is split by trial arm (control and intervention)

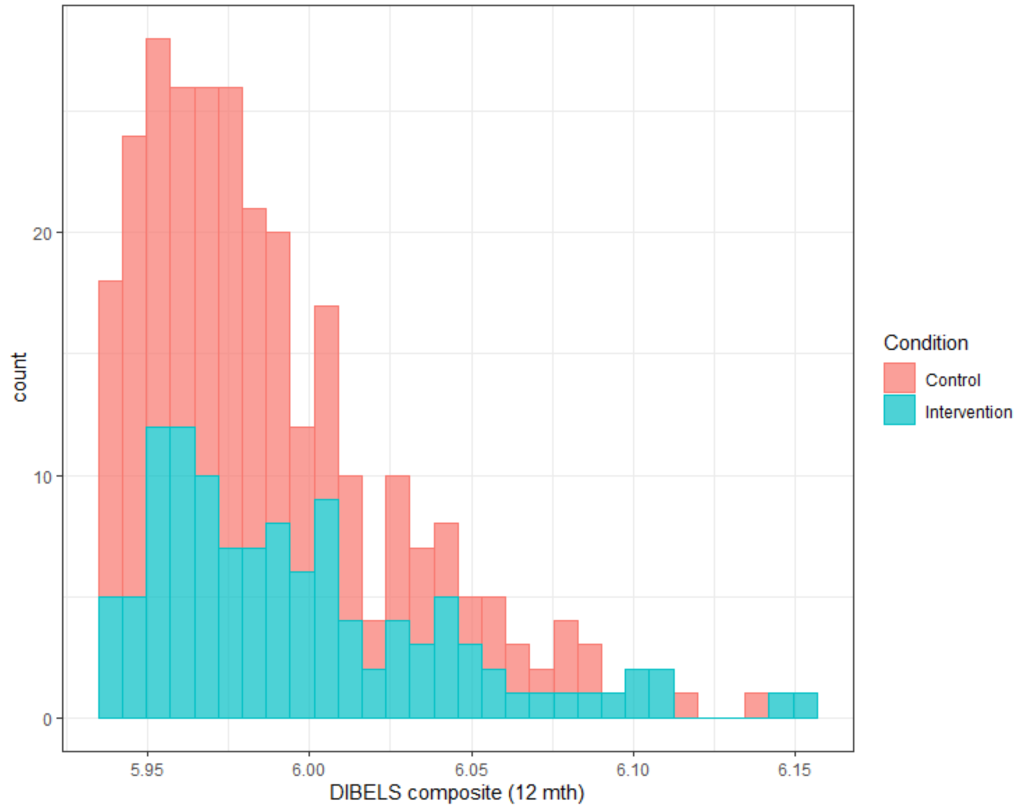


Table 11 reports the model parameter estimates from the primary outcome analysis using an linear mixed model (with and without robust estimation). Given the analysis of covariance (ANCOVA) design of the model, the intercept represents the expected value (or mean) of DIBELS® at 12 months when DIBELS® at baseline and school size covariates are both equal to zero. We were unable to find evidence against the hypothesis that the intervention and control (business as usual) were not significantly different from each other (β=-0.09 95% CI [-2.93, 2.75], t (44.69231)=-0.062, *p*=0.951). The mean difference between the control and intervention groups, controlling for baseline DIBELS® and school size, is -0.09 (Hedges' g=-0.01, 95% CI [-0.28, 0.26]). Although the intervention effect size is negative, it is clear that there is considerable uncertainty in the estimate reflected in the width of CI, which also contain zero.

We also fitted the linear mixed model using the robust scoring equations estimator, this improves the estimation of the standard errors and associated *p*-values. More specifically, the robust estimation is insensitive to small departures from idealised assumptions, required using standard estimation methods, which in this case incorporate the skew in the distribution. We find that the direction of the intervention effect is reversed and more in line with expectations for intervention (β=0.46, t (44.69231) =0.445, *p*=0.659; Hedges' g=0.06, 95% CI [-0.19, 0.31]) and, in theory, should provide more accurate estimate of the intervention effect size's SEs. However, we still fail to find any evidence that there is a difference between intervention and control (business as usual).

The unconditional ICC (95% CI) associated with school for the primary outcome, DIBELS® composite measures is 0.232 (0.094, 0.352), and the conditional ICC (95% CI) is 0.041 (0.005, 0.130). Robust estimates of the random intercept variance in the robust model were approximately zero (to three decimal points in R output), therefore, corresponding ICC were similarly near zero.

Table 11: Primary analysis for DIBELS® composite at 12 months using a two-level linear mixed model

	Primary analysis (two level)		Robust primary analysis	
Coefficient	*Estimates*	*CI*	*Estimates*	*CI*
Intercept	-15.70	-55.63, 24.24	-30.96*	-60.93, -0.99
DIBELS® baseline	1.05***	0.95, 1.15	1.09***	1.01, 1.17
School size	2.25	-2.66, 7.17	1.66	-1.78, 5.11
Intervention	-0.09	-3.00, 2.82	0.46	-1.58, 2.51
Random effects				

σ^2	110.85	66.72
τ_{00}	4.68 $_{School\_ID}$	0.00 $_{School\_ID}$
ICC	0.04	0.000
N	46 $_{School\_ID}$	46 $_{School\_ID}$
Observations	280	280
Marginal R^2 / Conditional R^2	0.624 / 0.639	0.756 / 0.756

* p<0.05; ** p<0.01; *** p<0.001.

Table 12: Primary analysis summary

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (SD)	Total n (intervention;control)	Hedges'g (95% CI)	*P*-value
12-months DIBELS® composite score	112 (38%)	398.49 (395.46, 01.52)	168 (16%)	398.58 (395.73, 01.44)	280 (112; 168)	-0.01 (-0.28, 0.26)	0.951

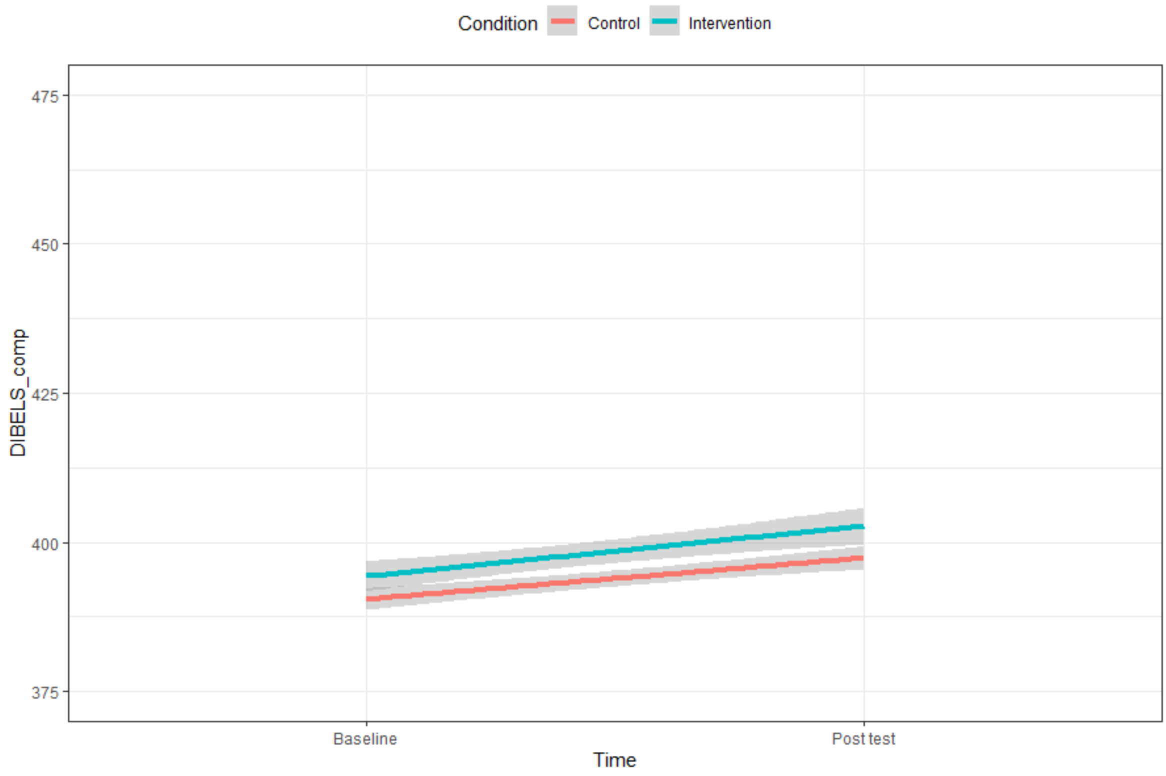
It may be more intuitive to look at this result using Figure 5, which shows the change over time on average is the same in both trial arms (bold lines) and virtually no difference on average between the trial arms (i.e. lines are parallel). An additional exploratory analysis using a time predictor and its interaction with trial arm in the linear mixed model rather than adjusting for baseline (Table 13), highlights that there is a change over time but no significant difference between the trial arms and intuitively is perhaps easier to understand (although this does address a slightly different research question to our main primary analysis). This analysis is performed using the classical approach rather than using a robust estimator.

Table 13: Additional analysis for DIBELS® composite at 12 months using a two-level linear mixed model with time and time x intervention interaction

	Supplementary analysis (two level)	
Coefficient	***Estimates***	***SE***
Intercept	389.75***	4.52
School size	0.91	4.53
Time	6.75***	0.81
Intervention	3.11	2.43
Time x intervention	0.31	1.28
Random effects		
σ^2	56.81	
τ_{00} $_{Pupil\_ID}$	103.26	
τ_{00} $_{School\_ID}$	55.41	
ICC	0.74	
N $_{School\_ID}$	55	
N $_{Pupil\_ID}$	378	
Observations	658	
Marginal R^2 / Conditional R^2	0.059 / 0.752	

* p<0.05; ** p<0.01; *** p<0.001.

Figure 5: Average pre- and post-test outcomes for each trial arm overlaid (bold lines). Standard errors are displayed around each line as the grey ribbons



Secondary analysis

Secondary outcome analyses have been organised into separate subsections for clarity. Each secondary outcome analysis reports the SAP specified analysis, the FSM analyses, and a modified analysis to account for zero inflation and over dispersion in the DIBELS® component scores (see individual 'Results' sections and Figure 6 below). After running the initial SAP specified models and checking for both zero inflation and over dispersion, it was clear that both were present in all DIBELS® component measures.

Figure 6: Histograms of the secondary outcome measures (split by trial arm)

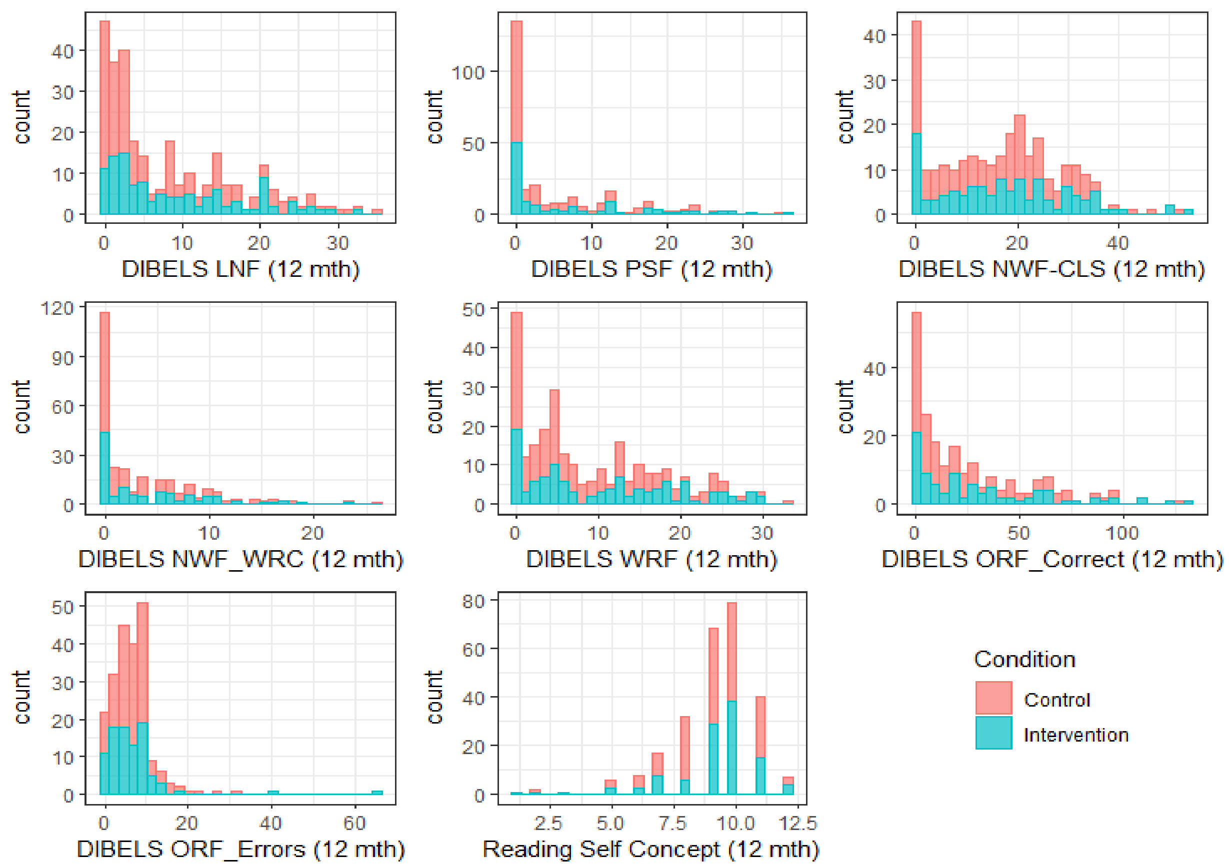


Table 14: Secondary analysis summary

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean[a] (SD)	n (missing)	Mean[a] (SD)	Total n (intervention; control)	Incidence rate ratio (CI)	*P*-value
Letter naming fluency (LNF)	115 (66)	1.95 (1.66, 2.25)	173 (28)	1.81 (1.52, 2.09)	288 (115; 173)	1.16 (0.86, 1.55)	0.325
Phonemic segmentation fluency (PSF)	112 (69)	0.528 (-0.30, 1.36)	166 (35)	0.320 (-0.50, 1.13)	278 (112; 166)	1.23 (0.57, 2.66)	0.597
Nonsense word fluency – correct letter sounds (NWF-CLS)	113 (68)	2.69 (2.44,2.94)	163 (38)	2.74 (2.50, 2.98)	276 (113; 163)	0.96 (0.75,1.2)	0.715
Nonsense word fluency – words read	113 (68)	1.132 (0.732, 1.53)	163 (38)	0.997 (0.612, 1.38)	276 (113; 163)	1.14 (0.78, 1.69)	0.495

correct (NWF-WRC)							
Oral reading fluency (ORF) –correct words	90 (91)	2.31 (1.72, 2.89)	126 (75)	2.79 (2.27, 3.31)	216 (90; 126)	0.62 (0.37, 1.04)	0.068
ORF – errors made	90 (91)	1.76 (1.51, 2.01)	126 (75)	2.02 1.79, 2.25)	216 (90; 126)	0.77 (0.60, 0.98)	0.037
Word reading fluency (WRF)	111 (70)	2.05 (1.82, 2.28)	159 (42)	2.09 (1.87, 2.31)	270 (111; 159)	0.96 (0.77, 1.20)	0.709
Outcome	**n** **(missing)**	**Mean** **(SD)**	**n** **(missing)**	**Mean** **(SD)**	**Total n** **(intervention; control)**	**Hedges' g** **(CI)**	***P*-value**
Reading Self-Concept Scale (RSCS)	109 (72)	8.55 (8.02, 9.08)	152 (49)	8.62 (8.09, 9.15)	268 (109; 152)	-0.04 (-0.35, 0.26)	0.785

[a] *Results are given on the log (not the response) scale*

Letter naming fluency (LNF)

The LNF was considerably skewed and contained an inflated number of zero scores, so the specified GLMM (Poisson) model was checked for overdispersion and zero inflation and was present in both cases (overdispersion: dispersion ratio=4.913, Pearson's χ^2=1351.160, p< 0.001; zero inflation: Observed zeros: 46, Predicted zeros: 6, Ratio: 0.13.) Given that we have a significant test of overdispersion and that the ratio of observed to predicted zeros is well outside the tolerance of 5% or a ratio of 0.95, we conclude that the number of zeros is inflated and there is some overdispersion, so need to be adjusted in the main analysis using a zero-inflated negative binomial model. This model provides a more accurate estimate of all parameter estimates and we can see the minor adjustment in the data in Table 15. All model estimates are presented as incident rate ratios (IRRs), which are the appropriate standard effect size measure for count models (as per SAP specification). An IRR of 1.0 indicates equal rates in both groups, an IRR greater than 1.0 indicates an increased rate in the intervention arm.

In all models reported in Table 15, we do not find evidence of a significant difference between intervention and control (business as usual) (IRR=1.16, 95% CI [0.86,1.55], p=0.325), regardless of whether we are adjusting for the moderating effect of FSM eligibility. As expected, we find that baseline LNF is a significant predictor (IRR=1.06 95% CI [1.06,1.07], p<0.0001).

The unconditional ICC (95% CI) associated with school for the secondary outcome, LNF (Poisson) is 0.739 (0.619, 0.796), and the conditional ICC (95% CI) is 0.623 (0.467, 0.698). For the zero-inflated negative binomial model, the unconditional ICC (95% CI) associated with school for the secondary outcome, LNF is 0.103 (R is unable to produce CIs for this model ICC—code does not exist currently), and the conditional ICC (95% CI) is 0.068. For the FSM individual only and FSM interaction models, the conditional ICCs are 0.09 and 0.07, respectively. Once more, all models using zero-inflated negative binomial do not report associated ICC CIs as the statistical software is not yet able to produce these intervals to the authors knowledge.

Table 15: Secondary analysis for DIBELS® letter naming fluency (LNF) at 12 months using a two-level LMM

	Primary LNF analysis		Overdispersed / zero inflated		FSM individuals only		FSM interaction	
Coefficient	*IRRs*	*CI*	*IRRs*	*CI*	*IRRs*	*CI*	*IRRs*	*CI*
(Intercept)	4.29***	2.59, 7.10	4.32***	2.44, 7.63	4.40***	1.93, 10.01	3.59*	1.64, 7.82
LNF baseline	1.06***	1.06, 1.07	1.08***	1.06, 1.10	1.09***	1.06, 1.12	1.08***	1.06, 1.10
School size	0.95	0.57, 1.56	0.93	0.55, 1.56	0.79	0.35, 1.78	0.88	0.51, 1.52
Intervention	1.16	0.86, 1.55	1.15	0.85, 1.57	1.32	0.89, 1.96	2.03	0.88, 4.65
FSM							1.17	0.83, 1.63
Intervention x FSM							0.67	0.39, 0.143
Zero-inflated model								
(Intercept)			0.01	0.00, 2205.56	0.00	-	0.01	0.00, 1419.04
Random effects								
σ^2	0.13		0.94		0.98		0.96	
τ_{00}	0.22 $_{School\_ID}$		0.07 $_{School\_ID}$		0.09 $_{School\_ID}$		0.07 $_{School\_ID}$	
ICC	0.62		0.07		0.09		0.07	
N	46 $_{School\_ID}$		46 $_{School\_ID}$		45 $_{School\_ID}$		45 $_{School\_ID}$	
Observations	280		280		152		273	
Marginal R^2 / conditional R^2	0.390 / 0.770		0.237 / 0.288		0.270 / 0.332		0.243 / 0.296	

** p<0.05   ** p<0.01   *** p<0.001.*

Phonemic segmentation fluency (PSF)

The PSF was considerably skewed and contained an inflated number of zero scores, so the specified GLMM (Poisson) model was checked for overdispersion and zero inflation, and was present in both cases (overdispersion: dispersion ratio=5.015, Pearson's χ^2=1298.979, p<0.001; zero inflation: Observed zeros: 124, Predicted zeros: 48, Ratio: 0.39) Given that we have a significant test of overdispersion and that the ratio of observed to predicted zeros is well outside the tolerance of 5% or a ratio of 0.95, we conclude that the number of zeros is inflated and there is some overdispersion, so need to be adjusted in the main analysis using a zero-inflated negative binomial model. This model provides a more accurate estimate of all parameter estimates and we can see the minor adjustment in the data in Table 16.

In models reported in Table 16, we do not find evidence of a significant difference between intervention and control (business as usual) (IRR=1.23 95% CI [0.57, 2.66], p=0.597), according to the primary analysis. However, when only considering those eligible for FSM, we do find some evidence of a significant effect of intervention over control (business as usual) (IRR=1.79, 95% CI [1.05, 3.05], p=0.032), indicating those eligible for FSM were 1.79 times more likely to score higher on the PSF than control. When considering this as a moderating effect of intervention in the primary analysis, FSM, trial arm (intervention), and their interaction were all significant predictors (FSM: IRR=1.60, 95% CI [1.05, 2.42], p=0.029; Trial arm: IRR=3.47, 95% CI [1.28, 9.43], p=0.0146; FSM x trial arm: IRR=0.52, 95% CI [0.27, 0.98], p=0.0433).

The unconditional ICC (95% CI) associated with school for the secondary outcome, Phonemic Segmentation Fluency (Poisson) is 0.868 (0.804, 0.901), and the conditional ICC (95% CI) is 0.832 (0.723, 0.881). For the zero-inflated negative binomial model, the unconditional ICC (95% CI) associated with school for the secondary outcome, Phonemic Segmentation Fluency is 0.204 (R is unable to produce CIs for this model ICC—code does not exist currently), and the conditional ICC (95% CI) is 0.151. For the FSM individual only and FSM interaction models, the conditional ICCs are 0.13 and 0.17, respectively. Once more, all models using zero-inflated negative binomial do not report associated ICC CIs as the statistical software is not yet able to produce these intervals to the authors knowledge.

Table 16: Secondary analysis for DIBELS® Phonemic segmentation fluency (PSF) at 12 months using a two-level LMM

	Primary PSF analysis		Overdispersed / zero inflated		FSM individuals only		FSM interaction	
Coefficient	*IRRs*	*CI*	*IRRs*	*CI*	*IRRs*	*CI*	*IRRs*	*CI*
(Intercept)	0.59	0.14, 2.52	3.10	0.87, 10.98	4.31*	1.25, 14.92	1.93	0.52, 7.15
PSF baseline	1.04***	1.04, 1.05	1.04***	1.02, 1.06	1.04**	1.02, 1.06	1.04 ***	1.02, 1.06
Intervention	1.23	0.57, 2.66	1.35	0.87, 2.07	1.79*	1.05, 3.05	3.47 *	1.28, 9.43
School size	3.66	0.87, 15.49	1.71	0.55, 5.32	1.02	0.30, 3.42	1.33	0.45, 3.94
FSM							1.60 *	1.05, 2.42
Intervention x FSM							0.52 *	0.27, 0.98
Zero-inflated model								
(Intercept)			0.69*	0.49, 0.98	0.60*	0.38, 0.97	0.70 *	0.50, 0.98
Random effects								
σ^2	0.32		0.86		0.93		0.86	
τ_{00}	1.59 $_{School\_ID}$		0.15 $_{School\_ID}$		0.14 $_{School\_ID}$		0.17 $_{School\_ID}$	
ICC	0.83		0.15		0.13		0.17	
N	46 $_{School\_ID}$		46 $_{School\_ID}$		45 $_{School\_ID}$		45 $_{School\_ID}$	
Observations	264		264		144		257	
Marginal R^2 / Conditional R^2	0.125 / 0.853		0.140/0.270		0.148/0.257		0.153/0.294	

** p<0.05 ** p<0.01 *** p<0.001.*

Nonsense word fluency - correct letter sounds (NWF - CLS)

The NWF for correct letter sounds was less skewed but did contain an inflated number of zero scores, so the specified GLMM (Poisson) model was checked for overdispersion and zero inflation and was again present in both cases (overdispersion: dispersion ratio=6.152, Pearson's χ^2=1568.643, p<0.001; zero inflation: Observed zeros: 38, Predicted zeros: 0, Ratio: 0). Given that we have a significant test of overdispersion and that the ratio of observed to predicted zeros is well outside the tolerance of 5% or a ratio of 0.95, we conclude that the number of zeros is inflated and there is some overdispersion, so need to be adjusted in the main analysis using a zero-inflated negative binomial model. This model provides a more accurate estimate of all parameter estimates and we can see the minor adjustment in the data in Table 17.

In all models reported in Table 17, we do not find evidence of a significant difference between intervention and control (business as usual) (IRR=0.96, 95% CI [0.75, 1.22], p=0.715), regardless of whether we are adjusting for the moderating effect of FSM eligibility. When adjusting the model to account for overdispersion and zero inflation, the direction of effect changes but this may be due to the small and statistically non-significant effect with the CI containing 1 (IRR=1.10 95% CI [0.94, 1.29], p=0.234). As expected, we find that baseline NWF-CLS is a significant predictor of NWF-CLS at follow-up (IRR=1.02, 95% CI [1.02, 1.03], p<0.0001).

The unconditional ICC (95% CI) associated with school for the secondary outcome, NWF-CLS (Poisson) is 0.770 (0.682, 0.824), and the conditional ICC (95% CI) is 0.750 (0.610, 0.797). For the zero-inflated negative binomial model, the unconditional ICC (95% CI) associated with school for the secondary outcome, NWF-CLS is 0.085 (R is unable to produce CIs for this model ICC—code does not exist currently), and the conditional ICC (95% CI) is 0.019. For the FSM individual only and FSM interaction models, the conditional ICCs are approximately zero (random intercept variance is near zero) and 0.005, respectively. Once more, all models using zero-inflated negative binomial do not report associated ICC CIs as the statistical software is not yet able to produce these intervals to the authors knowledge.

Table 17: Secondary analysis for DIBELS® nonsense word fluency – correct letter sounds (NWF-CLS) at 12 months using a two-level LMM

	Primary NWF-CLS analysis		Overdispersed / zero inflated		FSM individuals only		FSM interaction	
Coefficient	*IRRs*	*CI*	*IRRs*	*CI*	*IRRs*	*CI*	*IRRs*	*CI*
(Intercept)	11.4***	7.51, 17.60	11.35***	8.49, 15.18	8.98***	5.44, 14.83	9.75***	6.34, 14.99
NWF-CLS baseline	1.02***	1.02, 1.03	1.02***	1.02, 1.03	1.03***	1.02, 1.04	1.03***	1.02, 1.03
Intervention	0.96	0.75, 1.22	1.10	0.94, 1.29	1.08	0.88, 1.33	0.96	0.61, 1.52
School size	0.93	0.61, 1.43	1.13	0.85, 1.49	1.31	0.80, 2.13	1.21	0.90, 1.61
FSM							1.03	0.85, 1.26
Intervention x FSM							1.12	0.83, 1.52
Zero-inflated model								
(Intercept)			0.17***	0.12,0.24	0.16***	0.10, 0.26	0.17***	0.12, 0.24
Random effects								
σ^2	0.07		0.38		0.40		0.39	
τ_{00}	0.16 School_ID		0.01 School_ID		0.00016 School_ID		0.0021 School_ID	
ICC	0.71		0.02		<0.000		0.01	
N	46 School_ID		46 School_ID		44 School_ID		45 School_ID	
Observations	260		260		141		253	
Marginal R^2 / Conditional R^2	0.227 / 0.773		0.156 / 0.174		0.205 / 0.205		0.180 / 0.184	

** p<0.05; ** p<0.01; *** p<0.001.*

NWF-Nonsense word fluency - words read correctly (WRC)

The NWF-WRC for words read correctly was highly skewed and contained an inflated number of zero scores, so the specified GLMM (Poisson) model was checked for overdispersion and zero inflation, and was again present in both cases (overdispersion: dispersion ratio=4.316, Pearson's χ^2=1100.682, *p*<0.001; zero inflation: Observed zeros: 107, Predicted zeros: 24, Ratio: 0.22) Given that we have a significant test of overdispersion and that the ratio of observed to predicted zeros is well outside the tolerance of 5% or a ratio of 0.95, we conclude that the number of zeros is inflated and there is some overdispersion, so need to be adjusted in the main analysis using a zero-inflated negative binomial model. This model provides a more accurate estimate of all parameter estimates and we can see the minor adjustment in the data in Table 18.

In all models reported in Table 18, we do not find evidence of a significant difference between intervention and control (business as usual) (IRR=1.14 95% CI [0.78, 1.69], *p*=0.495), regardless of whether we are adjusting for the moderating effect of FSM eligibility. As expected, we find that baseline NWF-WRC is a significant predictor of NWF-WRC at follow-up (IRR=1.08, 95% CI [1.06, 1.09], *p*<0.0001). When only considering those with FSM eligibility status, we see some evidence of a significant effect of school size (IRR=3.87, 95% CI [1.30, 11.50], *p*=0.015), indicating pupils in larger schools are 3.87 times likely to score higher on the NWF-WRC if they are eligible for FSM.

The unconditional ICC (95% CI) associated with school for the secondary outcome, NWF-WRC (Poisson) is 0.651 (0.515, 0.740), and the conditional ICC (95% CI) is 0.543 (0.342, 0.625).

For the zero-inflated negative binomial model, both the unconditional and conditional ICC (95% CI) associated with school for the secondary outcome, NWF-WRC could not be calculated as the random effect variance is so small. Similarly, the FSM individual only and FSM interaction models could also not be calculated as the random effect variance was too small.

Table 18: Secondary analysis for DIBELS® nonsense word fluency – words read correct (NWF-WRC) at 12 months using a two-level LMM

	Primary NWF-WRC analysis		Overdispersed / zero inflated		FSM individuals only		FSM interaction	
Coefficient	**IRRs**	**CI**	**IRRs**	**CI**	**IRRs**	**CI**	**IRRs**	**CI**
(Intercept)	2.28*	1.16, 4.47	3.27***	1.97, 5.42	1.28***	0.43, 3.82	4.65***	1.98, 10.96
NWF-WRC baseline	1.08***	1.06, 1.09	1.09***	1.05, 1.13	1.07***	1.03, 1.11	1.09***	1.05, 1.13
Intervention	1.14	0.78, 1.69	1.10	0.84, 1.44	0.99	0.73, 1.35	0.77	0.33, 1.80
School size	0.98	0.50, 1.94	1.20	0.74, 1.97	3.87 *	1.30, 11.50	1.07	0.62, 1.83
FSM							0.86	0.59, 1.25
Intervention x FSM							1.26	0.73, 2.16
Zero-inflated model								
(Intercept)			0.52***	0.36, 0.75	0.68*	0.47, 0.99	0.53***	0.37, 0.75
Random effects								
σ^2	0.31		0.80		0.66		0.79	
τ_{00}	0.36 School_ID		<0.0001 School_ID		<0.0001 School_ID		<0.0001 School_ID	
ICC	0.54							
N	46 School_ID		46 School_ID		44 School_ID		45 School_ID	
Observations	260		260		141		253	
Marginal R^2 / Conditional R^2	0.134 / 0.604		0.154 / N/A		0.217 / N/A		0.151 / N/A	

** p<0.05; ** p<0.01; *** p<0.001.*

Oral reading fluency (ORF-WRC)

The ORF-WRC – words correct was again highly skewed and contained an inflated number of zero scores, so the specified GLMM (Poisson) model was checked for overdispersion and zero inflation, and was again present in both cases (overdispersion: dispersion ratio=10.090, Pearson's χ^2=1463.056, p<0.001; zero inflation: Observed zeros: 10, Predicted zeros: 2, Ratio: 0.20) Given that we have a significant test of overdispersion and that the ratio of observed to predicted zeros is well outside the tolerance of 5% or a ratio of 0.95, we conclude that the number of zeros is inflated and there is some overdispersion, so need to be adjusted in the main analysis using a zero-inflated negative binomial model. This model provides a more accurate estimate of all parameter estimates and we can see the minor adjustment in the data in Table 19.

In all models reported in Table 19, we do not find evidence of a significant difference between intervention and control (business as usual) (IRR=0.62, 95% CI [0.37, 1.04], p=0.068), regardless of whether we are adjusting for the moderating effect of FSM eligibility. As expected, we find that baseline ORF-WRC is a significant predictor (IRR=1.02, 95% CI [1.02, 1.02], p<0.0001), but we also find some evidence of the effect of school size (IRR=3.24, 95% CI [1.24, 8.47], p=0.0166). However, this school size effect reduces to a statistically non-significant level when models are adjusted for overdispersion and zero inflation, which can be confirmed with the coverage of the CIs containing 1.

The unconditional ICC (95% CI) associated with school for the secondary outcome, Oral Reading fluency – words read correctly (Poisson) is 0.926 (0.886, 0.947), and the conditional ICC (95% CI) is 0.923 (0.876, 0.945).

For the zero-inflated negative binomial model, the unconditional ICC (95% CI) was 0.057; however, conditional ICC (95% CI) associated with school for the secondary outcome, ORF-WRC could not be calculated as the random effect variance was so small. Similarly, the FSM individual only and FSM interaction models could also not be calculated as the random effect variance was too small.

Table 19: Secondary analysis for DIBELS® oral reading fluency – words read correct (ORF-WRC) at 12 months using a two-level LMM

	Primary ORF-WRC (correct) analysis		Overdispersed / zero inflated		FSM individuals only		FSM interaction	
Coefficient	**IRRs**	**CI**	**IRRs**	**CI**	**IRRs**	**CI**	**IRRs**	**CI**
(Intercept)	6.37***	2.47, 16.38	21.01***	9.47, 46.62	25.29**	4.57, 139.99	19.72***	6.49, 59.90
ORF-WRC baseline	1.02***	1.02, 1.02	1.02***	1.02, 1.03	1.02***	1.01, 1.03	1.02***	1.02, 1.03
Intervention	0.62	0.37, 1.04	0.88	0.65, 1.19	1.08	0.73, 1.61	1.49	0.62, 3.57
School size	3.24*	1.24, 8.47	0.96	0.42, 2.17	0.75	0.13, 4.24	0.86	0.35, 2.12
FSM							1.10	0.75, 1.60
Intervention x FSM							0.71	0.40, 1.24
Zero-inflated model								
(Intercept)			0.05 ***	0.02,0.14	0.07***	0.02-0.19	0.05***	0.02, 0.14
Random effects								
σ^2	0.06		1.51		1.49		1.54	
τ_{00}	0.67 $_{\text{School\_ID}}$		<0.001 $_{\text{School\_ID}}$		<0.001 $_{\text{School\_ID}}$		<0.001 $_{\text{School\_ID}}$	
ICC	0.92							
N	43 $_{\text{School\_ID}}$		43 $_{\text{School\_ID}}$		33 $_{\text{School\_ID}}$		42 $_{\text{School\_ID}}$	
Observations	150		150		78		148	
Marginal R^2 / Conditional R^2	0.332 / 0.949		0.174 / N/A		0.174 / N/A		0.180 / N/A	

** p<0.05; ** p<0.01; *** p<0.001. .*

Oral reading fluency (ORF) – errors made

The ORF – errors made, demonstrated little skew and did not appear to contain an inflated number of zero scores (when looking at the plot). However, inspection of the overdispersion and zero-inflation tests indicated both were present (overdispersion: dispersion ratio=2.168, Pearson's χ^2=314.375, p<0.001; zero inflation: Observed zeros: 8, Predicted zeros: 1, Ratio: 0.12). Given that we have a significant test of overdispersion and that the ratio of observed to predicted zeros is well outside the tolerance of 5% or a ratio of 0.95, we conclude that the number of zeros is inflated and there is some overdispersion, so need to be adjusted in the main analysis using a zero-inflated negative binomial model. This model provides a more accurate estimate of all parameter estimates and we can see the minor adjustment in the data in Table 20.

In models reported in Table 20, we find evidence of a significant effect of intervention over control (business as usual) (IRR=0.77, 95% CI [0.60, 0.98], *p*=0.0368) in our primary analysis, indicating a decrease in the intervention group (i.e. making fewer errors), but this evidence appears to disappear in all other models (regardless of whether we are adjusting for the moderating effect of FSM eligibility). We find that baseline ORF – errors made, is a significant predictor (IRR=1.02, 95% CI [1.00, 1.04], *p*<0.0001), but this finding appears to be inconsistent across models. We also find a marginal effect of school size when only considering those with FSM eligibility (IRR=0.45, 95% CI [0.21, 0.98], *p*=0.0449). However, this school size effect is generally statistically non-significant in all other models (reflected in the coverage of the CIs containing 1).

The unconditional ICC (95% CI) associated with school for the secondary outcome, ORF – errors made (Poisson) is 0.597 (0.422, 0.682), and the conditional ICC (95% CI) is 0.410 (0.200, 0.487).

For the zero-inflated negative binomial model, the unconditional ICC (95% CI) was 0.175; however, conditional ICC (95% CI) associated with school for the secondary outcome, ORF – errors made, could not be calculated as the random effect variance was so small. Similarly, the FSM individual only and FSM interaction models could also not be calculated as the random effect variance was too small.

Table 20: Secondary analysis for DIBELS® oral reading fluency (ORF) – errors made, at 12 months using a two-level LMM

	Primary ORF – errors made analysis		**Overdispersed / zero inflated**		**FSM individuals only**		**FSM interaction**	
Coefficient	*IRRs*	*CI*	*IRRs*	*CI*	*IRRs*	*CI*	*IRRs*	*CI*
(Intercept)	7.83***	5.11, 11.98	7.82***	4.74, 12.91	10.21***	4.45, 23.41	8.55***	4.23, 17.27
ORF – errors made baseline	1.02*	1.00, 1.04	1.02	0.99, 1.05	1.05*	1.00, 1.10	1.02	0.99, 1.05
Intervention	0.77*	0.60, 0.98	0.81	0.64, 1.04	0.80	0.58, 1.12	0.71	0.35, 1.46
School size	0.72	0.48, 1.09	0.72	0.46, 1.15	0.45*	0.21 ,0.98	0.74	0.45, 1.20
FSM							0.94	0.69, 1.27
Intervention x FSM							1.09	0.69, 1.73
Zero-inflated model								
(Intercept)			0.01	0.00,5.11	0.00	-	0.01	0.00, 2.55
Random effects								
σ^2	0.14		0.34		0.35		0.34	
τ_{00}	0.10 $_{School\_ID}$		0.01 $_{School\_ID}$		<0.0001 $_{School\_ID}$		0.01 $_{School\_ID}$	
ICC	0.41		0.03		-		0.03	
N	43 $_{School\_ID}$		43 $_{School\_ID}$		33 $_{School\_ID}$		42 $_{School\_ID}$	
Observations	150		150		78		148	
Marginal R^2 / Conditional R^2	0.115 / 0.478		0.072 / 0.100		0.148 / N/A		0.068 / 0.094	

** p<0.05; ** p<0.01; *** p<0.001.*

Word Reading Fluency (WRF)

The WRF was again highly skewed and contained an inflated number of zero scores, so the specified GLMM (Poisson) model was checked for overdispersion and zero inflation, and was again present in both cases (overdispersion: dispersion ratio=4.028, Pearson's χ^2=1015.018, *p*<0.001; zero inflation: Observed zeros: 45, Predicted zeros: 2, Ratio: 0.04). Given that we have a significant test of overdispersion and that the ratio of observed to predicted zeros is well outside the tolerance of 5% or a ratio of 0.95, we conclude that the number of zeros is inflated and there is some overdispersion, so need to be adjusted in the main analysis using a zero-inflated negative binomial model. This model provides a more accurate estimate of all parameter estimates and we can see the minor adjustment in the data in Table 21.

In all models reported in Table 21, we do not find evidence of a significant difference between intervention and control (business as usual) (IRR=0.96, 95% CI [0.77, 1.20], *p*=0.709), regardless of whether we are adjusting for the moderating effect of FSM eligibility. As expected, we find that baseline WRF is a significant predictor (IRR=1.08, 95% CI [1.06, 1.11], p<0.0001). When considering only those with FSM eligibility, we also find evidence of an effect of school size (IRR=2.08, 95% CI [1.07, 4.02], p=0.030).

The unconditional ICC (95% CI) associated with school for the secondary outcome, WRF (Poisson) is 0.751 (0.657, 0.812), and the conditional ICC (95%CI) is 0.481 (0.275, 0.570).

For the zero-inflated negative binomial model, the unconditional ICC (95% CI) associated with school for the secondary outcome, WRF is 0.085 (R is unable to produce CIs for this model ICC—code does not exist currently), and the conditional ICC (95% CI) is 0.020. For the FSM individual only and FSM interaction models, the conditional ICCs are 0.01 and 0.026, respectively. Once more, all models using zero-inflated negative binomial do not report associated ICC CIs as the statistical software is not yet able to produce these intervals to the authors knowledge.

Table 21: Secondary analysis for DIBELS® word reading fluency (WRF) at 12 months using a two-level LMM

	Primary WRF analysis		Overdispersed / zero inflated		FSM individuals only		FSM interaction	
Coefficient	**IRRs**	**CI**	**IRRs**	**CI**	**IRRs**	**CI**	**IRRs**	**CI**
(Intercept)	5.89***	4.00, 8.66	7.05***	4.86, 10.22	3.04***	1.56, 5.93	7.14***	4.03, 12.65
WRF baseline	1.08***	1.08, 1.09	1.08***	1.06, 1.10	1.08***	1.06, 1.11	1.08***	1.06, 1.10
Intervention	0.96	0.77, 1.20	0.99	0.81, 1.20	0.90	0.69, 1.18	0.76	0.42, 1.38
School size	0.80	0.54, 1.17	0.83	0.58, 1.20	2.08*	1.07, 4.02	0.88	0.59, 1.31
FSM							0.96	0.74, 1.23
Intervention x FSM							1.19	0.82, 1.74
Zero-Inflated Model								
(Intercept)			0.17***	0.11, 0.26	0.24***	0.15, 0.39	0.17***	0.11, 0.26
Random Effects								
σ^2	0.13		0.56		0.63		0.56	
τ_{00}	0.12 School_ID		0.01 School_ID		0.01 School_ID		0.026 School_ID	
ICC	0.48		0.02		0.01		0.03	
N	46 School_ID		46 School_ID		44 School_ID		45 School_ID	
Observations	257		257		140		250	
Marginal R^2 / Conditional R^2	0.511 / 0.746		0.288/0.302		0.289/0.296		0.292/0.311	

** p<0.05; ** p<0.01; *** p<0.001.*

Reading Self Concept (RSCS)

An LMM rather than an ordinal mixed model, as specified in the SAP, was fitted to the reading self-concept measure. This was due to a misunderstanding of the scoring criteria for the RSCS. Within the SAP, scoring was considered to provide a total score between 0 and 6, so an ordinal model seemed most appropriate. However, after review with the evaluation team, scoring was corrected to be between 0 and 12; therefore, an ordinal model was not suitable for this outcome. Instead, we have followed the same analysis as the primary outcome and reported accordingly.

Table 22 presents the results for the RSCS and including the FSM eligibility subgroup analyses. No evidence of a significant difference between intervention and control (business as usual) was found in all of the models in Table 21 (primary RSCS analysis, β=-0.07, 95% CI [-0.56, 0.42], t (27.837)=-0.275, p=0.785, hedges' g=-0.04, 95% CI [-0.35, 0.26]). We also note the level of uncertainty in the estimate given the width of the CIs and that they cover zero. The negative effect is likely due to the level of uncertainty and so does not reflect evidence of a negative trial.

We also fitted a robust LMM, given that some skew was observed in this outcome. The pattern of results remained unaltered, but the magnitude of the intervention effect was larger and showed a positive effect of intervention (β=0.13, 95% CI [-0.25, 0.50], t (27.84)=0.653, p=0.519; Hedges' g=0.088). As expected, baseline RSCS showed evidence as a significant predictor across all models (β=0.21, 95% CI [0.10, 0.32], t (238.38)=3.735, p<0.001; Hedges' g=0.13, 95% CI [0.06, 0.2]).

The unconditional ICC (95% CI) associated with school for the secondary outcome, RSCS is 0.160 (0.061, 0.276), and the conditional ICC (95% CI) is 0.006 (0.006, 0.220). Robust estimates of the random intercept variance in the robust models (primary, FSM only, and FSM*intervention interaction) were approximately zero (to 3dp in R output), therefore, corresponding ICC were similarly near zero.

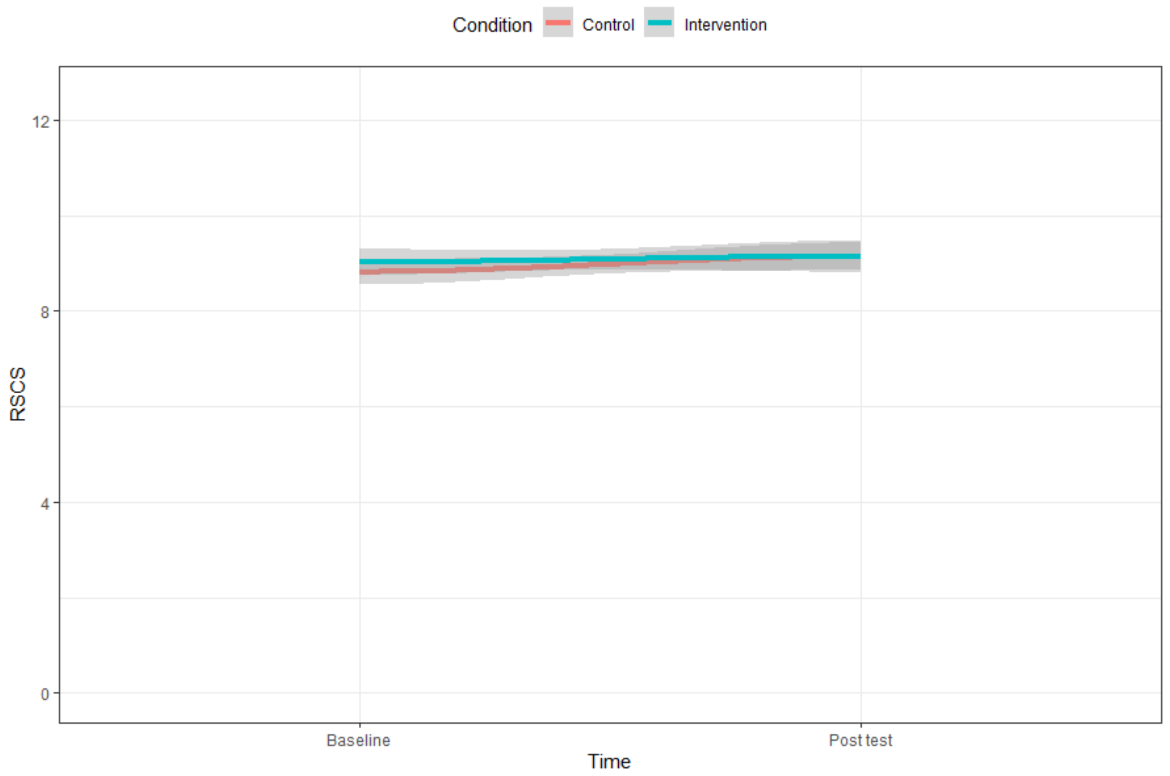
Table 22: Secondary analysis for the Reading Self-Concept Scale (RSCS) at 12 months using a two-level LMM

	Primary RSCS analysis		Primary RSCS (robust)		FSM individuals only		FSM interaction (robust)	
Coefficient	*Estimates*	*CI*	*Estimates*	*CI*	*Estimates*	*CI*	*Estimates*	*CI*
(Intercept)	6.03***	4.67, 7.40	6.28***	5.11, 7.45	4.86***	2.98, 6.73	6.31***	4.86, 7.75
RSCS total baseline	0.21***	0.10, 0.32	0.21***	0.11, 0.31	0.27***	0.13, 0.41	0.20***	0.09, 0.30
Condition (intervention)	-0.07	-0.56, 0.42	0.13	-0.25, 0.50	0.17	-0.33, 0.67	0.02	-1.15, 1.19
School size (large)	1.33**	0.40, 2.26	1.19**	0.44, 1.94	2.06**	0.62, 3.50	1.28**	0.46, 2.10
FSM							-0.02	-0.53, 0.48
Intervention x FSM							0.07	-0.70, 0.85
Random effects								
σ^2	2.41		2.04				2.09	
τ_{00}	0.20 School_ID		0.00 School_ID				0.00 School_ID	
ICC	0.08		0.00				0.00	
N	46 School_ID		46 School_ID				45 School_ID	
Observations	243		243		132		236	
Marginal R^2 / Conditional R^2	0.094 / 0.165		0.104 / 0.104		N/A		0.099 / 0.099	

** p<0.05; ** p<0.01; *** p<0.001.*

We do observe that, when plotting the pre-post-test changes, on average the intervention group appears to start with higher scores at baseline than the control group (Figure 7). There is then relatively little change, on average in the intervention group, but the controls catch up.

Figure 7: Mean pre-post-test changes in both trial arms. Red line is the control arm and blue is the intervention arm. Standard errors are shown as grey ribbons



Analysis in the presence of non-compliance

An instrumental variables approach was used to evaluate whether differing levels of compliance differed to the primary ITT analysis of DIBELS® composite at 12 months. Three compliance ratings were outlined in the protocol based on a six-point checklist: i) full compliance (all items adhered); ii) partial compliance – scenario A; and iii) partial compliance – scenario B. These compliance ratings were used as the instrumental variable and adjusted estimates of the intervention effect were estimated under each compliance scenario (Table 23) and compared to our primary ITT analysis.

The pattern of evidence across the different models in Table 23 was consistent; however, the magnitudes of intervention effect were different across the models (despite not being statistically significant and potentially reflecting the uncertainty in the estimates). Larger effects were observed in the full compliance and partial compliance scenarios with the largest effect seen with partial compliance scenario B (β=2.09, 95% CI [-1.25, 5.42], p=0.219; Hedges' g=0.14, 95% CI [-0.08, 0.38]) and smallest effect in the ITT. Partial compliance scenario A showed a smaller difference (β=0.48, 95% CI [-4.06, 5.03], p=0.835; Hedges' g=0.02, 95% CI [-0.21, 0.26]) than full compliance (β=1.07, 95% CI [-4.04, 6.17], p=0.681; Hedges' g=0.04, 95% CI [-0.18, 0.28]). Note Cohen's d effect sizes were converted to Hedges' g following Borenstein *et al.* (2011).

Table 23: Comparing compliance estimates of intervention effects between intention-to-treat and different levels of compliance. A two-level LMM is used for the primary analysis, but all compliance instrumental variable regressions report robust standard errors instead

	Primary		IV full		IV partial A		IV partial B	
Coefficient	*Estimates*	*CI*	*Estimates*	*CI*	*Estimates*	*CI*	*Estimates*	*CI*
(Intercept)	-15.70	-55.63, 24.24	-11.48	-51.69, 28.72	-12.98	-52.70, 26.74	-8.88	-47.98, 30.23
DIBELS® composite baseline	1.05***	0.95, 1.15	1.04***	0.94, 1.14	1.04***	0.94, 1.15	1.03***	0.93, 1.13
School size	2.25	-2.66, 7.17	2.66	-1.78, 7.11	2.63	-1.80, 7.07	2.72	-1.74, 7.17
Condition (intervention)	-0.09	-3.00, 2.82	1.07	-4.04, 6.17	0.48	-4.06, 5.03	2.09	-1.25, 5.42

Random effects				
σ^2	110.85			
τ_{00}	4.68 $_{School\_ID}$			
ICC	0.04			
N	46 $_{School\_ID}$			
Observations	280	280	280	280
Marginal R^2 / Conditional R^2	0.624 / 0.639	0.624 / 0.620	0.625 / 0.621	0.622 / 0.618

** p<0.05; ** p<0.01; *** p<0.001.*

Missing data analysis

The first step in the imputation process was to investigate if any variables were associated with a binary indicator of missingness in the primary outcome. We dummy coded a binary (1= missing, 0= not) missingness indicator and used this as the outcome variable in a logistic regression with predictors: DIBELS® baseline, FSM eligibility, SEND status, and English as an Additional Language (EAL) status. Table 24 present results from this regression and appeared to show that none of these variables were associated with missingness in the primary outcome.

Following guidance from van Buuren (2018) and Austin *et al.* (2021), we included the variables include in the primary analysis as default. Therefore, we proceeded to impute missing component scores of the DIBELS® (LNF, PSF, NWF, ORF, and WRF) using the primary adjustment set of covariates from the primary analysis, and any available component level data among these as they are likely correlated to the composite score and with each other (Plumpton *et al.*, 2016). Therefore, the DIBELS® composite score could be calculated from the complete component data. On this basis, any missingness in component scores used information from the remaining components given their correlations among each other. Exceptions were when individual components were too highly correlated and caused the imputation models estimation problems as a result of multicollinearity (R mice [Multivariate Imputation by Chained Equations] package identifies and deals with this automatically by omitting one of the predictors). Specifically, this affected the ORF measures (number correct, number of errors, and accuracy), and in addition, we found that the WRF was also relatively highly correlated.

Table 24: Logistic regression with missing data status of DIBELS® composite at 12 months as primary outcome and predictors of missingness: SEND, baseline DIBELS®, EAL as per SAP

	DIBELS® composite (missing or not)	
Coefficient	*Odds ratios*	*SE*
(Intercept)	76.34	382.64
DIBELS® composite baseline	0.98	0.01
FSM	1.29	0.36
SEND SpLD	0.35	0.39
SEND MLD	1.01	0.55
SEND SLD	1.04	0.39
SEND PMLD	0.00	0.00
SEND SLCN	0.76	0.31
SEND ASD	1.84	0.68
SEND Other	1.62	0.99
SEND PD	0.72	0.50
SEND SEMH	0.00	0.00
EAL	1.43	0.60
Observations	340	
R^2 Tjur[a]	0.042	

*[a] Tjur, T. (2009); * p<0.05; ** p<0.01; *** p<0.001. ASD=Autism Spectrum Disorder; DIBELS®=Dynamic Indicators of Basic Early Literacy Skills®, Eighth Edition; EAL=English as an Additional Language; FSM=free school meals; MLD=Moderate Learning Difficulty; PD=Physical Disability; PMLD=Profound and Multiple Learning Disability; SAP=statistical analysis plan; SE=standard error; SEMH=Social, Emotional, and Mental Health; SEND=special educational needs and disability; SLCN=Speech, Language, and Communication Needs; SLD=Severe Learning Difficulty; SpLD=Specific Learning Difficulty.*

Table 25 shows the comparison of primary analysis model with the pooled results from 20 imputed datasets. The pattern of findings is unaltered in the pooled results compared to the primary analysis but there is a noticeable difference in the value of the intercepts. Reasoning for this discrepancy between the primary and the imputed results relates to the composite scoring and the volume of zeros in the components scores used to calculate the composite. When the imputation is used, scores are moved away from zero in the individual measures (zero is given if no score recorded, or missing, in the composite), but when imputed, the composite score increases substantially. Therefore, genuine zeros for no score points scored or a discontinuation are not imputed as they are correctly recorded as zero, only genuinely recorded missing values are imputed in this analysis. To be able to calculate composite scores for the primary analysis, missing values were inserted as zeros, so that they made no contribution to the overall composite score. The comparisons of the missing data analysis and the primary analysis, then compares when these scores are treated as a zero contribution or imputed. This artificially increases the difference in intercept and to some extent the other estimated parameters as the whole distribution is artificially shifted as a result. Method of imputation varied for different variables: i) component measures used either a zero-inflated negative binomial or negative binomial depending on the extent of the zero inflation (‘zinb’ or ‘nb’ in R mice package; Note, imputation using zinb for all component scores gave some convergence issues, so distributions with less severe zero inflation used negative binomial instead); ii) school size, school year, trial arm, EAL, and FSM used predictive mean matching (‘pmm’ in R mice package). Convergence was assessed by plotting the trace plots and through visual inspection determined stationarity (see Appendix F). The number of iterations was also increased from the default to ensure stability, from 20 to 100. It must be noted that both ORF measures had around 45% missing data, so imputation was more challenging for these variables and is likely the biggest contributor to the change. In terms of effect sizes and associated uncertainty, we report Hedges’ g= -0.09 and 95% CI [-0.34, 0.19], which when compared to the primary analysis effect size, Hedges’ g=-0.01, 95% CI [-0.28, 0.26] is relatively similar, Considering that the CI contains zero, we cannot rule out that no intervention effect is present. The effect size is tiny when information is imputed and the CIs are sufficiently wide to indicate considerable uncertainty.

Table 25: Missing data sensitivity analyses comparing results from primary analysis to pooled imputed results from 20 imputed datasets

	Primary analysis		**Imputed analysis (pooled results)**	
Coefficient	***Estimates***	***SE***	***Estimates***	***CI***
Intercept	-15.70	-54.80, 23.83	79.953^{**}	29.60, 118.31
DIBELS® baseline	1.05^{***}	0.95, 1.15	0.831^{***}	0.72, 0.94
School size	2.25	-2.59, 7.08	-2.494	-8.73, 3.75
Intervention	-0.09	-2.93, 2.75	-1.054	-4.31, 2.20
σ^2	110.85		124.87	
τ_{00}	4.68 School_ID		15.697 School_ID	
ICC	0.04		0.112	

** p<0.05; ** p<0.01; *** p<0.001.*

Figure 8 and Figure 9 show the density plots of the original distributions (blue) and the imputed datasets (red) at baseline and post-test, respectively. It can clearly be seen that the imputations do not significantly deviate from the original distributions.

Figure 8: Density plots showing the distributions of the original data (blue) and the 20 imputed data sets (red) at baseline

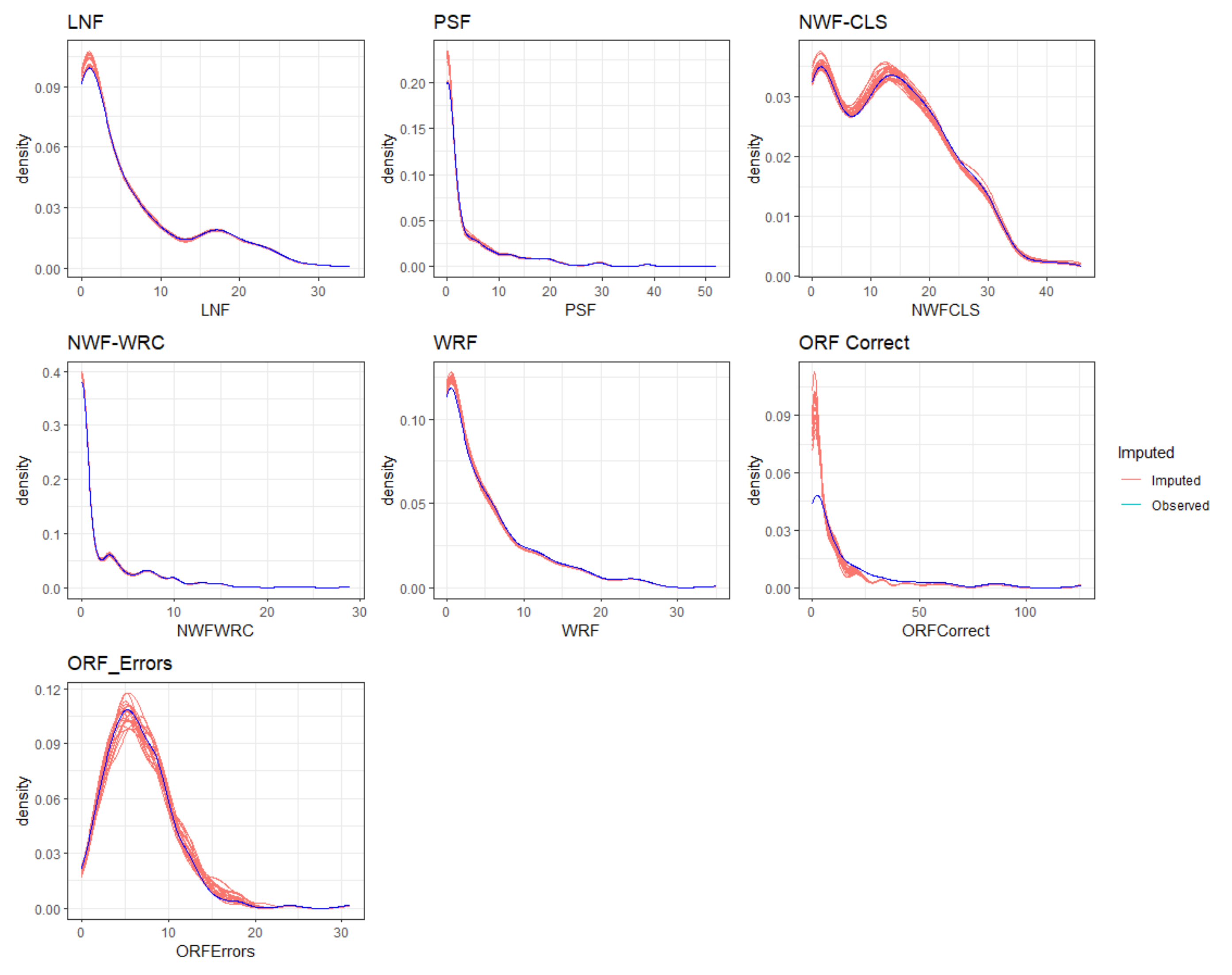
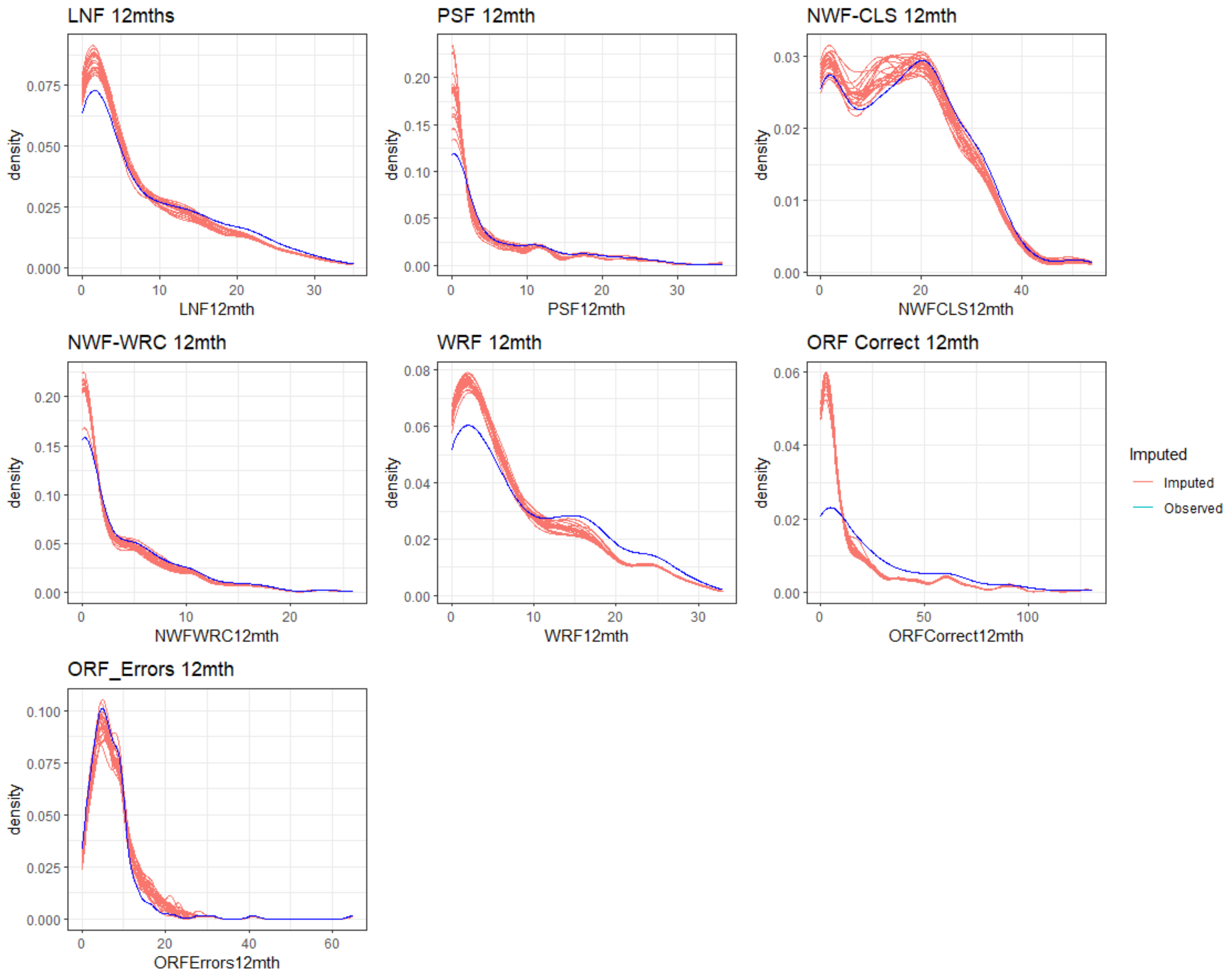


Figure 9: Density plots showing the distributions of the original data (blue) and the 20 imputed data sets (red) at post-test



Sub-group analyses

Planned subgroup analyses were conducted with pupils that were eligible for FSM in both primary and secondary outcomes (secondary outcomes have been reported in the secondary analysis section for brevity). Given our results from the primary analysis were reported with robust estimates, we also report the FSM subgroup analyses using robust estimators. Across all analyses (Table 26), we were still unable to find evidence against the hypothesis that intervention and control (business as usual) are equal (Primary, β=-0.09, 95% CI [-3.00, 2.82], t (44.69)=-0.062, *p*=0.951, Hedges' g=-0.01, 95% CI [-0.28, 0.26]; FSM only, β=0.67, 95% CI [-2.28, 3.63], t (38.71)=0.447, *p*=0.657, Hedges' g=0.09, 95% CI [-0.3, 0.48]; FSM moderation, β=-0.40, 95% CI [-4.45, 3.66], t (229.35)=-0.193, *p*=0.991, Hedges' g=-0.05, 95% CI [-0.55, 0.45]), but the magnitude of effect was somewhat larger than in the primary analysis; however all CIs were relatively wide and contained zero.

Table 26: Subgroup analyses for free FSM status comparing primary analysis, primary analysis with only those eligible for FSM, and FSM as a moderator in the primary analysis

	Primary analysis		FSM individuals only (robust)		FSM interaction (robust)	
Coefficient	***Estimates***	***CI***	***Estimates***	***CI***	***Estimates***	***CI***
Intercept	-15.70	-55.63, 24.24	-37.98***	-77.28, 1.32	-30.95*	-61.13, -0.76
DIBELS® baseline	1.05 ***	0.95, 1.15	1.10***	1.00 ,1.20	1.09***	1.01, 1.16
School size	2.25	-2.66, 7.17	3.42	-2.60, 9.44	2.46	-1.14, 6.06
Intervention	-0.09	-3.00, 2.82	0.67	-2.28, 3.63	0.83	-2.23, 3.90
FSM					0.28	-2.33, 2.88

Intervention x FSM					-0.40	4.45, 3.66
Random effects						
σ^2	110.85		51.80		66.07	
τ_{00}	4.68 School_ID		6.18 School_ID		~0.00 School_ID	
ICC	0.04		0.11		~0.00	
N	46 School_ID		45 School_ID		45 School_ID	
Observations	280		152		273	
Marginal R^2 / Conditional R^2	0.624 / 0.639		0.781 / 0.805		0.761 / 0.761	

** p<0.05; ** p<0.01; *** p<0.001.*

Additional analyses and robustness checks

The primary analysis was extended with the addition of FSM, pupil primary need (within their identification as SEND) and EAL covariates to test if the robustness of any intervention effect was not conditional (Table 26). This analysis does not appear to support the number of parameters estimated for the given sample size as the width of CIs around the estimates is very wide. In addition, when fitting the model, SEND categories with very low cell counts were removed as they caused additional problems in the model estimation (PMLD; multi-disabled sensory impaired; visual impairment; and hearing impairment). An additional analysis was conducted removing the SEND variables and their interactions from the model, reducing the number of parameters to estimate alleviating issues with lack of sample size to sensibly estimate parameters. This was motivated by concern for the original analysis by reviewers and the logic of inclusion of SEND classifications in a sample that by definition is defined as having a SEND status being in a special school context.

For the original SAP defined analysis, the point estimate of the trial arm indicator variable increased (β=12.03, 95% CI -6.79, 30.85], t (122.21)=1.259, *p*=0.30, Hedges' g=1.15, 95% CI [-0.86, 1.45]), there is considerable uncertainty as reflected in the width of the CIs and likely as a result of insufficient sample size. The analysis did, however, indicate weak evidence of a significant effect of SEND for Social, Emotional, and Mental Health (SEMH) differs from those without that particular SEND (β=11.50, 95% CI [1.52, 21.47] t (231.31)=2.271, *p*=0.024; Hedges' g=1.11, 95% CI [0.11, 2.04]), noting the substantial coverage of the CI, but not in the interaction between it and the intervention arm (β=-14.23, 95% CI [-29.22, 0.75], t (239.45)=-1.871, *p*=0.0705; Hedges' g=-1.37, 95% CI [-2.8, 0.07]). Both results should be very cautiously interpreted as they likely reflect the insufficiency of this model, so we do not draw any firm conclusions from this result on that basis.

Following the primary analysis, we also re-ran the analysis with robust estimation due to the skewed distribution. We found that change in the intervention effect size was relatively similar but there were additional small significant effects of SEND category interactions with intervention, (MLD, β=-9.93, 95% CI [-18.13, -1.73], t (208.06)=-2.374, *p*= 0.018; Hedges' g=-1.26 95% CI [-2.3, -0.22]; ASD, β=-7.52, 95% CI [-13.33, -1.70], t (146.07)=-2.532, *p*=0.012; Hedges' g=-0.95 95% CI [-1.69, -0.22]; SEMH, β=-12.36, 95% CI [-23.94, -0.78], t (239.45)=-2.092, *p*=0.037; Hedges' g=-1.57, 95% CI [-3.04, -0.1]). Similarly, to the SAP specified model, this model is likely problematic as the sample size is unlikely to support the number of parameters estimated.

The additional analysis omitting the SEND variables has been added to the results in Table 27, for a more accurate estimate of the EAL and FSM. We found that the point estimate of the trial arm indicator variable increased (β=3.25, 95% CI [-4.36, 10.86], t (122.21) =0.838, *p*=0.408, Hedges' g=0.40, 95% CI [-0.53, 1.33]); however, the intervention effect still shows no evidence of a difference between trial arms.

Table 27: Additional analyses that extend the primary analysis and robust analyses with the inclusion of additional covariates: SEND, EAL, and their interactions with trial arm

	Additional primary analysis		Additional robust analysis		*Additional robust analysis (no SEND)*	
Coefficient	*Estimates*	*CI*	*Estimates*	*CI*	*Estimates*	*CI*
(Intercept)	-25.25	-70.06, 19.57	-43.05**	-76.57, -9.52	*-29.32*	*-60.38, 1.74*
DIBELS® composite baseline	1.06***	0.95, 1.17	1.11***	1.02, 1.19	*1.08****	*1.00, 1.16*
Condition (intervention)	12.03	-6.79, 30.85	9.13	-4.12, 22.39	*3.25*	*-4.36, 10.86*

School size (large)	8.36	-1.57, 18.28	6.43	-0.19, 13.05	*4.16*	*-1.51, -9.82*
FSM	-0.42	-3.83, 3.00	0.21	-2.43, 2.84	*0.20*	*-2.46, 2.87*
SEND SpLD	0.79	-8.71, 10.28	0.96	-6.29, 8.21		
SEND MLD	5.87	-3.17, 14.91	4.13	-2.85, 11.11		
SEND SLD	-0.31	-5.88, 5.25	-0.21	-4.39, 3.96		
SEND SLCN	-0.36	-6.32, 5.61	-0.99	-5.44, 3.47		
SEND ASD	4.51	-1.06, 10.08	3.60	-0.57, 7.76		
SEND Other	2.97	-5.25, 11.19	2.44	-3.74, 8.62		
SEND PD	-1.72	-9.61, 6.17	-1.92	-7.89, 4.04		
SEND SEMH	11.50*	1.52, 21.47	6.78	-0.88, 14.44		
EAL	-1.84	-6.95, 3.26	-2.37	-6.26, 1.53	*-1.95*	*-5.91, 2.00*
Condition (intervention) * School size (large)	-4.95	-20.05, 10.16	-2.46	-12.76, 7.84	*-2.58*	*-10.35, 5.19*
Condition (intervention) * FSM	1.28	-4.03, 6.60	0.02	-4.10, 4.13	*-0.21*	*-4.35, 3.92*
Condition (intervention) * SEND SpLD	2.32	-16.45, 21.08	3.27	-10.13, 16.68		
Condition (intervention) * SEND MLD	-11.03*	-21.81, -0.25	-9.93*	-18.13, -1.73		
Condition (intervention) * SEND SLD	-3.14	-11.78, 5.49	-2.50	-8.78, 3.79		
Condition (intervention) * SEND SLCN	-4.24	-12.41, 3.92	-3.44	-9.57, 2.68		
Condition (intervention) * SEND ASD	-8.38*	-16.21, -0.56	-7.52*	-13.33, -1.70		
Condition (intervention) * SEND Other	-8.61	-22.33, 5.10	-6.30	-16.54, 3.93		
Condition (intervention) * SEND PD	-3.86	-15.34, 7.62	-3.16	-11.93, 5.60		
Condition (intervention) * SEND SEMH	-14.23	-29.22, 0.75	-12.36*	-23.94, -0.78		
Condition (intervention) * EAL	0.70	-8.29, 6.89	2.14	-3.73,8.01	*-0.12*	*-5.93, 5.68*
Random effects						
σ^2	102.57		62.07		*66.53*	
τ_{00}	6.66 School_ID		0.00 School_ID		0.00 School_ID	
ICC	0.06		0.00		*0.00*	
N	44 School_ID		44 School_ID		44 School_ID	
Observations	269		269		*270*	
Marginal R^2 / Conditional R^2	0.656 / 0.677		0.777 / 0.777		*0.758 / 0.758*	

** p<0.05; ** p<0.01; *** p<0.001.*

Implementation and Process Evaluation

A total of 27 schools and 181 pupils were randomised to the implementation arm post-baseline data collection. One school (nine pupils) could not commit to the study post-randomisation. At the start of the academic year 2021/2022, 26 schools and 175 pupils agreed to take part in HERiSS. Four schools and 67 pupils in the intervention arm did not engage with the programme. For the 22 schools and the 108 pupils that did engage in the intervention arm, it was a challenging year with just ten pupils able to complete at least half of the programme. That notwithstanding, all teachers/TAs interviewed said that, given the opportunity, they would use HER® in the future. The IPE focuses on the factors that contributed to this lack of success and apparent contradiction in schools' experience of HERiSS.

Compliance

Key finding: Compliance was poor.

Compliance criteria included schools' attendance of the training provided, engagement with ISO support, and recommendations including use of the HERiSS support manual and timetabling three sessions per week per pupil. These are key implementation factors that facilitate Change process 1: Upskilling teaching staff to teach children with SEND to read and Change process 3: Teaching children with SEND early reading skills, outlined in the logic model.

Data gathered by the delivery team and ISO visits indicate that compliance was poor:

- six out of 26 schools (23%) were fully compliant (see Table 5 and compliance definition;
- two schools were Partial scenario A and scenario B compliant;
- no schools met Partial scenario A only;
- nine schools were Partial scenario B compliant only; and
- ten schools (38%) met none of the compliance criteria.

Eight schools provided no pupil tracking forms at all for ISOs to review during their visits and no school provided a full set.

Training

Training is a critical component of the HERiSS intervention, and an underlying question in the HERiSS logic model: Can teaching staff be upskilled to teach pupils with SEND the early reading skills necessary to become proficient at reading?

Training consisted of a set of materials and videos sent to schools at the start of the academic year 2021/2022, followed by attendance at two webinars in which the delivery team went through a detailed explanation of how HER® works and what was expected of schools, as well as answering any concerns they may have had.

Teachers/TAs and ISOs were complimentary about the HERiSS training using words like *'helpful'*, *'really good'*, and *'thorough'*. However, in the interview data seven teachers/TAs (out of 13) and four ISOs (out of five) observed that the training was a lot to take in, in a short space of time and potentially overwhelming. Participants commented on the fact that everything was received at the start of the academic year, which is the busiest time of year. It was clear that there was some confusion on the part of schools about how the training was meant to work. Not all schools appreciated the need to look through the training materials sent to schools before the webinar. Similarly, those who had not logged into HER® prior to the initial webinar also had difficulty understanding. One participant did point out, however, that the format meant she could go back and work through the training in her own time, which was helpful.

Three participants felt that the training was not sufficiently concrete, and it would have helped to have had more demonstrations, but this included one participant who had not been able to log in before the training session. Irrespective of any initial difficulties, six participants observed that once they had started using the programme the training made sense.

It was clear from one teacher/TA interview that not everyone who had been involved in the delivery of HER® had taken part in the training despite this being a requirement of participation. This was also commented on by one ISO.

Three ISOs reported that levels of engagement in the training varied with one observing that engagement in the webinars was not predictive of engagement in implementation. It was further noted by one ISO that some schools were sceptical during training and lacked 'buy-in' from the get-go, a key implementation factor (described below).

Using the activities outlined in the HER® manual

HERiSS includes a support manual developed by the delivery team with suggested activities and adaptations to the programme to help in cases where pupils need a bit of extra support.

Seven participants cited examples when they had used the manual and felt that it was a useful resource. The troubleshooting sections and motivational resources were particularly appreciated. Only one participant reported having to make further adaptations, but these were additional flashcards relevant to her pupil. The lack of a need to make further adaptations was cited by some as one of the factors that facilitated implementation (see below).

However, these data did not correspond with the data from the ISO interviews. One ISO said that there was no evidence of the schools that she was supporting having used the manual and two others reported instances of the school's manual having been lost or locked away in a cupboard. One ISO, (ISO, Belinda,[5]) who had been involved in the development of the manual reported that, with hindsight, the manual was not sufficiently user-friendly for teachers/TAs to use effectively. She suggested that having on online version of the manual with embedded videos to illustrate suggestions might be a helpful way forward.

Engagement with ISO support and recommendations

The delivery team had some difficulty getting schools to engage at the outset and this was reflected in ISO comments. One ISO did not join the team until January 2022 and was brought in specifically to help with two schools that were geographically located nearby. All four other ISOs spoke of the difficulties of organising visits. Some of this was due to Covid-19 (see context below), but other reasons given included a difficulty getting past gatekeepers who were often unaware of the project, teachers/TAs being able to find a slot in their timetable, and the general pressures that all schools were under.

Once school visits had been set up (either *in situ* or remote) all but one teacher/TA were also complimentary about the ISO support received using terms such as *'helpful'*, *'really, really good'*, *'useful'*, *'professional'*, and *'knowledgeable'*. Teachers/TAs appreciated the timeliness of responses and also how positive and encouraging the ISOs were. Four participants felt that they would have liked more ISO visits; in three cases this was because the visits were valued.

One TA was however a bit more sceptical about the support received noting that: '*it was teaching an old dog new tricks*' (TA, Jane)

In turn, ISOs reported mixed relationships with schools. In some schools they felt appreciated, in others being able to do their role was more difficult, largely because of the challenges that schools were facing (see section 'Context', below).

ISOs reported that hands-on, concrete support was particularly appreciated and the need to keep things as easy for teachers/TAs as possible was noted. This is consistent with the comments noted in respect of the support manual and the training. For this reason, ISOs reported that they felt that the *in situ* visits were most helpful and were scheduled where possible. Most ISOs gave examples of evidence of schools following up on their suggestions and felt that schools appreciated the practical/hands-on advice offered; however, two ISOs also reported that they were not always told what was in fact happening. One talked of '*empty promises*' (ISO, Kim) and another that it was '*quite frustrating with some of the schools as well, because it's almost easier when someone says "no, we're not doing it for two weeks for this reason" than being told that something would happen'* (ISO, Emily)

All ISOs were sympathetic to the pressures that schools were under. For two ISOs, this led to a direct sense of conflict between the competing demands of trying to support schools and make things as easy for them as possible and responsibility from a research point of view that the intervention was delivered as planned, resulting in ISOs not insisting in some cases that certain things happen. The view was taken that it was more productive to support schools with what they *were* doing rather than criticise what they were not.

[5] Pseudonyms have been given for all interview participants.

Three ISOs also noted a tension between themselves and schools over the importance of motivating pupils. Some schools are reported to have been reluctant to adopt reward systems on top of those being used regularly in the classroom. Yet, in the pupil interviews seven out of seven pupils who responded to the 'collecting stars' symbol (part of the motivation system in HER®) said that this was the aspect they liked the most.

Timetabling three sessions per week per pupil (dosage)

Key findings: Dosage was very low.

Although six schools were fully compliant timetabling three HER® sessions per pupil per week, the mean number of actual logins per pupil per week was much lower at 1.4 ranging from 0.21 to 5.7. The number of pupils logging in for a mean of three or more times per week was three although eight other pupils logged in for an average of between two and a half and three times per week.

The mean number of episodes per week achieved across all pupils was 0.8 ranging from 0 to 5.8. The number of pupils achieving three episodes per week overall was three.

Only two pupils completed the programme all the way to episode 100 but only one achieved episode 100 to criterion. Forty-five pupils did not get past episode nine. Figure 10 shows the number of pupils and the episodes achieved to criterion in groups of ten.

Figure 10: Number of pupils and episodes achieved to criterion in groups of ten

Teachers/TAs were very aware that even despite timetabling, and even in the schools that were very well organised, pupils were not taking part in three sessions per week. The reasons for this are discussed in the sections on 'Facilitators and barriers to implementation', 'Time constraints' and 'Context' (below). For three teachers, this led to a feeling of regret—of not having been able to implement HERiSS as they had hoped.

> *I think for me personally….I haven't done the project justice.* (Teacher, Oonagh)

Three ISOs also reported being frustrated that pupils were not being given the opportunity to succeed through lack of scheduling.

Fidelity

Key findings: Fidelity was low.

Fidelity

Fidelity, recorded directly from HER® was low:

Repeating episodes

Episodes were repeated, where appropriate, at 80% fidelity for 23 pupils (30% of applicable cases) and were not repeated, where appropriate, for 86 pupils (70% of applicable cases).

Benchmark assessments

Benchmark assessments were completed, where appropriate, at 80% fidelity for 25 pupils (39%). At between 50 and 80% fidelity benchmark assessments were completed for 17 pupils (27%) and no benchmark assessments were conducted for 22 pupils (34%) despite being required.

Given the low compliance noted above in respect of training, it is possible that the lack of fidelity is related to a lack of schools' understanding what was expected of them. However, a key factor is likely to have been the context within which HERiSS was conducted (see section 'Context', below).

Facilitators and barriers to implementation

Key findings: There were five key interrelated facilitators/barriers to implementation: Pupil response to the intervention and related pupil eligibility; School/individual buy-in including Senior Leadership Team support, and organisation; Ease of use; and the Context within which HERiSS was delivered including Covid-19, logistics, time constraints, competing priorities, and staff turnover including key decision makers.

Pupil response to the intervention

Pupil response was highlighted by ISOs as a significant source of either motivation for, or loss of confidence, in HERiSS for teachers/TAs and this finding was reflected in the teacher/TA interviews. All teachers/TAs spoke of how much their pupils enjoyed HER®, and in particular how engaging they found it, and in turn how motivating this was for them. This was echoed by pupils who were positive about HER®, with 76.64% of symbols across pupils being placed under 'Like'.

Although pupil enjoyment was observed by all participants, one TA spoke of the difficulty of implementing HERiSS when a pupil was less enthusiastic noting in one case that implementation became increasingly difficult as the pupil did not want to be differentiated from his peers. The school tried running the programme with others joining in, but this meant that the programme could not be delivered to him, as outlined by the delivery team, because they were not able to focus on his needs and tailor the pace to suit his learning.

In another instance, peer reaction led to more pupils being added to the programme:

> *I have other pupils in school now that I've had to put on to a Headsprout® account that aren't part of the project but because they've seen other pupils doing it and they all want a taste of it, they all want a go and that's what we want, we want kids wanting to read and wanting to take part*. (TA, Julia)

Pupil eligibility

Pupil response was in turn a function of pupil eligibility.

Pupil eligibility included pupils in Key Stage 1 and/or Key Stage 2 who did not have a reading ability beyond the level of HER® and could:

- sit at a computer for up to ten minutes;
- understand and follow one- or two-step instructions;
- imitate spoken sounds/words;
- respond to feedback (praise or correction); and
- use some self-initiated speech (in English) (single words to short sentences).

Five teachers/TAs reported that, with hindsight, some of the pupils chosen were not suitable, and eight felt that they had other pupils who may have benefited from the programme. Teachers/TAs were asked whether this was because the eligibility criteria were unclear or had not been understood. Only three teachers/TAs felt that the eligibility criteria were a bit too general but most agreed that it was only once they were familiar with the programme, they could pick out the pupils who were most likely to benefit from taking part.

Pupil eligibility also relates to another factor—use of time, outlined below. Two participants questioned whether continuing to work with pupils who had been selected for HERiSS was the best use of their time and their pupils' time.

ISO interviews also identified the importance of pupil eligibility. Three ISOs observed that some pupils were not suitable for HERiSS and that was difficult for staff.

> *I lost one school because they just didn't have the faith in it and I think that it was largely because somehow with the eligibility criteria the wrong students had been selected to do it.* (ISO, Emily)

Issues around pupil eligibility that were cited included (individual responses): confusion on the part of some teachers/TAs as to who had selected pupils, a suggestion that in some cases pupils had been selected by another member of staff who had left (see below), schools saying that they had put some pupils forward but did not get parental consent so made substitutions, other reasons for substitutions such as pupils leaving the school, changing classroom, and therefore location, and two instances where the person doing the eligibility check appeared to not know the pupils very well.

School/individual buy-in including Senior Leadership Team support

Four ISOs spoke about the importance of school buy-in either as a facilitator to engagement or, in its absence as a barrier, and that a lack of buy-in made engagement difficult all the way through. One ISO noted that as early as attendance in the training webinars (see above) it was possible to differentiate between those schools willing to try and those that were more sceptical.

Individual staff buy-in was often a reflection of school buy-in but not always. Three ISOs reported a lack of motivation for HERiSS on the part of some teachers/TAs, one suggesting that it was because of all the issues that schools were facing (see below), another that it was not seen as important, and a third that it was simply seen as '*annoying*' (ISO, Kim)

Factors that may have influenced individual buy-in suggested by ISOs included individual confidence around their IT skills and in their ability to deliver the programme, and fear of the unknown. The importance of the need to become *'familiar'* with the programme was picked up by ISOs: some schools found the initial set up cumbersome, and, in those cases, this delayed the start of engagement.

Senior Leadership Team support

Engagement was reported by ISOs to be particularly good when the Senior Leadership Team was actively involved in the delivery or had appointed someone responsible for coordinating HERiSS (see also sections 'Organisation' and 'Logistics', below).

Six teachers/TAs reported that the support they had from the school Senior Leadership Team was very helpful. Support included providing office space, allocating time, and juggling staffing to facilitate delivery and listening to pupils read as well as broader encouragement. One school used HERiSS as part of their school improvement plan, another demonstrated a HERiSS session during an Office for Standards in Education, Children's Services and Skills (Ofsted) inspection.

Two TAs and one teacher, however, felt that they had not had much support from their Senior Leadership Team but did not provide any examples of the impact that this had had on their experience of the programme. All three participants were not involved in the decision to take part in HERiSS and had had the project passed on to them.

One ISO observed that poor buy-in at the top sometimes led to poor internal communication meaning that those responsible for delivery did not fully understand what was expected of them.

Organisation

Four ISOs singled out good organisation as being key to good engagement on the part of schools. This was reflected in the teacher/TA interviews by those for whom implementation went well:

> *We got into quite a good system to deliver it, but children would get in straight off the bus, they would come down, their iPads would be ready on the table, they'd log themselves in, and we'd almost just float and oversee what was going on"* (Teacher, Fatima)

Good examples of organisation tended to reflect having one person in the school responsible for coordinating HERiSS or having consistent staff involved in delivery.

Ease of use

A factor related to both pupil response and individual buy-in was how easy HER® was to use. Teachers/TAs spoke of the fact that because HER® was IT based, pupils did not see the programme as '*work*' so were enthusiastic to do HER® sessions. Ease of use also meant that not much teacher involvement was required, and it was possible for pupils to use independently.

One teacher also spoke of how easy it was to monitor pupil progress.

For three teachers/TAs, however, HER® was too different to business as usual (see below). For all three this was specifically in relation to the order in which letter sounds were taught and meant that although HER® itself was easy to implement, it gave them extra work running catch-up sessions with pupils for any sounds missed that the rest of the class was working on.

Context

The context in which HERiSS was conducted was challenging and this had a direct impact on implementation. All teachers/TAs spoke of the day-to-day challenges faced by special schools at the best of times. Many pupils have comorbid health conditions. This leads to frequent pupil absences, as well as the need to have trained staff members on hand in the classroom for various medical needs and the difficulties of staffing around this. Staff spoke too of the challenges of teaching pupils with SEND all of whom have different needs and different learning styles. Meeting individual needs often requires one to one or small group work, which poses logistical challenges for resources including the use of space around the school. The lack of interventions that are SEND-specific was noted, meaning that staff are often having to find time to adapt programmes or make their own materials.

Covid-19

A difficult situation in normal times was compounded by Covid-19. All but one teacher/TAs talked about the problems wrought by the pandemic. Ten commented on specific problems due to staffing. A comment by a teacher was typical:

> *Staffing absences however, has been an issue, unfortunately, Headsprout® has been the one thing that got dropped.* (Teacher, Katy)

Another six talked about pupil absences also having an impact, which meant that pupils were not available to take part in scheduled HER® sessions.

Logistics

Five teachers/TA spoke about the logistical difficulties of organising implementation within this context. Problems cited included the difficulties of finding distraction free spaces within the school and organising staffing to access these, having access to space/staff at the right moments of the day that namely, corresponded with individual pupils' timetables, accessing the equipment needed, accessing the internet, and staffing.

Covid-19 posed logistical problems too, which as noted above, was a barrier to implementation and had a direct impact on schools' ability to deliver HERiSS. The need to maintain pupils and staff in bubbles meant that teachers/TAs were limited in the spaces that they could use around the school; early on the in the pandemic the need to sanitise equipment in between pupil use, placed additional time pressures on staff.

Two participants spoke of the lack of consistency and changing routines wrought by Covid-19 leading to increased challenging behaviour among pupils.

These examples cited above point out the practical problems staff faced, but the language used reflects the level of stress that staff were under.

> *We've got COVID again now, we've so many staff off this week it's been horrendous, it's been really hard.* (TA, Denise)

> *Staffing, even now you've still got staffing challenges, if you haven't got any staff you can't do it, and that was quite frustrating.* (TA, Angela)

Issues raised by school staff in the teacher/TA interviews were echoed by ISOs, who talked about staff being *'over stretched'* and under *'stress'*.

Time constraints

Not surprisingly the challenges outlined above led to time pressure—a significant barrier to implementation. The difficulties of fitting the training into the start of the school year have already been noted. This continued into the academic year with four teachers/TAs speaking of the difficulty of fitting HERiSS into daily activities. This was partly because of the pressures facing special schools generally but was also clearly exacerbated by the additional pressures of Covid.

Competing priorities

Time pressure meant that schools had to prioritise activities. For those who were working with pupils where their suitability for the study may have been questionable (see section 'Pupil eligibility') teachers/TAs and ISOs questioned whether spending time on HER® was right for that pupil.

Four teachers/TAs observed that carving out HERiSS time was not always the best use of available resources especially as they were also having to deliver those aspects of the curriculum not covered by HERiSS and were also delivering phonics programmes.

Three ISOs cited competing demands of their existing phonics programme as putting schools under pressure. This was in partly philosophical in that schools could not see how HER® met the government requirement for all SEND pupils to have access to a phonics programme and partly practical as an added time pressure to the challenges noted above.

Staff turnover including key decision makers

The pressures of staff have already been noted. Staffing issues had another significant impact on the delivery of HERiSS. ISOs reported that many schools had a change of staff at the beginning of or part way through the academic year and this sometimes meant that those who had originally signed up to take part in HERiSS or who chose the pupils to take part were no longer there.

> *The new literacy lead had told me that she didn't know anything about what was going on, anything about the project, she didn't even know the pupils who most of the pupils were.* (ISO, Belinda, Line 210)

This will have clearly had an impact on school and staff buy-in noted above. But even where the decision makers may have still been in the school, ISOs noted at least four occasions in which the staff responsible for implementation left and it was difficult to find replacements.

This issue was not directly picked up in teacher/TA interviews although two said that they only got involved because of a change of staff, one noting that it left her '*a little lost'* (TA, Sophie, Line 31) and another that '*it sort of got left to me*' (Teacher, Katy, Line 154), and others noted that choosing pupils to be involved had been done by someone else no longer a part of the team.

Teacher/TA and pupil perceptions of HER®

Key findings: Teachers/TAs and pupils were positive about HER® but acknowledged that it is not right for everyone.

Teachers/TAs and pupils positive about HER®

Notwithstanding the challenges outlined above, teachers/TAs were positive about HER®. All teachers/TAs said that they would be happy to use HER® in the future. Two reported that it had exceeded their expectations while seven said that it met their expectations. Four had no real expectations but commented that they were interested in trying something new (see subsection 'Teacher/TA perceptions of participating in a Randomised Controlled Trial below). One said that she was considering using it for her own son.

Overall, pupils were also positive about HER®, with 76.64% of symbols across pupils being placed under 'Like', 19.18% of symbols placed under 'Not sure', and 10.96% of symbols placed under 'Don't like'. One pupil asked if he could carry on with the programme over the summer.

Pupils enjoyed 'collecting stars' the most followed by 'The teacher helping me', 'Doing things myself', 'The characters', 'Getting things right', 'The games', and 'The pictures'. Six out of eight enjoyed 'Learning words' and 'Learning how to read'.

Six teachers/TAs commented that it was a '*nice programme'*. The interface was praised including the graphics and animations—it was easy to use and for the pupils did not feel like work. They appreciated the comprehensiveness of HER®—namely that, everything was there in the programme, there was no need for further adaptations—a point reflected in the comments about the support manual. One teacher also talked about the way that sounds were put into words and then picked up again in the books. The way that the programme works was highlighted by some participants including its structure, the responsiveness to the pupil, and the repetition, which worked well for some learners. One participant pointed out that HER® was great for pupils who do not have pencil skills who therefore struggle with other interventions, which ask pupils to *'circle'* things.

What teachers/TAs particularly appreciated was the pupil response, notably enjoyment, progress, and an increase in self-confidence in reading.

Pupil enjoyment

Pupil enjoyment has already been discussed above. All teachers/TAs shared examples of this with an enthusiasm that contrasted with the tone and vocabulary used when discussing the pressures that they were under:

> *I've never seen them so motivated to actually want to do the work and read the book and everything.* (TA, Sophie)

> *And oh my gosh the first session he loved it*. (TA, Denise)

Pupil progress

Eleven teachers/TAs talked about the evident progress they saw pupils make.

> *His mum cried because I recorded a video of him reading the book, and when I showed his mum, she cried, she was like, 'I never ever thought my child would read*'. (TA, Geeta)

Pupil self-esteem

Along with progress, six teachers/TAs commented on an increase in pupil confidence and self-esteem and a further teacher talked about the increased independence and enjoyment her pupils got from reading.

> *It almost gave some of them a confidence that kind of said that, 'well, I can read*'. (Teacher, Fatima)

Two teachers talked about their perceived impact on other skills: increased talking/sounding out; increased confidence in asking for help; and one example of improved writing skills.

HER® not right for everyone

Three teachers/TAs were an unqualified *'yes'* about using HER® in the future; others were caveated with the observation that it would be helpful to have as an intervention to work with some pupils. It was clear from responses that they all felt that HER® is a very good resource to have available, but it is not something that can be delivered to everyone.

> *I would happily, you know with the right pupils.* (Teacher, Katy)

This was consistent with attitudes to teaching pupils with SEND to read, which was described as *‘challenging’* but *‘rewarding’.* The challenge noted by all is that SEND pupils are all different and require differentiated ways of working that meet individual needs.

> *We’ve looked into quite a lot of methods because we’ve discovered children do not read the same, and one child does not read the same as another child*. (Teacher, Fatima)

It is also consistent with those who were less convinced about the progress that pupils had made. One TA noted that there was no real progress—in fact their pupils were put back onto phonics catch up at the end of the year and three teachers/TAs reported mixed progress with it working for some of their pupils but not for others. One did not talk about progress either way but noted that she did not think that the pupils that had been selected were suitable for the programme. And one who was supporting one pupil said that the pupil had regressed over the year (nothing to do with HERiSS) but that he had enjoyed the programme.

Teachers/TAs alike demonstrated passion in their commitment to finding something that works: *‘just keep trying’* (TA, Denise); and *‘we don’t give up’* (Teacher, Katy). The reward comes with finding something that works.

There was a concern expressed by three teachers/TAs that the programme made some pupils anxious about getting things wrong. One of these and two others also observed that it would have been helpful if some pupils could have started at a level that better reflected their skills—the earlier episodes were too easy for them. And although there were positive comments about the pace of the programme and the responsiveness to the learner, for some pupils the programme moved too fast with wordy instructions that did not give them time to process what was being asked of them, and for others it was too slow.

Pupils also expressed a concern about *‘getting things wrong’,* with five placing the symbol under ‘Don’t like’.

HER® on its own insufficient

One participant however felt that it could potentially be used with all pupils but that was caveated with the proviso that it was used alongside phonics business as usual. All others spoke of it as part of their resources around literacy—elsewhere the fact that it is not a comprehensive literacy programme has been noted.

Aspects of HER® that were a barrier to implementation

Nearly all teachers/TAs highlighted a concern about the ‘American’ feel to the programme and ISOs reported that schools had made comments to them about this. The accent was the most frequently raised issue but the different terminology for punctuation and the different fonts (pupils with SEND can have difficulty generalising different fonts) were also mentioned. For some the accent was a problem, for others it was something that they feel they come across regularly because of the influence of television.

Notwithstanding the comments about the ease of the interface, teachers/TAs and ISOs reported issues with technology: initial difficulties logging in; the programme freezing occasionally; problems with moving on with episodes 3 and 56; and difficulties printing out the books. One suggested that it would be useful to have a progress bar within each episode as that may have encouraged some pupils to finish an episode within a session.

Business as usual

Key findings: All schools reported using multiple literacy programmes, including phonics, pre- and post-intervention. There was a greater emphasis placed on phonics post-intervention because of the introduction of government guidance in respect of phonics at the start of the academic year 2021/2022. Schools reported a lack of SEND-specific programmes.

Survey

A pre- and post-intervention survey was sent to teachers/TAs in all schools (intervention and control) to establish what business as usual was for literacy before and during the intervention period. Topics included standard literacy activities and the changes over the year in approaches and/or programmes. Pre-intervention, 55 members of staff across the schools completed the teaching-as-usual survey between April 2021 and June 2021. When describing the types of literacy lessons, teachers/TAs spoke about a range of activities, with whole-class, small group, and 1:1 sessions, covering a variety of topics. Thirty-five teachers/TAs (63.6%) mentioned phonics, 12 teachers/TAs (21.8%) mentioned adapting programmes and teaching to effectively support their pupils, and 13 teachers/TAs (23.6%) reported one to one instruction. Most teachers/TAs (n=45, 81.8%) reported using one or more reading programmes, such as Read Write Inc. and Bug Club. These were both whole-class and targeted programmes, with several highlighting the need to adapt these programmes for them to be suitable for their pupils—one stating that the programme they use was *'heavily adapted'*. Very few reading programmes were designed for children with SEND. Exceptions included two programmes specifically for children with Down syndrome, one for dyslexia, and one resource for children who are visually impaired.

Several teachers/TAs spoke about these programmes and approaches being relatively new to the school, with 20 teachers/TAs (36.4%) indicating that the reading programmes had changed since before March 2020, and 21 teachers/TAs (38.2%) reporting a change in the way literacy was delivered. Although most teachers/TAs reported some use of phonics teaching, there was a sense that several schools had been in the process of reviewing or introducing the programmes and approaches, with one teacher stating: *'prior to this there was inconsistency in approaches and schemes used across school'* and another stating: *'we had no phonics programmes in use in the school before lockdown'*. Eight teachers/TAs reported introducing or planning to introduce new programmes in 2021.

Forty-two members of staff completed the teaching-as-usual survey post-intervention between May 2022 and August 2022. When asked about the types of literacy lessons in schools, there was a noticeable increased focus on phonics, across both intervention and control schools, with 32 teachers/TAs (76.2%) mentioning phonics specifically. Teachers/TAs continued to report adapted learning and overall reported an increase in frequency of literacy activities (i.e. move to daily phonics sessions). This appears to closely tie to the DfE guidance on SSP programmes, with some schools introducing new phonics programmes and others moving toward programmes recommended by the DfE, sometimes mid-year. This most commonly included Read Write Inc. (n=8) and Bug Club (n=9), both on the recommended DfE list.

While pre-intervention most schools reported an element of phonics teaching, post-intervention schools reported an increased focus on phonics, whether by standardising their provision, reviewing the programmes used, or introducing new programmes or frameworks. Most teaching staff reported that the way they delivered literacy in school had not changed during the school year (n=28, 66.7%) but just under one-third (n=12, 28.6%) reported that the reading programme used had changed. For example, one member of staff stated: *'we have introduced ELS [essential letters and sounds] based on the SSP phonics release'*. One teacher/TA noted: *'We have steered more towards Twinkl SSP and reading whilst Floppy's phonics was under review'*, and another stated: *'Due to the DfE recommendations that schools use an SSP programme from April 2022, we were forced to change mid-year how we teach reading across school'.* One school noted that their use of a particular phonics programme was linked to funding, stating: *'we have had funding for a phonics scheme and have now committed to using only Read, Write Inc.'*. Several schools spoke about phonics becoming a focus or priority for the school, in part due to DfE recommendations. Finally, one school noted that they had changed the phonics scheme that they used and in addition had also purchased Headsprout® ready for September 2022 to target specific learners. Both pre- and post-intervention, teachers/TAs spoke of adapting teaching to the needs of their pupils and using a variety of programmes and tools within schools for both literacy in general and for phonics specifically.

Data from the pre- and post-intervention surveys do not suggest that control (business as usual) schools accessed HER®. However, the DfE phonics validation process conducted and published in April 2021 meant that several schools, both intervention and control, reviewed or made changes to their phonics provision following the list of validated SSP programmes.

Interviews

Teachers/TAs also discussed business as usual in more in-depth interviews. All teachers/TAs reported being involved in the delivery of phonics programmes as part of their usual literacy activities, which cover reading, writing, comprehension, speaking, and listening skills. Common programmes include Jolly Phonics and Bug Club, but all participants spoke of using many differentiated interventions according to the needs of their pupils. Teachers/TAs were asked about the impact of the government phonics policy changes introduced at the start of the academic year 2022/2023. Most commented that there was minimal impact as their school was already teaching phonics, but five participants said that it had prompted a review of available programmes, and of these, three said that their school had decided to invest in 'Read Write Inc.' phonics.

Teachers/TAs also noted that there is little or no literacy support specifically targeting children with SEND and in particular a few talked about the lack of programmes or materials that motivate children with SEND. For many this was one of the reasons for signing up to the HERiSS project.

Teachers/TAs were asked about how HER® compares with business as usual. One of the obvious things noted was that it is not a comprehensive literacy package and does not include comprehension and writing, which other programmes do. This meant that pupils on the HERiSS programme also needed to participate in those literacy activities, which were not included to keep up with the rest of the class.

The other main difference is that it was entirely computer-based. Teachers/TAs appeared to like this aspect noting the ease of the interface, how easy it was to engage pupils, and the fact that pupils could, for the most part, work independently.

Teachers/TAs also liked the fact that it was comprehensive—and that there was no need for any further adaptations as is often the case with other programmes being used.

> …*and then for us as teachers, we're adapting all the time, we had to do no adaptations, it was there, it was ready, it was accessible, it was, yea you know, it was really easy to deliver, and it was pitched really appropriately for those pupils who tried it, once we got those skills in, once we got like the mouse skills and the attention skills in, it was pitched perfectly for our pupils*. (Teacher, Katy, Line 234)

One teacher pointed out that HER® was great for pupils who do not have pencil skills who therefore struggle with other interventions, which ask pupils to *'circle'* things.

Four teachers/TAs noted however that the order of teaching sounds was different to business as usual. For one this posed a problem vis-à-vis the rest of the class, one felt that the flash cards in the manual were too different to other resources and could potentially confuse pupils, while others just noticed the difference.

Recruitment and retention

School recruitment

There were two school recruitment phases for the HERiSS project:

Recruitment Phase 1: October 2019 – end of March 2020 (pre-pandemic)

Initial advertising of the project was via the EEF Project announcement (i.e. mailing lists and Twitter) and the project's Twitter account (@HERISSProject). Both provided contact details, links to the project webpage (hosted by Bangor University), and the Expression of Interest form.

Eligible special schools were sent information about the project directly and invited to participate, via email in the first instance. This information included a project flyer and a brief project overview document.

Eligible special schools were identified as follows:

- must include Key Stage 1 and/or Key Stage 2 pupils;
- must not be a provision exclusively for severe and/or profound and multiple learning disabilities, physical disabilities and complex medical needs, visual or hearing impairment; and
- must have at least five eligible pupils in the appropriate key stages.

Schools were initially targeted in a defined geographic region (North West and West Midlands) for pragmatic reasons—to facilitate the *in situ* support provided by ISOs. Initially 308 schools were contacted.

If schools expressed an interest, arrangements were made for a follow-up phone call to further explain the project, including the pupil eligibility criteria, project timeline, randomisation, and the process for formally signing up to the project. If schools were still interested in participating, they were sent the pupil eligibility screening form, school information sheet and privacy notice, and the MOU. When the signed MOU was returned to the delivery team, the school entered a 'recruited schools' database, and were informed that the delivery team would be in touch with details regarding pupil data, consent, and the pre-test assessments in due course.

The experience of the delivery team was that success rates of email correspondence with schools was mixed. Most schools contacted could only be done so via the general school office email address, which required the receptionist to forward the email to the headteacher and literacy lead. Greater success was achieved when the delivery team followed up with a phone call—only once this was done were emails forwarded to the right person, although not in all cases. It was common to have to call a school several times to get to speak to the receptionist. This turned out to be a crucial role of the recruitment team. Once the recruitment team was able to speak to the right person directly, the response from schools was positive.

Follow-up calls were made throughout Phase 1 to schools who had not responded, schools who had expressed an interest, and schools who had been sent an MOU but not returned it. It continued to prove very difficult to get through to the headteacher or literacy lead and scheduling a call often took several weeks, and in many cases even then they would end up not being available when the scheduled call was made. Schools in Phase 1 took up to 91 days from initial contact to being sent the MOU. Delays did not appear to reflect lack of interest, as many schools did go on to sign up—it just took time, not surprising perhaps given the context in which special schools operate.

Time delays were also experienced with schools returning completed MOUs. A similarly intensive schedule of chaser emails and calls was concurrently in place for the dozens of schools who had received an MOU, but not yet returned or declined. It took up to 56 days for signed MOUs to be returned, and by the end of Phase 1, 47 schools had returned a signed MOU, but there were still 19 MOUs outstanding.

Recruitment Phase 2: October 2020 – June 2021 (during pandemic)

Based on the lessons learned from recruitment Phase 1, Phase 2 began with a postal mail out to headteachers and literacy leads of all schools who had not signed up in Phase 1. At the same time, an initial email was sent to all schools, which was followed up with a phone call within a few days.

Throughout February 2021 and March 2021, additional schools were contacted to increase the number of recruited schools. These were schools across Leeds and West Yorkshire, and as far as Hull and Lincolnshire. These extensions were facilitated by the decision for the pre- and post-test assessments to be conducted remotely, thus not requiring the evaluation team's assessors to be travelling to schools. The combined total contacted over Phases 1 and 2 was 365 schools.

The difficulties of the time it took for schools to respond was longer in Phase 2 taking up to 200 days from initial contact to being sent the MOU. This was a challenging time for all schools despite the end of the Covid-19 lockdown.

Reflections on recruitment

In addition to directly targeting schools, the delivery team also contacted known networks such as the regional Whole School SEND leads for the North West and the West Midlands on the basis that schools may have seen networks as trusted sources of information. However, with the lack of direct, in-person contact with these networks, there was limited opportunity for directly talking to them about the project and the team noted that there was not a marked change associated with when the information was shared across these networks.

The most effective recruitment strategy was following up initial contacts with schools with telephone calls.

Thoughts provided by the delivery team on future recruitment include:

- Increasing the recruitment period would have the most significant impact on recruitment and implementation:
    - building up specific networks of key stakeholders with formal roles in recruitment process;
    - allowing for internal delays;
    - allowing for direct involvement in determining eligibility of pupils and collating of pupil data; and
    - allowing for earlier randomisation and longer period of implementation planning with schools on an individual basis prior to implementation start dates (in the current project, schools were randomised in the last few weeks of the school year, and early August in some cases, leaving no time for detailed implementation planning).

This would also require greater staffing levels during the recruitment phase.

- Consideration of how schools could be supported with the issue of capacity to commit to delivery—additional funding to directly support appropriate staffing of the intervention should be considered, especially for efficacy trials. It is strongly recommended that this is someone who will remain in the school and continue to bring that expertise to the school following the trial. However, directly allocating funding to supporting the delivery of the intervention would have enabled more schools to participate and would also have supported implementation further.

- Reconsider pupil eligibility—while the eligibility criteria pertaining to pre-requisite skills are important, the exclusion of pupils beyond Key Stage 2 should be reconsidered:

    - many schools who were unable to participate due to insufficient pupil numbers reported would have eligible pupils in Key Stage 3;

    - several schools in Phase 2 were lost due to the fact eligible pupils had moved on to Key Stage 3; and

    - several schools who are participating would have had more eligible pupils if they had been permitted to include pupils in Key Stage 3.

The main reason for not including Key Stage 3 pupils in this trial was due to the increased logistical difficulties envisaged when trying to support delivery with pupils across different key stages, and therefore potentially different sites. However, a number of participating schools did in fact have pupils on different sites despite being in the same key stages, so it is unlikely to be all that different to the current trial in that respect. Additional sites do bring more logistical challenges, and logistical challenges were found to be a key factor in implementation, but once identified, are not insurmountable.

Teacher/TA perceptions of participating in an RCT

Teachers/TAs were unanimous in their enthusiasm about taking part in a piece of research using words like '*exciting'*, '*interesting',* and *'enjoyed'* with many repeating the fact raised when talking about their attitudes to teaching pupils with SEND to read that there is a dearth of research in this area.

> …*because there isn't anything there for special needs. So, to try something it was interesting, it was good, I enjoyed it*. (TA, Denise)

Seven specifically spoke about having been given the opportunity to contribute to evidence-based practice and repeated how *'exciting'* this was as well as a sense that their input would be worthwhile.

Five participants spoke about the benefits of taking part and cited pupil progress but also the fact that participation exposed them to new ideas.

One teacher described using participation as part of their School Improvement Plan and that the monitoring of HER® allowed them to evidence this.

Two teachers said that participation had brought extra work for the school, but no-one reported any drawbacks or regrets from taking part.

Cost

A full economic cost-effectiveness analysis of the intervention has not been included in this study. However, the following cost data were collected to enable schools to decide whether to invest in the programme.

Table 28: Cost of delivering Project A

Item	Type of cost	Cost per school	Total cost over three years	Total cost per pupil per year over three years
Initial teacher/TA training	Time commitment from school staff	Per member of staff: Length of webinars: 2 x 1.5-hours seminars Length of all training videos: 2 hours 45 minutes for all core training videos		
Teacher/TA delivery	Time commitment from school staff	Per school: Average of 2.5 members of staff Three times per week x 20 minutes		
Implementation support	Travel costs for ISO visits	£1,392.00		
	Costs of telephone support	£232.00		
HER® annual license fee	Running cost per school—up to 36 pupils	£195.00	£585.00	£16.25 (assume all 36 pupils)
Printed resources	HER® support manual, all benchmark stories, and master copies of fluency practice sheets x two sets of each	£148.00	£148.00	£4.11 (maintaining above assumption)
	Additional printed stories	£100.00	£300.00	£8.33 (maintaining above assumption)
Total		£2,067.00	£1,033.00	£28.69 (maintaining above assumption)

Conclusions

Table 29: Key conclusions

Key conclusions
1. Children in special schools receiving Headsprout Early Reading® in Special Schools (HERiSS) made no additional progress in reading, on average, compared to children in other special schools. This result has a low security rating
2. There is some evidence of HERiSS having a positive impact on oral reading fluency. However, this finding should be interpreted with caution, as there is considerable uncertainty around the result, with data from 45% of the pupils missing from this analysis
3. There was the equivalent of one month's additional progress found in reading for pupils eligible for free school meals (FSM) who received HERiSS compared to FSM eligible pupils who did not receive the intervention. For these pupils, there was also evidence of a positive impact on phonemic segmentation fluency (the ability to break down a word into its phonetic components). However, due to there being a limited number of FSM pupils in the evaluation, these results are less secure than the headline findings
4. A total of 55 schools and 382 pupils were recruited to this trial, which is one of the largest randomised controlled trials to be funded in special schools in the United Kingdom. Difficulties of implementation in special schools including pressures on staffing, time constraints, meeting individual pupil needs, pupil and staff absence, and logistics were exacerbated by Covid-19, meaning the intervention could not be delivered as intended
5. Factors that facilitated implementation included Senior Leadership Team buy-in and support, pupils' positive response to the intervention, ease of use of the online programme, and good logistical organisation in schools

Impact evaluation and IPE integration

Evidence to support the logic model

Previous studies have indicated that it is possible to teach pupils with SEND early reading skills (Grindle *et al.,* 2013, 2021; Tyler *et al.,* 2015a, 2015b). The difference between these studies and the HERiSS evaluation was that HER® was either delivered or supported by a team of trained researchers rather than teachers/TAs who have other responsibilities within the classroom. The aim of HERiSS was to test the effectiveness of delivering HER® to pupils with SEND by teachers/TAs, supplemented with additional support: training; a support manual; and fortnightly implementation support for schools.

The HERiSS logic model accordingly identifies three change processes:

- Change process 1: Upskilling teaching staff to teach children with SEND to read;
- Change process 2: Adaptation of a reading programme to provide sufficient support for delivering that programme effectively for children with SEND; and
- Change process 3: Teaching children with SEND early reading skills.

These three processes are dependent upon implementation factors on the part of schools, teachers, ISOs, and pupils.

Schools

Schools were required to:

- provide cover for training;
- ensure teacher attends training (support for TA + understanding of intervention);
- enable support visits;
- sufficient information and communications technology (ICT) equipment;

- contingency for staff attrition or illness; and
- recommended staff resource allocated to delivering the intervention (enabling perseverance with challenges).

The IPE analysis highlights the considerable pressures that schools were under—in part due to the normal challenges faced by special schools but these were clearly exacerbated by Covid-19. While schools did facilitate training it is not clear that all who were involved in the delivery of HER® had in fact attended training and there is evidence of changes in staff or staff stepping in to cover absences without having attended the training.

There is evidence too that schools did not always facilitate support visits—in some cases this was because of a school-wide policy due to Covid-19 but there were instances cited by ISOs of difficulties getting past gatekeepers. It was suggested that this was sometimes because those members of staff were not aware of the project.

The provision of ICT equipment does not appear to have been an issue for schools—issues arose mostly around the logistics of finding space and freeing up staff to be able to support pupils at the best times for them.

Providing contingencies for staff attrition or absence was one of the biggest barriers for schools—levels of absence during the pandemic were such that HERiSS was not a priority.

Teachers

Teachers were required to:

- attend and engage with training;
- engage with implementation support;
- engage with training related to specific strategies;
- use the implementation manual to support use of these strategies;
- communicate difficulties to ISOs;
- engage with support from ISOs;
- schedule sessions for three episodes per week;
- listen to pupils read stories within the programme;
- conduct benchmark assessments every ten episodes (including scoring, inputting scores, and making instructional decisions); and
- conduct fluency activities in response to episode, benchmark, or reading data.

Although most teachers/TAs did attend and engage with training it was clear that the timing of the training was not helpful for many, coming as it did at the busiest time of the school year. It was also clear that for some there was confusion in terms of what was expected of them prior to attending the webinars, others had difficulties logging in to HER® before the webinars and both situations meant that teachers/TAs were not able to make the most of them despite agreeing that, overall, the training had been very good.

Evidence of the use of the support manual was mixed—some teachers/TAs reported not needing to use it and ISO's noted occasions when the manual was missing but both groups also provided evidence of using the manual.

Despite being complimentary about ISO support and appreciating the hands-on concrete support, it was clear from the ISO interview data that teachers/TAs did not/could not always follow through with suggestions. In many instances ISOs observed that this was because of the difficulties that staff were facing.

In the end, only six out of 26 schools (23%) were fully compliant, and ten schools (38%) met none of the compliance criteria and although the magnitudes of intervention effect were different across the compliance scenarios (despite not being statistically significant and potentially reflecting the uncertainty in the estimates). Larger effects were observed in the full compliance and partial compliance scenarios with the largest effect seen with partial compliance scenario B (Hedges' g=0.14, 95% CI [-0.08, 0.38]) and smallest effect in the ITT. Partial compliance scenario A showed a smaller difference (Hedges' g=0.02, 95% CI [-0.21, 0.26]) than full compliance (Hedges' g=0.04, 95% CI [-0.18, 0.28]). Note Cohen's d effect sizes were converted to Hedges' g following Borenstein *et al.* (2011). Partial compliance scenario B reflects the most flexible criteria from the three scenarios, so this increased flexibility to compliance likely reflects the larger effect. In all cases, the analyses do not provide evidence to reject the hypothesis that business as usual is the same as Headsprout® plus business as usual, and acknowledging the uncertainty in the estimates reflected in the CIs, which all contain zero.

Teachers/TAs and ISOs acknowledged that the fidelity and dosage activities did not happen (see section 'Pupil' below) citing staff and pupil absence, logistical issues, time pressure, and competing priorities.

Given the findings above it is questionable whether Change process 1: Upskilling teaching staff to teach children with SEND, occurred.

ISOs

ISOs were required to:

- provide high-quality implementation support, including supervision and feedback for school staff;
- closely monitor online data and provide timely feedback and advice to schools;
- provide high-quality implementation support, including support to make effective use of additional strategies in response to implementation challenges; and
- provide high-quality implementation support, including support responding appropriately to intervention data and making effective use of fluency activities.

ISO support was a key component (along with the support manual) of Change process 2: Adaptation of a reading programme to provide sufficient support for delivering that programme effectively for children with SEND. Teacher/TA interviews suggest that the support received was high quality and timely, but it was clear too that the support was not always followed through by schools. Many ISOs acknowledged the frustration of knowing that HER® was not being implemented as planned despite their best efforts and two highlighted a tension in their role between trying to be as supportive and positive to schools as possible and maintaining the rigour of the research process—namely, ensuring that implementation was being done as planned.

It is also questionable, given the above, whether Change process 2: Adaptation of a reading programme to provide sufficient support for delivering that programme effectively for children with SEND, occurred.

Pupils

Pupils were expected to be supported to:

- engage with at least three sessions per week;
- read the stories within the programme;
- complete benchmarks;
- engage in fluency activities when required; and
- engage with any additional strategies needed.

The dosage and fidelity data clearly show that the above activities did not happen. The number of pupils logging in for a mean of three or more times per week was three, with eight other pupils logging in for an average of between two and a half and three times per week. The mean number of logins was 1.4. Even schools that timetabled three sessions a week were often unable to deliver them. Fluency was achieved with fidelity for 23 pupils (30% where appropriate) and benchmark activities for 25 pupils (37% where appropriate).

The dosage was clearly a factor of the challenges highlighted above; fidelity may have been due to a lack of understanding of what was expected of them by teachers/TAs given the low level of compliance.

Only ten pupils reached episode 40 or over. It is not surprising then that in neither the primary or secondary outcomes did we find any evidence of differences between business arms for the reading skills, reading self-concept, and the different components of reading fluency of Key Stage 1 and Key Stage 2 pupils in special schools. In almost all case, we observed very small effect sizes (Hedges' g <0.1; or very small changes in the IRR that were close to 1) and most estimates had considerable uncertainty reflected in wider CIs that contained 0 or 1 (continuous outcome or count outcome, respectively). Change process 3: Teaching children with SEND early reading skills, did not occur.

Interpretation

The successful recruitment of 55 schools and 382 pupils makes this trial one of the largest SEND intervention studies to be funded in special schools in the UK. Informed consent was obtained from pupils' primary care givers and assent was obtained from pupils prior to data collection. These too are significant. HERiSS has shown that it is possible to recruit special schools to an RCT, that the process of randomisation is acceptable to schools, that it is possible to obtain assent from pupils with SEND, and that it is possible to collect data from pupils with SEND online. These are all important achievements given the paucity of research into interventions for pupils with SEND.

Schools were enthusiastic about taking part in a piece of research focusing on SEND pupils welcoming the chance to contribute to an evidence base. Teacher/TA feedback in respect of their participation resonated with discussions about teaching pupils with SEND, which highlighted a lack of differentiated programmes for SEND pupils a significant challenge for teachers.

We did not find evidence against the hypothesis that the Headsprout® intervention and business as usual perform equally on improving the reading skills of Key Stage 1 and Key Stage 2 pupils in special schools as measured by the composite score of the DIBELS® instrument (Hedges' g=-0.01, 95% CI [-0.28,0.26]). The findings suggest that we do not have sufficient evidence to conclude whether the intervention is an improvement over the control (business as usual) as there is considerable uncertainty around our estimate reflected in the CI. The interval contains zero, which would indicate identical means of the trial arm groups cannot be ruled out, so there is also a lack of evidence in the direction of intervention effect.

In most secondary outcomes (reading self-concept and the different components of reading fluency), again we did not find evidence against the hypothesis that the Headsprout® intervention and business as usual perform equally. A possible exception would be ORF (IRR=0.77, 95% CI [0.60, 0.98] SE=0.1) as a reduction in errors made was found when comparing the trial arms. The CI is relatively modest, although the interval does border on the edge of 1, which would indicate no difference in the IRR between trial arms. However, even when adjusting the analyses to account for some complications arising from the data structure, the evidence of a difference between intervention and control (business as usual) was weak. In most case, the observed effects were often very small and had considerable uncertainty reflected in the width of the CIs. The evidence suggests that it is not possible to conclude that the intervention provides any improvement compared to business as usual (control).

To ensure that any findings were not conditional on other variables of interest, we updated the primary analysis to include SEND, FSM, EAL, and whether they moderated the intervention effect (i.e. an interaction between trial arm and any of the additional covariates). The following disability and educational needs indicators appeared to show small negative effects (indicating that the intervention was less likely to be effective in these individuals): Moderate learning disabilities; Autism spectrum disorder; and social, emotional, and mental health needs were; however, this is potentially due to the chance of baseline imbalance of these SEND categories as proportionally more individuals in the intervention arm had these conditions than the control arm (business as usual). Therefore, these findings should be interpreted very cautiously and that no adjustment for multiple comparisons has been included. An additional analysis was conducted removing the SEND variables and their interactions from the model, reducing the number of parameters to estimate alleviating issues with lack of sample size to sensibly estimate parameters. This was motivated by concern for the original analysis by reviewers and the logic of inclusion of SEND classifications in a sample that by definition is defined as having a SEND status being in a special school context.

The level of loss to follow-up was proportionally larger than we anticipated at the outset of the project when designing the study. This was potentially due to the lack of evidence in the context of large-scale interventions in special schools. There was a noticeably larger amount of attrition in the intervention arm schools (as noted above one school cited the capacity to commit to the intervention post-randomisation, two schools did not officially withdraw but cited *'staffing crisis'* and *'conflict with current literacy programme'* as reasons respectively for not participating and two schools never engaged with either the delivery team or the evaluation team). When investigating the missingness and testing the consistency of the findings after allowing for missingness, we found that on average across the entire sample of individuals, scores in the composite measure increased. Reasoning for this discrepancy between the primary and the missing data analysis results relates to the composite scoring and the volume of zeros in the components scores used to calculate the composite. The missing data analysis imputes missing scores based on all other known information, so that a more complete data set is available to analyse. When the imputation is used, scores are moved away from zero in the individual measures (zero is given if no score recorded, or missing, in the composite), but when imputed, the composite score increases substantially (please note that only missing values are imputed and not genuine zero values). It must be noted that both ORF measures had around 45% missing data, so imputation was challenging for these variables and is likely the biggest contributor to the change. Overall, the missing data analyses and the primary analysis results did not differ as a lack of evidence remained against the hypothesis that the intervention and the control (business as usual) were equal.

When considering subgroup analyses for FSM eligibility, again most analyses for both primary and secondary outcomes were non-significant. PSF was one exception that showed within the FSM subgroup, showing some evidence that a difference was present between intervention and control (business as usual), correspondingly when FSM was a moderator in the primary analysis. These were small but marginally significant effects and is somewhat consistent with previous evidence from Grindle *et al.* (2021). However, if corrections for multiple comparisons are made, this effect is likely non-significant and unlikely provides evidence for a difference between trial arms.

In general, most analyses found that accounting for the classroom and school clustering showed only marginal explanation of information in the model, which reflects the context of the special schools (within school heterogeneity is likely quite high) and the small number of individuals in each cluster.

Notwithstanding the difficulties outlined above and the evidence to suggest differences between intervention and control (business as usual) across measures, the experience of staff and pupils using HER® was positive. Teachers/TAs noted how much pupils enjoyed the programme, the progress that they made, and with that an increase in confidence. Teachers appreciated the ease of use and the comprehensiveness of the programme, which meant that they were not required to make adaptations. All staff who spoke about their experiences reported that they would like to use HER® in the future but were keen to stress that while HER® works for some pupils it is not suitable for all. This was consistent with teachers/TAs' attitudes to teaching children with SEND to read—that every learner is different and what works for one may not work for another. It is not surprising therefore that pupil eligibility was a significant barrier to implementation—many schools reported that with hindsight they had not put the right pupils on the programme and that they had others whom they felt would have benefited.

Teachers/TAs also noted that HER® is not a comprehensive literacy programme and would never therefore replace their current literacy provision but that it would fit into and sit alongside their existing resources well. All schools reported being involved in the delivery of several literacy programmes including phonics as business as usual.

Limitations and lessons learned

Conducting research with pupils with SEND is challenging. This was the first such large-scale RCT and there are lessons to be learned both for the delivery and evaluation teams.

Attrition

Baseline assessments were conducted during May 2021 – July 2021 (Summer Term 2021). The intervention did not start until September 2021, in some cases five months after baseline data were collected.

Potential confounding variables include:

- attrition between baseline data collection and the start of the intervention with some pupils not returning to school; and
- changing staff/changing roles at schools impacted an understanding of what schools expected:

- engaging with schools and identification of an appropriate contact person was a problem throughout, from recruitment, to implementation, and data collection stages. This was not a lack of will on the part of schools—simply a reflection of the time and staffing pressures that they are under (highlighted throughout this report).

The trial had a relatively high rate of attrition. One school with ten pupils withdrew post-randomisation citing an inability to commit because of staffing/resources. Post-testing seven schools were lost (60 pupil). Three schools simply did not respond, others cited problems such as not being able to fit the session in before the end of term, a conflict with the school's current literacy programme, a staffing crisis, and post-tests being too difficult for schools. Even within the schools that did engage in post-testing, pupils were lost for various reasons although no pupils were officially withdrawn from the study: moving school; long-term absence; schools deciding that pupils were not eligible (Key Stage 4); and pupil selectively mute/non-verbal. In addition, there were many instances of missing data where no scores were recorded largely because of a difficulty of engaging pupils on the day of testing.

Contamination or other concurrent interventions

The MOU did not explicitly state that schools in the intervention group were not to offer pupils' using HER® other reading interventions. Had implementation worked as planned this may have been a significant limitation.

The teaching-as-usual survey was designed to capture any changes to business as usual. What is clear is that schools use several different interventions with pupils, many of which cover all aspects of literacy and others focusing on phonics. It is highly unlikely that HER® was the only exposure to reading for any pupil.

Data collection

Baseline and post-intervention data were collected with pupils supported by teachers/TAs. This was to help pupils feel as comfortable as possible and would have been the case even if assessments had been conducted *in situ*. It was even more important with the assessments being conducted online because teachers/TAs were able to sort out any technical issues as well as inform the research assistant if the pupil was beginning to become distressed. There were recorded instances where research assistants felt that teachers/TAs had prompted pupils. Where possible the assessments were paused and the member of staff reminded that it was important to know what the pupils' independent responses would be, and the assessment was repeated with a second set of materials. However, this was only in those instances where teacher/TA assistance was obvious to the research assistant and others may have been missed.

It is possible that sharing more information with schools about the nature of the assessments may have helped. The information shared was primarily aimed at reassuring teachers/TAs that pupils would not be put in a stressful situation and that they had the right to stop at any point where the pupils appeared to be uncomfortable. It did not share information about the purpose of the assessment or that, as standardised assessments, they may seem a bit unusual. During the assessments a number of teachers/TAs questioned the assessments and felt that they did not reflect the pupils' true ability, and this may have influenced the support given to pupils.

Capturing implementation fidelity

Data regarding fidelity and dosage come primarily from the HER® software. Other data regarding compliance came from ISO reports. All ISO data collection was cross-checked by the delivery team. The evaluation team were unable to independently verify the data from ISO reports. The triangulation process used to cross reference data sources however, was designed to mitigate against this limitation.

Measurement of outcomes

Data for primary and secondary outcomes were collected by a team of research assistants recruited for the project. All research assistants received training in data collection methods, engaging with and motivating pupils with SEND, and safeguarding. Research assistants sent in practice videos to Dr Flynn and Dr Denne, which were assessed prior to the start of data collection. Research assistants were not assigned to pupils until they met the required fidelity level. Fidelity checks were also taken mid-point through the data collection (at baseline and post-intervention) to check against procedural drift. All research assistants were asked to add notes to each data collection session indicating whether there were specific difficulties with the data collection session—IT issues, pupils not appearing to be engaged, and difficulties understanding pupil responses. Data collection sheets were all crossed-checked by the evaluation team.

Despite these measures in place, it is possible that there will have been variation in the accuracy of data collection. These analyses were not pre-specified in the SAP and hence, were not conducted retrospectively. This may be a useful addition in future studies to pre-specify a sensitivity analysis to investigate this issue.

Furthermore, the adapted version of the DIBELS® that was used for online delivery is not validated. It was not within the scope of this evaluation to validate the measure. This should be considered in any future research.

Selective reporting and data availability

All schools in the intervention group were invited to take part in the teacher/TAs interviews. Responses were all from those schools that had engaged in the study despite attempts to invite schools that had not engaged. The experiences and perceptions of teachers/TAs taking part is therefore skewed in favour of those who had taken part at least to some extent. Including interviews with ISOs was an attempt to mitigate against this.

Reflections

Given the issues encountered in delivering this intervention, future studies may benefit from greater allocation of resources to support delivery/evaluation. The special school environment has many additional challenges compared to mainstream schools and the potential for greater heterogeneity among individuals within schools and a higher propensity for loss to follow-up, should be a high priority for design of future studies. Accordingly, interventions may need more extensive piloting within the special schools' environment and more tightly controlled selection of participants via extended inclusion/exclusion criteria.

References

Allor, J.H., Mathes, P.G., Roberts, J.K., Cheatham, J.P. and Champlin, T.M. (2010) 'Comprehensive Reading Instruction for Students With Intellectual Disabilities: Findings From the First Three Years of a Longitudinal Study'. *Psychology in the Schools,* 47(5), 445–466. https://doi.org/10.1002/pits.20482

Altman, D.G., and Bland, J.M. (2005) 'Treatment Allocation by Minimisation'. *BMJ,* 330(7495), 843. https://doi.org/10.1136/bmj.330.7495.843

Angrist, J.D. and Imbens, G.W. (1995) 'Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity'. *Journal of the American Statistical Association*, 90(430), 431–42. https://doi.org/10.1080/01621459.1995.10476535.

Austin, P.C., White, I.R., Lee, D.S., and van Buuren, S. (2021) 'Missing Data in Clinical Research: A Tutorial on Multiple Imputation'. *Canadian Journal of Cardiology,* 37(9), 1322–31. https://doi.org/10.1016/j.cjca.2020.11.010

Barnett, A.G. and Dobson, A.J. (2008) *'An Introduction to Generalized Linear Models* (3rd ed.)'. New York: Chapman and Hall/CRC

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015) 'Fitting Linear Mixed-Effects Models Using lme4'. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01

BERA (British Educational Research Association). (2018) *'Ethical Guidelines for Educational Research'.* London, UK: BERA. Available at: https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018 (accessed 28 September 2023).

Bland, J.M. (2017) 'Minim: Allocation by Minimisation in Clinical Trials'. Available at: https://www-users.york.ac.uk/~mb55/guide/minim.htm (accessed 26 September 2023).

Borenstein, M., Hedges, L.V., Higgins, J.P. and Rothstein, H.R. (2011) '*Introduction to Meta-Analysis'*. West Sussex, UK: John Wiley & Sons.

Bradford, S., Shippen, M.E., Alberto, P., Houchins, D.E. and Flores, M. (2006) 'Using Systematic Instruction to Teach Decoding Skills to Middle School Students With Moderate Intellectual Disabilities'. *Education and Training in Developmental Disabilities*, 41(4), 333–43. http://psycnet.apa.org/record/2006-21512-002

Bradshaw, J., Gore, N. and Darvell, C. (2018) 'Supporting the Direct Involvement of Students With Disabilities in Functional Assessment Through Use of Talking Mats®'. *Tizard Learning Disability Review*, 23(2), 111–16. https://doi.org/10.1108/TLDR-01-2018-0004

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Maechler, M. and Bolker, B.M. (2017) 'glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling'. *The R Journal*, 9(2), 378–400. doi:10.32614/RJ-2017-066

Browder, D.M., Wakeman, S.Y., Spooner, F., Ahlgrim-Delzell, L. and Algozzine, B. (2006) 'Research on Reading Instruction for Individuals With Significant Cognitive Disabilities'. *Exceptional Children*, 72, 392–408.

Chapman, J.W. and Tunmer, W.E. (1995) 'Development of Young Children's Reading Self-Concepts: An Examination of Emerging Subcomponents and Their Relationship With Reading Achievement'. *Journal of Educational Psychology*, 87(1), 154–67. https://doi.org/10.1037/0022-0663.87.1.154

Coyne, M.D., Kame'enui, E.J., Simmons, D.C. and Harn, B.A. (2004) Beginning Reading Intervention as Inoculation or Insulin: First-Grade Reading Performance of Strong Responders to Kindergarten Intervention. *Journal of Learning Disabilities,* 37(2) 90–104. http://dx.doi.org/10.1177/00222194040370020101

Data Protection Act 2018, c. 12. Available at: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted (Accessed: 30 Nov 2022)

Demack, S. (2019) Does the classroom level matter in the design of educational trials? A theoretical & empirical review. EEF Research Paper 003, May 2019. https://educationendowmentfoundation.org.uk/projectsand-evaluation/evaluating-projects/eef-research-papers/ Department for Education (DfE). (2018) '*Special Educational Needs in England: January 2018'.* London: DfE. Available at: https://www.gov.uk/government/statistics/special-educational-needs-in-england-january-2018 (accessed 26 September 2023).

Dong, N. and Maynard, R. (2013) '*PowerUp!:* A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies'. *Journal of Research on Educational Effectiveness*, 6(1), 24–67. https://doi.org/10.1080/19345747.2012.673143

Education Endowment Foundation (EEF). (2021) '*Phonics'.* Available at: https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/*phonics* (accessed 26 September 2023).

Education Endowment Foundation (EEF). (2018) '*Statistical Analysis Guidance for EEF Evaluations'.* London: EEF. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/EEF_statistical_analysis_guidance_2018.pdf?v=1630566986 (accessed 26 September 2023).

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C.A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R. and Wissel, S. (2016) *'Foundational Skills to Support Reading for Understanding in Kindergarten Through 3rd Grade'.* (NCEE 2016-4008). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Available at: https://ies.ed.gov/ncee/wwc/PracticeGuide/21 (accessed 26 September 2023).

Fox, J., Kleiber, C. and Zeileis, A. (2021) '*ivreg: Instrumental-Variables Regression by '2SLS', '2SM', or '2SMM', with Diagnostics. R Package Version 0.6-1'*. Available at: https://CRAN.R-project.org/package=ivreg (accessed 26 September 2023).

Gelman, A. and Hill, J. (2006) '*Data Analysis Using Regression and Multilevel/Hierarchical Models (Analytical Methods for Social Research*)'. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Germain, R. (2004) 'An Exploratory Study Using Cameras and Talking Mats to Access the Views of Young People With Learning Disabilities on Their Out-of-School Activities'. *British Journal of Learning Disabilities,* 32(4), 170–74. https://doi.org/10.1111/j.1468-3156.2004.00317.x

Good, R.H., Kaminski, R.A., Shinn, M., Bratten, J., Shinn, M., Laimon, D., Smith, S. and Flindt, N. (2004) '*Technical Adequacy of DIBELS: Results of the Early Childhood Research Institute on Measuring Growth and Development (Technical Report, No. 7)'.* Eugene, OR: University of Oregon.

Gough, P.B. and Tunmer, W.E. (1986) Decoding, Reading, and Reading Disability. *RASE: Remedial and Special Education*, 7(1), 6–10. https://psycnet.apa.org/doi/10.1177/074193258600700104

Grindle, C.F., Carl Hughes, J., Saville, M., Huxley, K. and Hastings, R.P. (2013) 'Teaching Early Reading Skills to Children With Autism Using Mimiosprout Early Reading'. *Behavioural Interventi*ons, 28(3), 203–24. 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., Jackson Brown, F. and 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(4), 334-344. https://doi.org/10.1111/1471-3802.12531

Hayes, R.J. and Moulton, L.H. (2017) '*Cluster Randomised Trials* (2nd ed.)'. New York, NY: Chapman and Hall/CRC. https://doi.org/10.4324/9781315370286

Hedges, L.V. (2007) 'Effect Sizes in Cluster-Randomized Designs'. *Journal of Educational and Behavioural Statistics,* 32(4), 341–370. https://doi.org/10.3102/1076998606298043

Hedges, L.V. (2011). 'Effect Sizes in Three-Level Cluster-Randomized Experiments'. *Journal of Educational and Behavioural Statistics*, 36(3), 346–380. http://www.jstor.org/stable/29789487

Hsieh, H.F. and Shannon, S.E. (2005) 'Three Approaches to Qualitative Content Analysis'. *Qualitative Health Research,* 15(9), 1277–88. https://doi.org/10.1177/1049732305276687

Huffstetter, M., King, J.R., Onwuegbuzie, A.J., Schneider, J.J and 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(4), 279–98. https://doi.org/10.1080/10824669.2010.532415

Jiang, Y. and Small, D. (2014) '*ivpack: Instrumental Variable Estimation'*. R Package Version 1.2. Available at: https://CRAN.R-project.org/package=ivpack (accessed 26 September 2023).

Joseph, L.M. and Seery, M.E. (2004) 'Where is the Phonics?: A review of the Literature on the Use of Phonetic Analysis With Students With Mental Retardation'. *Remedial and Special Education*, 25(2), 88–94. https://doi.org/10.1177/07419325040250020301

Kahan, B.C. and Morris, T.P. (2012) 'Improper Analysis of Trials Randomised Using Stratified Blocks or Minimisation'. *Statistics in Medicine*, 31(4), 328–40. **https://doi.org/10.1002/sim.4431**

Kasim, A., Xiao, Z., Higgins, S., & Troyer, E. D. (2017). eefanalytics: Analysing Education

Trials [Computer software manual]. (R package version 1.0.6).

Katims, D.S. (2000) 'Literacy Instruction for People With Mental Retardation: Historical Highlights and Contemporary Analysis'. *Education and Training in Mental Retardation and Developmental Disabilities*, 35(1), 3–15.

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

Kuznetsova, A., Brockhoff, P.B. and Christensen, R.H.B. (2017) 'lmerTest Package: Tests in Linear Mixed Effects Models'. *Journal of Statistical Software*, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Layng, T.J., Twyman, J.S. and Stikeleather, G. (2003) 'Headsprout Early Reading™: Reliably Teaching Children to Read[1]'. *Behavioural Technology Today*, 3, 7–20. https://behavior.org/resources/191.pdf

Legislation.gov.uk. (2018) *'Data Protection Act 2018 c. 12'.* Available at: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted (accessed 28 September 2023).

Marchand-Martella, N.E., Slocum, T.A. and Martella, R.C. (2004) '*Introduction to Direct Instruction'*. London, UK: Pearson Education Inc.

Moerbeek, M. and Schie, S.V. (2019) 'What Are the Statistical Implications of Treatment Non-Compliance in Cluster Randomized Trials: A Simulation Study'. *Statistics in Medicine*, *38*(26), 5071–84. https://doi.org/10.1002/sim.8351

Morena, L. (2011) '*Evaluating the Impact of Headsprout on the Reading Achievement of English Language Learners'*. Doctoral dissertation. Athens, GA: University of Georgia.

Murphy, J. Gray, C.M., van Achterberg, T., Wyke, S. and Cox, S. (2010) 'The Effectiveness of the Talking Mats Framework in Helping People With Dementia to Express Their Views on Well-Being'. *Dementia*, 9(4), 454–72. https://doi.org/10.1177/1471301210381776

Parkin, Kennedy, Long, Hubble and Powell, (2018). 'Support for People With a Learning Disability'. *Briefing paper Number: 07058 House of Commons.* Available at: **https://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN07058** (accessed 26 September 2023).

Plumpton, C.O., Morris, T., Hughes, D.A. and White, I.R. (2016) 'Multiple Imputation of Multiple Multi-Item Scales When a Full Imputation Model is Infeasible'. *BMC Research Notes*, 9, 45. https://doi.org/10.1186/s13104-016-1853-5

Qualtrics© Available at: https://www.qualtrics.com

Raudenbush, S.W. and Bryk, A.S. (2002) *'Hierarchical Linear Models. Applications and Data Analysis Methods (2nd ed.)'.* Thousand Oaks, CA: Sage Publications.

Reichow, B., Lemons, C.J., Maggin, D.M. and Hill, D.R. (2019) 'Beginning Reading Interventions for Children and Adolescents With Intellectual Disability'. *Cochrane Database of Systematic Reviews*, 12, CD011359. https://doi.org/10.1002/14651858.CD011359.pub2

Roberts-Tyler, E.J., Hughes, J.C. and Hastings, R.P. (2019) 'Evaluating a Computer-Based Reading Programme With Children With Intellectual Disabilities: Feasibility and Pilot Research'. *Journal of Research in Special Educational Needs*, 20(1), 14–26. https://doi.org/10.1111/1471-3802.12458

Rose, J. and Department for Children, Schools and Families (DCSF). (2009) '*Identifying and Teaching Children and Young People With Dyslexia and Literacy Difficulties'.* Available at: http://dera.ioe.ac.uk/14790/7/00659-2009DOM-EN_Redacted.pdf

Rubin, D.B. (1996) 'Multiple Imputation After 18+ Years'. *Journal of the American Statistical Association,* 91(434), 473–89. https://doi.org/10.2307/2291635

Schieffer C., Marchand-Martella N.E., Martella R.C., Simonsen F.L. and Waldron-Soler K.M. (2002) 'An Analysis of the Reading Mastery Program: Effective Components and Research Review. *Journal of Direct Instruction*, 2(2), 87–119.

Sermier Dessemontet, R., Moser Opitz, E. and Schnepel, S. (2020) 'The Profiles and Patterns of Progress in Numerical Skills of Elementary School Students With Mild and Moderate Intellectual Disability'. *International Journal of Disability, Development and Education*, 67(4), 409–23. https://doi.org/10.1080/1034912X.2019.1608915

Sermier Dessemontet, R., Martinet, C., de Chambrier, A.-F., Martini-Willemin, B.-M. and Audrin, C. (2019) 'A Meta-Analysis on the Effectiveness of Phonics Instruction for Teaching Decoding Skills to Students With Intellectual Disability'. Educational Research Review, 26, 52–70. https://doi.org/10.1016/j.edurev.2019.01.001

Storey, C., McDowell, C. and Leslie, J.C. (2020) 'Headsprout Early Reading for Specific Literacy Difficulty: A Comparison Study'. *Journal of Behavioural Education,* 29(3), 619–33. https://doi.org/10.1007/s10864-019-09336-7

Talking Mats Limited. (2014) '*Effectiveness Framework of Functional Communication'*. Available at **https://www.talkingmats.com/wp-content/uploads/2014/09/EFFC-2014.pdf** (accessed 26 September 2023).

Tonkin-Crine, S., Anthierens, S., Hood, K. Yardley, L., Cals, J.W., Francis, N.A., Coenen, S., van der Velden, A.W., Godycki-Cwirko, M., Llor, C., Butler, C.C., Verheij, T.J., Goossens, H. and Little, P.; GRACE INTRO/CHAMP consortium. (2015) 'Discrepancies Between Qualitative and Quantitative Evaluation of Randomised Controlled Trial Results: Achieving Clarity Through Mixed Methods Triangulation'. *Implementation Science,* 11(66), 1–8. https://doi.org/10.1186/s13012-016-0436-0

Tjur, T. (2009) 'Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination'. *The American Statistician,* 63(4), 366–72. https://www.jstor.org/stable/25652317

Twyman, J.S., Layng, T.V.J. and 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'. *Behavioural Technology Today*, 6, 1–19.

Tyler, E.J., Hughes, J.C., Beverley, M. and Hastings, P. (2015a) 'Improving Early Reading Skills for Beginning Readers Using an Online Programme as Supplementary Instruction'. *European Journal of Psychology of Education*, 30(3), 281–94. https://doi.org/10.1007/s10212-014-0240

Tyler, E.J., Hughes, J.C., Wilson, M.M., Beverley, M., Hastings, R.P. and Williams, B.M. (2015b) '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

van Buuren, S. (2018) *‘Flexible Imputation of Missing Data, Second Edition’.* Boca Raton, FL: Chapman and Hall/CRC. https://doi.org/10.1201/9780429492259

Wickham, H., Averick, M., Bryan, J., Chang, W., D’Agostino McGowan, L., François, R., Grolemund G., Hayes, A., Henry, L., Hester, J. Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. and Yutani, H. (2019) ‘Welcome to the Tidyverse’. *The Journal of Open Source Software*, 4(43), 1686, https://doi.org/10.21105/joss.01686

Appendix A: EEF cost rating

Figure A1: Cost rating

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

Appendix B: Security classification of trial findings

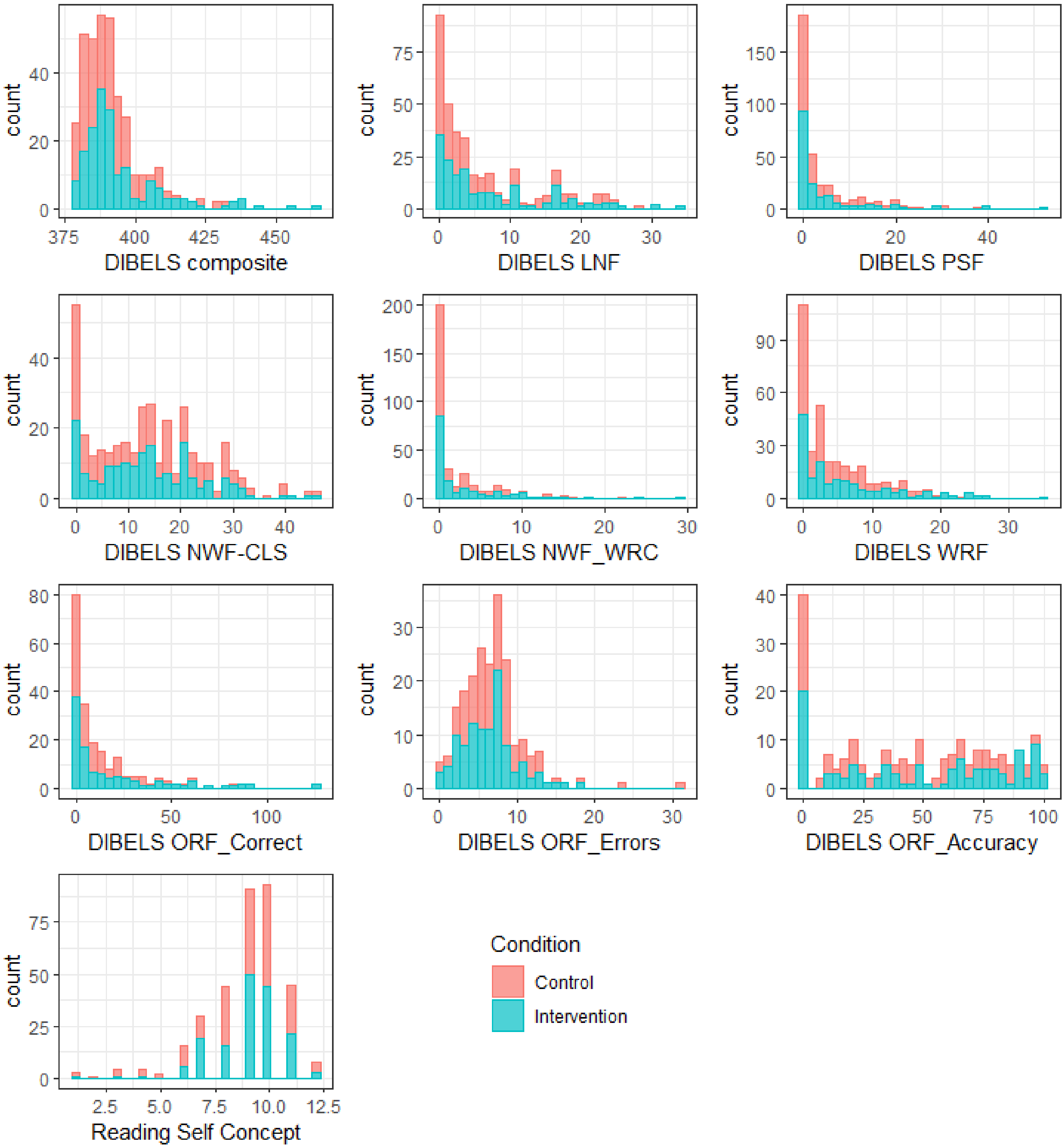
Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	MDES	Attrition					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%					
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%	3		Adjustment for threats to internal validity [X]		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					1
0	No comparator	>=0.6	>50%					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Moderate	Clustered randomised design, baseline imbalance in primary and secondary outcome (ES>0.3)
Threat 2: Concurrent Interventions	Low/Moderate	From the interview data there were schools using many differentiated interventions or other HERS programme although earlier on in the report, the authors did discuss how most were not SEND specific.
Threat 3: Experimental effects	Low/moderate	Business as usual survey didn't suggest control schools accessing HER®. Schools did however review or make changes to phonics provision following a list of validated SSP programmes.
Threat 4: Implementation fidelity	High	Compliance was poor with low dosage and fidelity.
Threat 5: Missing Data	Moderate/High	High levels of attrition for the primary outcome (27%) and 45% for secondary outcomes – some evidence of differential attrition.
Threat 6: Measurement of Outcomes	Low	Measures were blind to treatment allocation. These were delivered online and there hasn't been much discussion whether any validation work has been carried out on the alternative form of delivery.
Threat 7: Selective reporting	Low	Evaluation team provided substantial detail on a wide range of analyses.

- **Initial padlock score:** 3
- **Reason for adjustment for threats to validity**: [-2] Padlocks – Substantial attrition (evidence of differential attrition), poor implementation fidelity.
- **Final padlock score:** initial score adjusted for threats to validity = [1] Padlock due to substantial differential missing data and implementation fidelity

Appendix C: Histograms of baseline variables

Figure C1: Histograms of the baseline variables distributions



Appendix D: Reading Self-Concept Scale

Adapted Reading Self-Concept Scale

Researcher Score Sheet

Instructions:

I'm going to read some things to you. They are all about reading. I want you to tell me if they are what you think about reading (thumbs up) or not what you think about reading (thumbs down).

Item questions

Tick the box that corresponds with the participant's response.

Question	Yes	No
I need help to read * (D)		
I can read words (C)		
Reading is easy (D)		
Reading makes me happy (A)		
I get words wrong when I read* (C)		
I don't like reading (A)*		

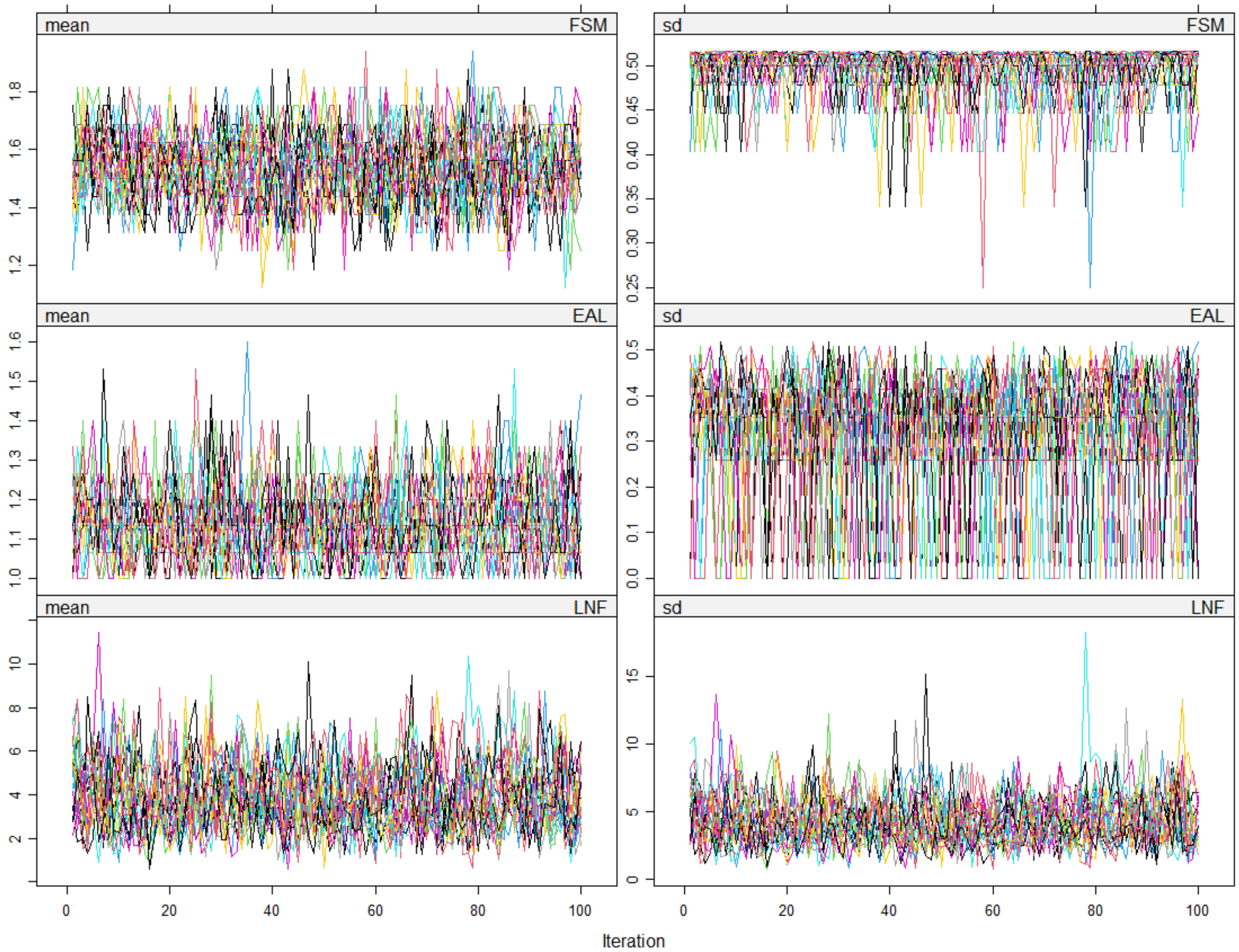
*Note that items are reverse scored.

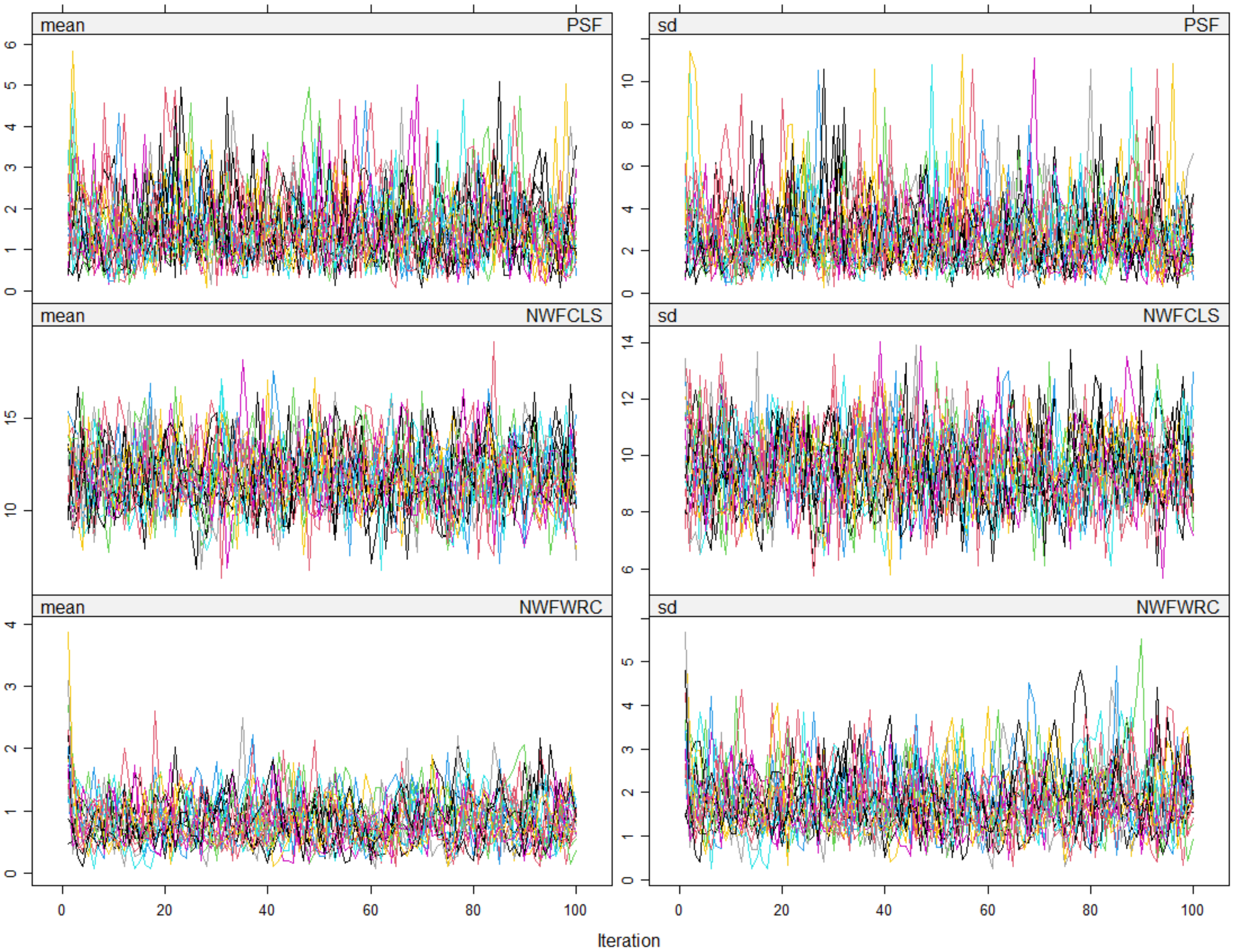
A – attitude towards reading

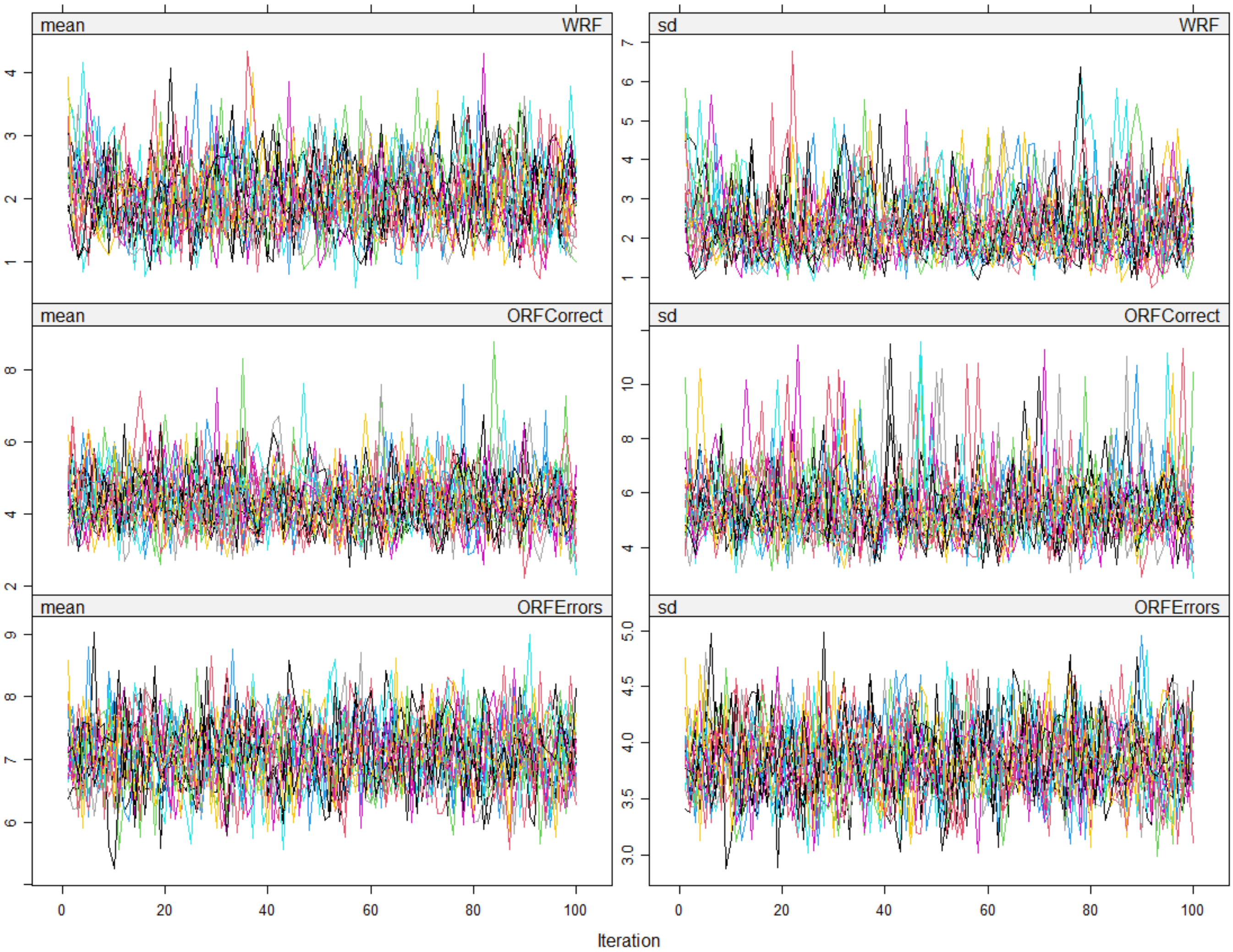
D – perception of level of difficulty

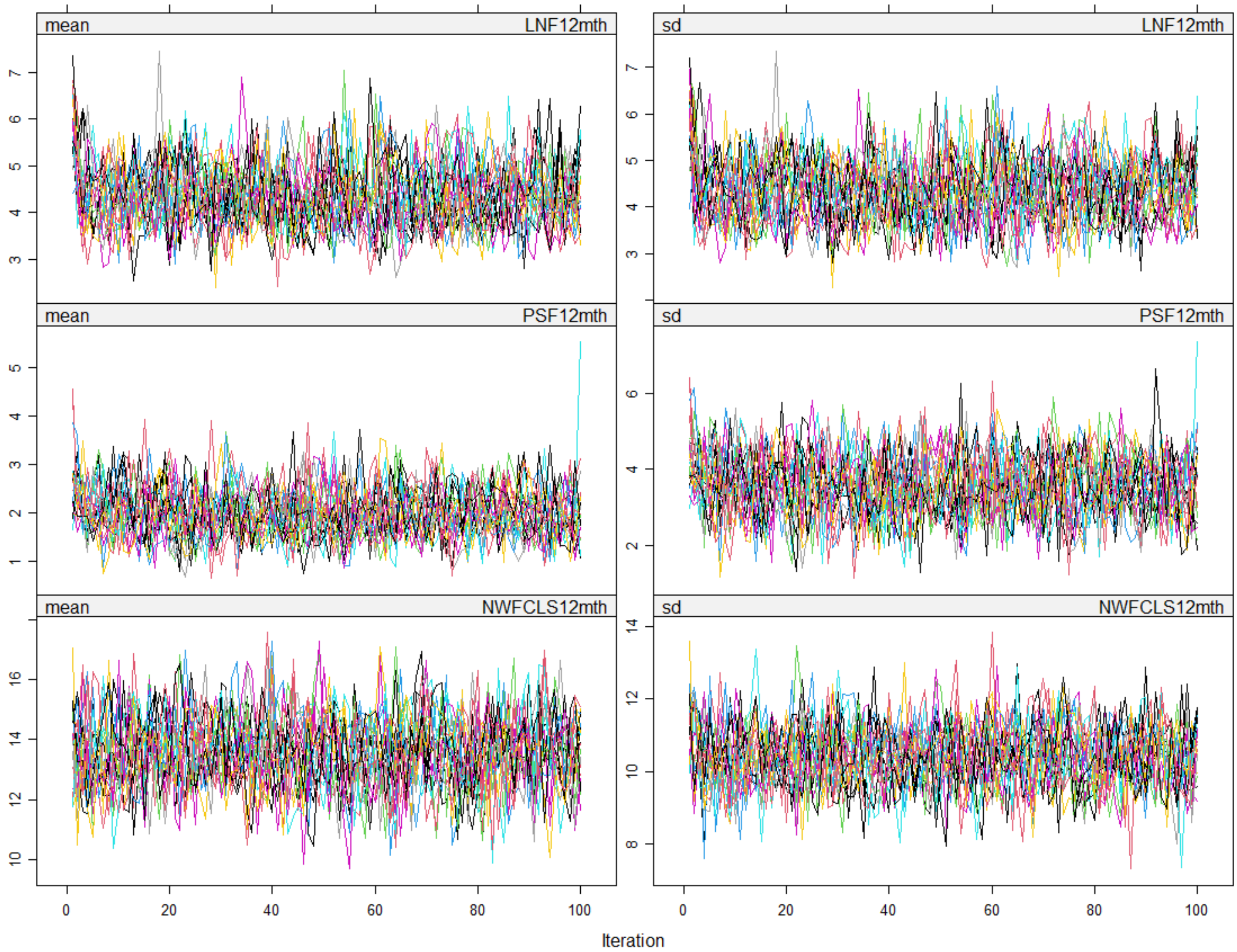
C – perception of competence

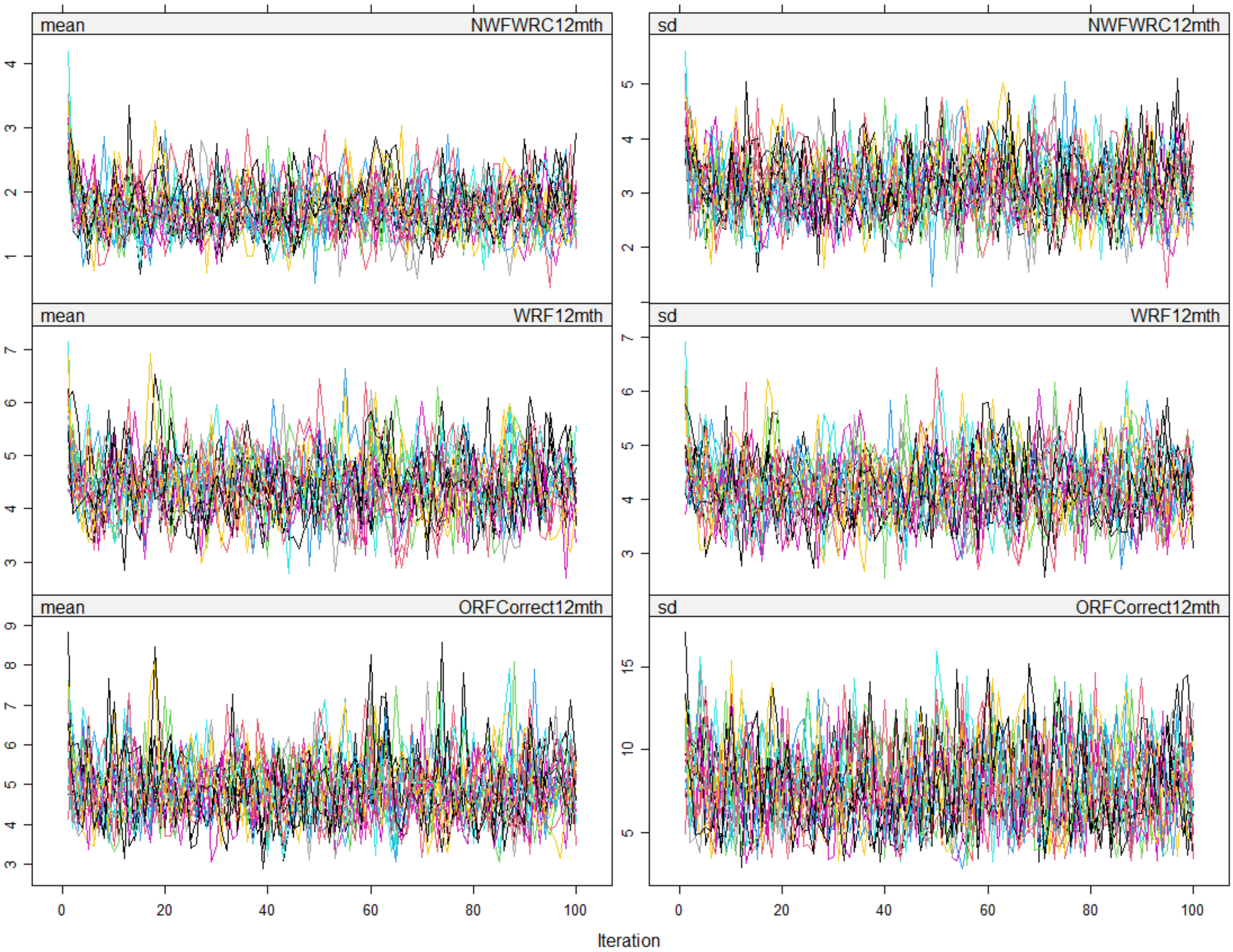
Appendix E: Trace plots for multiple imputation convergence

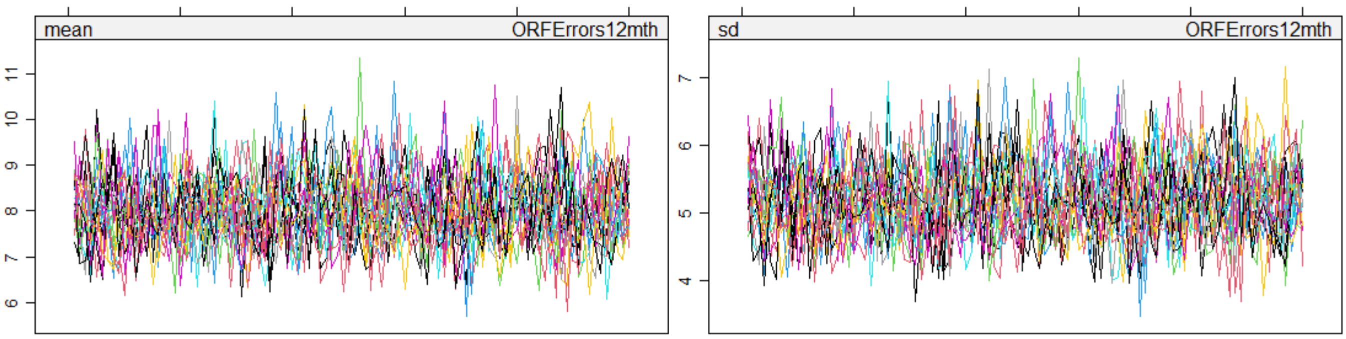



mean
PSF
sd
PSF
mean
NWFCLS
sd
NWFCLS
mean
NWFWRC
sd
NWFWRC
Iteration


mean
WRF
sd
WRF
mean
ORFCorrect
sd
ORFCorrect
mean
ORFErrors
sd
ORFErrors
Iteration


mean
LNF12mth
sd
LNF12mth
mean
PSF12mth
sd
PSF12mth
mean
NWFCLS12mth
sd
NWFCLS12mth
0
20
40
60
80
100
Iteration


mean
NWFWRC12mth
sd
NWFWRC12mth
mean
WRF12mth
sd
WRF12mth
mean
ORFCorrect12mth
sd
ORFCorrect12mth
Iteration


mean
ORFErrors12mth
sd
ORFErrors12mth

Appendix F: Parents' information sheet and consent form

PARENT/GUARDIAN INFORMATION SHEET

Version 1.4, 28/01/2021

Title of Project: Headsprout® Early Reading in Special Schools (HERiSS)

Delivery team (Bangor University): Dr Emily Roberts-Tyler, Professor Carl Hughes, Dr Corinna Grindle (independent consultant), Dr Claire McDowell (Ulster University)

Evaluation team (University of Warwick): Dr Samantha Flynn, Dr Louise Denne, Professor Richard Hastings, Dr Rebecca Morris, & Dr Tom Bailey

If you need help to read this information sheet, please contact a member of the project team using the contact details at the end of this information sheet.

Introduction

Headsprout® Early Reading is a computer-based, targeted reading intervention designed to teach beginning readers the skills and strategies necessary for efficient, fluent reading, taking beginning readers to a mid Y3 reading age. The delivery team at Bangor University have already done some research testing Headsprout® in a small number of special schools, and it is now ready to be tested with a large number of special schools across England to evaluate how effective it is in helping pupils with special educational needs and disabilities (SEND) to learn to read.

Your child's school is being invited to take part in this larger research study and they have chosen your child to be included in the study. Before you decide if you are happy for your child to be involved in the research study, you need to understand why the research is being done and what it would involve for your child's school. Please take the time to read the following information carefully.

Please ask us if there is anything that is not clear, or if you would like more information. Take time to decide whether you wish to take part.

Who can take part?

We will be recruiting 110 special schools across England for this study. Special schools in England can take part in this study if they have pupils in key stage 1 and/or 2 and are able to allocate sufficient staff to deliver the programme.

What is Headsprout®?

Headsprout® is a computer-based programme that focuses on building fluency in essential early reading skills (such as decoding and blending) by giving children lots of opportunities to practice, until the skill becomes fluent. Pupils work their way through online lessons ('episodes'), which are supplemented by additional 1:1 fluency activities as required. Although Headsprout® was originally designed for typically developing children, a team from Bangor University have piloted the programme in special schools with some encouraging results.

What does taking part in the evaluation involve for my child?

Whether schools will receive Headsprout® or be in the control group (and not receive Headsprout®) will be decided at random (like tossing a coin). Half of the schools will receive Headsprout®, and the other half will be in the control group. This is so that we can compare the reading abilities of pupils in the study who received Headsprout® to those who did not. It is the best way to show whether there are any differences between the two groups.

The Headsprout® schools will be required to follow the programme with the selected pupils (as described above). The control group schools will continue to teach as normal.

Before your child can be part of the evaluation, we will need your consent for them to take part. The form is attached to this information sheet and needs to be completed and returned to the school to let us know if you are happy for your child to take part.

If you are happy for your child to take part, the evaluation team at the University of Warwick will complete reading tests with your child on two occasions (between May and July in 2021, and again in 2022) in all schools, regardless of whether they received Headsprout® or not. This is to see if there are any differences between pupils who do Headsprout® and those who don't. These tests will be completed by trained testers and will be conducted online using

Microsoft Teams and with the help of school staff. They are short but we are allowing 45 minutes per pupil, to include settling in time and breaks.

We will also collect information about whether English is their first language, whether they are eligible to receive Free School Meals, and their SEND.

What does Headsprout® involve?
Children work through activities in an online programme which adapts instruction in response to their answers. Headsprout® involves three computer-delivered lessons weekly, with stories and benchmark assessments at various stages. Activities are designed to be engaging and resemble computer games. There are 80 computer-delivered lessons and, depending on individual children and their needs, sessions typically take between 10 and 30 minutes. Stories are available throughout the programme, and can be read within the online episodes as well as outside of the Headsprout® sessions (these are available as printable books). Benchmark assessments take place after every 10 episodes, and take the form of a story. School staff listen to each child read the story individually and rate the reading. These ratings and other monitoring activities are used to decide if the child needs to work on the additional 1:1 fluency activities.

Are there any other parts to the study?
We will also ask some pupils from Headsprout® schools to have an interview with a researcher to tell us more about their experiences of Headsprout®. The interviews with pupils will be undertaken using Talking Mats® to ensure that children with different levels of communication abilities are able to share their experiences with the evaluation team. Again these interviews will be conducted online using Microsoft Teams. We will send you further information about the interviews at the time, so that you can decide whether or not you would want your child to take part in this additional part of the research.

How will information about my child be handled?
All information about your child will be handled in confidence and only members of the delivery and evaluation teams will have access to identifiable data. Study data stored at Bangor University and the University of Warwick will be kept separate from personal information (e.g., names). Study data will be kept securely for 10 years in line with the University of Warwick's policies.

Research data will be pseudonymised as quickly as possible after data collection. This means all direct and indirect identifiers will be removed from the research data and will be replaced with a participant number. The key to identification will be stored separately and securely to the research data to safeguard your identity.

At the end of the trial, the University of Warwick will share the data with EEF's data archive processor through secure data portals, where it will be encrypted and saved to secure servers. At that point EEF will become the data controller and the University of Warwick will no longer have any responsibility for it. For further information, please read the attached privacy notice. You may also, should you wish, refer to the University of Warwick Research Privacy Notice which is available here:

https://warwick.ac.uk/services/idc/dataprotection/privacynotices/researchprivacynotice or by contacting the Information and Legal and Compliance Team at GDPR@warwick.ac.uk.

Does my child have to take part in this study?
No, you are free to choose whether or not you want your child to take part in the study. If you do not want your child to take part then the evaluation team at the University of Warwick will not complete reading tests with your child. You can ask questions about the study using the contact details below.

Your child can stop taking part in the study at any time, without giving a reason, and without affecting you/them in any way. If you, or your child, decide to withdraw from the study, you will be able to choose to have their data removed as well, provided that this is before data analysis takes place (scheduled for July 2022). If the study is stopped for any other reason, we will inform you.

What will happen to the results of this study?
At the end of the evaluation, a report of the research results will be completed and sent to the Education Endowment Foundation who are paying for the study. Results will be published in scientific journals and presented at scientific meetings. Your child will not be identified in any report, publication or presentation. Once the research study is complete, we will send you a summary of the results via your child's school.

What are the potential benefits?
Schools in both groups (Headsprout® and control) will be contributing to the evidence base about teaching children with SEND to read using an online programme, and this may help more children with SEND in the future.

If your child is in the Headsprout® group, then they will be able to use the programme during their time at school.

What is the timetable for the study?
The evaluation team from the University of Warwick will contact schools in early May 2021 to set up online sessions to carry out some reading tests with the pupils. After these tests have been completed, schools will be randomly chosen to either receive Headsprout® or be in the control group.

For schools who are receiving Headsprout®, they will start using Headsprout® with pupils from September 2021 for 8 months.

The evaluation team from the University of Warwick will contact all schools in the study (both Headsprout® and control schools) to carry out some more online reading tests with pupils in late May to early July 2022.

What organisations are involved in this study?
The delivery of Headsprout® is being led by researchers at Bangor University, and the research evaluation is being carried out by the University of Warwick. The study is being funded by the Education Endowment Foundation.

Who has reviewed the study?
This study has been reviewed and given favourable opinion by the University of Warwick's Humanities and Social Science Research Ethics Committee (HSSREC): **REF: HSSREC 37/19-20**

Contact details
To speak to a member of the delivery team about the Headsprout® programme, please contact Emily Roberts-Tyler using the contact details below.

To speak to a member of the evaluation team about the reading tests and interviews, please contact Louise Denne or Samantha Flynn using the contact details below.

Project Team contact details

Emily Roberts-Tyler	Email: **e.j.tyler@bangor.ac.uk** Telephone: 01248 383962
Louise Denne	Email: **l.denne@warwick.ac.uk**; Telephone: 024 7652 3638
Samantha Flynn	Email: **s.flynn@warwick.ac.uk**; Telephone: 07823 362152

Who should I contact if I wish to make a complaint?

Any complaint about the way you have been dealt with during the study or any possible harm you might have suffered will be addressed. Please address your complaint to the person below, who is a senior University of Warwick official entirely independent of this study:

Head of Research Governance
Research & Impact Services
University House
University of Warwick
Coventry
CV4 8UW
Email: **researchgovernance@warwick.ac.uk**
Tel: 02476 575733

If you wish to raise a complaint on how we have handled your personal data, you can contact our Data Protection Officer who will investigate the matter: **DPO@warwick.ac.uk.**

If you are not satisfied with our response or believe we are processing your personal data in a way that is not lawful you can complain to the Information Commissioner's Office (ICO).

PARENT/GUARDIAN CONSENT FORM

Title of Project: Headsprout® Early Reading in Special Schools (HERiSS)

Delivery team (Bangor University): Dr Emily Roberts-Tyler, Professor Carl Hughes, Dr Corinna Grindle (independent consultant), Dr Claire McDowell (Ulster University)

Evaluation team (University of Warwick): Dr Samantha Flynn, Dr Louise Denne, Professor Richard Hastings, Dr Rebecca Morris, & Dr Tom Bailey

Would you like your child to take part in our reading study?

If No:	If Yes:
1. Please complete the red box over the page. 2. Please return the form to school.	1. Please complete the green box below. 2. Please return the form to school.

	Please initial box
I have read and understood the information sheet for this study. I have been given the opportunity to ask any questions I want to ask.	
I understand that my child's participation is voluntary and that I am free to withdraw my consent before July 2022 without giving any reason.	
I understand that it will not be possible to identify any personal data from this study in publications (e.g., journal articles, conference presentations). I agree to my child's pseudo-anonymised data being used in this way.	
I understand that my child's school may or may not be selected to receive the Headsprout® Early Reading programme and that this is to be decided at random by the research team.	
I understand that my child will take part in reading tests delivered online at their school on two occasions, once in May-July 2021 and again in May-July 2022.	
I understand that the pseudo-anonymised data collected about my child will include whether English is their first language, whether they are eligible to receive Free School Meals, and their SEND will be collected	
I understand that my child's data in relation to this study will be securely stored for 10 years, in line with the University of Warwick's policies.	
I **agree** that my child can take part in the above study.	

Name of child: ______________________________

______________________ ____________ ____________________

Name of Parent Date Signature of parent

	Please initial box
I do **not** want my child to take part in the above study	

Name of child: ____

______________ ________ __________________
Name of Parent Date Signature of parent

Appendix G: School information sheet, privacy notice, and Memorandum of understanding

SCHOOL INFORMATION SHEET

Version 1.4, 28/11/2021

Title of Project: Headsprout® Early Reading in Special Schools (HERiSS)

Delivery team (Bangor University): Dr Emily Roberts-Tyler, Professor Carl Hughes, Dr Corinna Grindle (independent consultant), Dr Claire McDowell (Ulster University)

Evaluation team (University of Warwick): Dr Samantha Flynn, Dr Louise Denne, Professor Richard Hastings, Dr Rebecca Morris, & Dr Tom Bailey

Introduction

Headsprout® Early Reading is a computer-based, targeted reading intervention designed to teach beginning readers the skills and strategies necessary for efficient, fluent reading, taking beginning readers to a mid Y3 reading age. The delivery team at Bangor University have already done some research testing Headsprout® in a small number of special schools, and it is now ready to be tested with a large number of special schools across England to evaluate how effective it is in helping pupils with special educational needs and disabilities (SEND) to learn to read.

Your school is being invited to take part in this larger research study. Before you decide, you need to understand why the research is being done and what it would involve for your school. Please take the time to read the following information carefully.

Please ask us if there is anything that is not clear, or if you would like more information. Take time to decide whether you wish to take part.

Who can take part in the study?

We will be recruiting 110 special schools in England for this study. Special schools in the West Midlands or North West of England can take part in this study if they have pupils in key stage 1 and/or 2, and are able to allocate two teachers or Teaching Assistants (TAs) to support delivery of the programme at least 3 times per week.

The school must have at least five pupils in key stage 1 or 2 who meet the eligibility criteria for the study. Schools will be provided with a brief task to help identify eligible pupils. However, the following list provides an outline of the eligibility criteria. Pupils must:

1. be able to sit at a computer for a period of time (approximately 10 minutes)
2. understand and follow one or two-step instructions (e.g., 'clap your hands and turn around')
3. imitate spoken sounds or words (e.g., 'say eeee', and pupil says 'eeee')
4. respond to feedback (praise or correction)
5. use some self-initiated speech (including single words and short sentences).
6. have minimal reading skills.

What is Headsprout®?

Headsprout® is a computer-based programme that focuses on building fluency in essential early reading skills (such as decoding and blending) by giving children lots of opportunities to practice, until the skill becomes fluent. Pupils work their way through online lessons ('episodes'), which are supplemented by additional 1:1 fluency activities as required. Although Headsprout® was originally designed for typically developing children, a team from Bangor University have piloted the programme in special schools with some encouraging results.

What does Headsprout® involve?

Pupils work through activities in an online programme which adapts instruction in response to their answers. Headsprout® involves three computer-delivered lessons weekly, with stories and benchmark assessments at various stages. Activities are designed to be engaging and resemble computer games. There are 80 computer-delivered lessons and, depending on individual children and their needs, sessions typically take between 10 and 30 minutes. Stories are available throughout the programme, and can be read within the online episodes as well as outside of the Headsprout® sessions (these are available as printable books). Benchmark assessments take place after every 10 episodes, and take the form of a story. School staff listen to each child read the story individually and rate the reading.

These ratings and other monitoring activities are used to decide if the child needs to work on the additional 1:1 fluency activities.

The study involves implementing Headsprout® with between 5 and 15 pupils. Based on a combination of small group episode delivery and 1:1 work as required, implementation of Headsprout® with the maximum 15 pupils would require between 10 and 20 hours per week. The time required would typically be at the lower end of this range, however, some flexibility to respond with 1:1 input as required is important for high quality implementation. It is therefore necessary for two staff members to be allocated to Headsprout® delivery for this study, to enable reliable implementation each week.

What does taking part in the study involve?

Whether schools will receive Headsprout® or be in the control group (and not receive Headsprout®) will be decided at random (like tossing a coin). Half of the schools will receive Headsprout®, and the other half will be in the control group. This is so that we can compare the reading abilities of pupils in the study who received Headsprout® to those who did not. It is the best way to show whether there are any differences between the two groups.

The Headsprout® schools will be required to follow the programme with the selected pupils (as described above). The control group schools will continue to teach as normal.

The evaluation team at the University of Warwick will complete reading tests with pupils on two occasions (between April and July in 2021, and again in May-July 2022) in all schools, regardless of whether they received Headsprout® or not. This is to see if there are any differences between pupils who do Headsprout® and those who don't. These tests will be completed by trained testers remotely (using Microsoft Teams) and should take around 45 minutes per pupil, including settling in time and breaks. The evaluation team will arrange the best times to complete the testing with each school.

We will also ask some teachers and pupils from Headsprout® schools to have an interview with a researcher to tell us more about their experiences of taking part in the research, and Headsprout®. The interviews with pupils will be undertaken remotely with Microsoft Teams using Talking Mats® (a picture based question response system) to ensure that children with different levels of communication abilities are able to share their experiences with the evaluation team. Further information about the interviews will be provided to teachers and pupils at the time to help them to make the decision about taking part.

How will the information about my school and the pupils be handled?

All information about your school and pupils will be handled in confidence and only members of the delivery and evaluation teams will have access to identifiable data. Study data stored at Bangor University and the University of Warwick will be kept separate from personal information (e.g., names). Study data will be kept securely for 10 years in line with the University of Warwick's policies.

Research data will be pseudonymised as quickly as possible after data collection. This means all direct and indirect identifiers will be removed from the research data and will be replaced with a participant number. The key to identification will be stored separately and securely to the research data to safeguard pupil and teacher identity.

At the end of the trial, the University of Warwick will share the data with EEF's data archive processor through secure data portals, where it will be encrypted and saved to secure servers. At that point EEF will become the data controller and the University of Warwick will no longer have any responsibility for it. For further information, please refer to the University of Warwick Research Privacy Notice which is available here: **https://warwick.ac.uk/services/idc/dataprotection/privacynotices/researchprivacynotice** or by contacting the Information and Legal and Compliance Team at **GDPR@warwick.ac.uk**.

Does my school have to take part in this study?

No, you are free to decide whether or not you want your school to take part in this study using the information in this information sheet. You can also ask us questions about it, using the contact information below.

Your school and/or pupils can withdraw from the study at any time, without giving a reason, and without affecting you/them in any way. If you do decide to withdraw from the study, you will be able to elect to have your data removed as well, provided that this is before data analysis takes place (scheduled for July 2022). If the study is stopped for any other reason, we will inform you.

What will happen to the results of the study?

At the end of the evaluation, a report of the research results will be completed and sent to the Education Endowment Foundation who are paying for the study. Results will be published in scientific journals and presented at scientific meetings. Your school and pupils will not be identified in any report, publication or presentation. Once the research study is complete, we will send you a summary of the results.

What are the potential benefits for my school?
If your school is randomly chosen to receive the Headsprout® Early Reading programme, then you will be provided with pupil licences for all children who are taking part in the study.

If your school is randomly chosen to be in the control group (and not to receive Headsprout®), your school will instead receive two payments totalling £1000, £250 on completion of pre-test assessments, and £750 on completion of post-test assessments.

Schools in both groups (Headsprout® and control) will be contributing to the evidence base about teaching children with SEND to read using an online programme, and this may help more children with SEND in the future.

Does it cost anything to take part?
If your school is randomly chosen to receive the Headsprout® Early Reading programme, then you will be required to allocate three members of staff (including the staff who will be delivering the intervention, and at least one teacher) to attend a training session for 1 day in September 2021 to learn how to support pupils to do the programme. They will then go on to support the pupils in your school to do Headsprout®, with a manual and implementation support provided by the delivery team throughout the study. This will take the form of monthly school visits, with phone supervision and email support provided in between these visits. Schools chosen to receive Headsprout® will also make a contribution to the training and implementation support costs of £200.

If your school is randomly chosen to be in the control group (and not to receive Headsprout®), there will be no costs to your school.

What is the timetable for the study?
Schools will be recruited between December 2019 and April 2021. Schools will then decide which pupils will take part in the study (based on the eligibility criteria above). Parents of these pupils will be asked to read some information about the study and consent to their child taking part by early April 2021.

The evaluation team from the University of Warwick will remotely carry out some reading tests with the pupils in recruited schools starting in mid-April 2021. After these tests have been completed, schools will be randomly chosen to either receive Headsprout® or be in the control group; all schools will be informed of their allocation by early July 2021.

For schools who are receiving Headsprout®, there will be training for TAs and teachers in early September 2021, and they will start using Headsprout® with pupils from September 2021 for 8 months.

The evaluation team from the University of Warwick will carry out some more remote reading tests with pupils in all schools in the study (both Headsprout® and control schools) in late May to early July 2022.

Who is involved in this study?
The delivery of Headsprout® is being led by researchers at Bangor University, and the research evaluation is being carried out University of Warwick. The study is being funded by the Education Endowment Foundation.

Who has reviewed the study?
This study has been reviewed and given favourable opinion by the University of Warwick's Humanities and Social Science Research Ethics Committee (HSSREC): HSSREC 37/19-20

How do I enrol my school in this trial?
Places in this trial are limited and will be allocated on a 'first come first serve' basis. If you are willing and able to participate in this trial, please contact Emily Roberts-Tyler using the contact details below.

Contact details
For further information about Headsprout® and the planned delivery within schools, please contact Emily Roberts-Tyler using the contact details below.

For further information about the evaluation (i.e., the reading tests and interviews), please contact Louise Denne or Samantha Flynn using the contact details below.

Project Team contact details

Emily Roberts-Tyler	Email: **e.j.tyler@bangor.ac.uk** Telephone: 01248 383962
Louise Denne	Email: **l.denne@warwick.ac.uk**; Telephone: 024 7652 3638
Samantha Flynn	Email: **S.Flynn.1@warwick.ac.uk**; Telephone: 07823 362152

Who should I contact if I wish to make a complaint?

Any complaint about the way you have been dealt with during the study or any possible harm you might have suffered will be addressed. Please address your complaint to the person below, who is a senior University of Warwick official entirely independent of this study:

Head of Research Governance
Research & Impact Services
University House
University of Warwick
Coventry
CV4 8UW
Email: **researchgovernance@warwick.ac.uk**
Tel: 02476 575733

If you wish to raise a complaint on how we have handled your personal data, you can contact our Data Protection Officer who will investigate the matter: **DPO@warwick.ac.uk.**

If you are not satisfied with our response or believe we are processing your personal data in a way that is not lawful you can complain to the Information Commissioner's Office (ICO).

Headsprout® Early Reading® in Special Schools (HERiSS)

Privacy notice for parents and carers

1. Why are we collecting this data?

Your child's school is participating in a research project testing the impact of an early reading programme called 'Headsprout® Early Reading'.

The University of Warwick is evaluating the Headsprout® programme. The evaluation will be conducted as a randomised controlled trial; this means that of all schools that apply to take part in the trial, half will be randomly selected to receive the programme, and half will continue with their normal practice (like tossing a coin). The University of Warwick will be collecting data about pupil's reading skills and confidence from all schools that take part in the trial. They will also be interviewing pupils who did Headsprout® about their thoughts and feelings about it.

The Education Endowment Foundation (EEF) has commissioned and funded the HERiSS project. The University of Warwick and Bangor University (joint data controllers; responsible for keeping the data safe and using it appropriately) will decide what happens to children's data as part of the evaluation. When the trial is completed and the data is submitted to the archive, the University of Warwick and Bangor University will delete all children's data, and the Education Endowment Foundation (EEF) will become the data controller.

2. What personal data is being collected for this project?

Bangor University will collect the following personal data about each child, either from the school or the National Pupil Database (NPD):

- The child's name, school year, Unique Pupil Number (UPN), gender, whether English is an additional language, special educational needs and/or disabilities (SEND) information, and eligibility for Free School Meals. This data will be kept up to date during the project.

The University of Warwick will collect the following data about each child, directly from them via an on-line platform, Microsoft Teams:

- Reading skills test scores (at the start and end of the project)
- How they feel about reading (at the start and end of the project)
- For children who used Headsprout®: How they felt about using it.

Bangor University will collect the following data about each child using Headsprout®, from the Headsprout® programme:

- How far pupils got with the programme, and how they used it.

3. What is the legal basis for processing this data?

We will only use your personal data where we have a lawful basis for doing so. For this project, this will be your informed and written consent for your child to participate in this project. We will not collect any data about your child unless you agree that it is ok for us to do so.

4. Who will personal data be shared with?

Bangor University will collect some data, as described in section 2, and share this with the University of Warwick. They will then destroy that data, so that only the research team at the University of Warwick will see your child's data during the evaluation.

At the end of the trial, the University of Warwick will share the data with EEF's data archive processor through secure data portals, where it will be encrypted and saved to secure servers. At that point EEF will become the data controller and the University of Warwick will no longer have any responsibility for it.

5. Is personal data being shared outside of the UK?

Your child's personal information will not be transferred outside of the UK.

6. How long will personal data be retained?

Bangor University will delete any data within one month of securely transferring it to the University of Warwick.

The University of Warwick will share all the data with EEF's data archive processor within three months of the trial completion. The University of Warwick will keep all data until July 2032 so that we can finish writing up the results and do further analysis where appropriate

Data is not kept longer than is necessary and is deleted in accordance with University of Warwick's internal policy.

7. Can I stop my child's personal data being used?

The University of Warwick handles personal data in accordance with the rights given to individuals under data protection legislation. If at any time you wish us to withdraw your child's data or correct errors in it, please contact **HERiSS@warwick.ac.uk**

In certain circumstances, data subjects have the right to restrict or object to processing. They also have the right to make a subject access request to see all the information held about them. The University of Warwick will cooperate fully when a subject access request (SAR) is made. To exercise these rights, please contact our Compliance Team (infocompliance@warwick.ac.uk).

8. Who can I contact about this project?

For any queries contact **HERiSS@warwick.ac.uk**

If you have a concern about the way this project processes personal data, we request that you raise your concern with the University of Warwick in the first instance. Alternatively, you can contact the Information Commissioner's Office, the body responsible for enforcing data protection legislation in the UK, at https://ico.org.uk/concerns/.

9. Last updated

We may need to update this privacy notice from time to time, so we recommend that you revisit this information every now and then. This version was last updated on 28th January 2021.

Memorandum of Understanding in relation to participation in the Headsprout® Early Reading in Special Schools (HERiSS)

*Please sign both copies, retaining one for your own records and returning the second copy to Emily Roberts-Tyler via email (**e.j.tyler@bangor.ac.uk**) or post (School of Education, Bangor University, Eifionydd, Normal Site, Bangor, LL57 2PZ)*

School Name: ____

School Address: ____

This Memorandum of Understanding (MOU) sets out the roles and responsibilities of schools participating in, and the parties involved in delivering and evaluating, the Headsprout® Early Reading in Special Schools (HERiSS).

This document is being sent to your school because you have indicated interest in participating in the project.

1. *The project teams and their roles*

This MOU will refer to the Headsprout® Early Reading programme as the "**Intervention**".

The Education Endowment Foundation (EEF) are the funders of the project.

Bangor University (the delivery team) will be responsible for school recruitment, training teachers and TAs in schools in the Headsprout® group to use the intervention and providing supervision to these teachers and TAs.

The University of Warwick (the evaluation team) will independently evaluate the intervention's impact on pupil outcomes (namely, their reading ability and reading self-concept). They will also explore teachers' and pupils' experiences of the intervention.

2. *Communication with parents*

All participating schools will distribute an information sheet (containing a link to an online privacy notice) and consent form to parents of between 5 and 15 pupils in KS1 and KS2 who meet the study eligibility criteria (provided by the delivery team). The information sheet and consent form for parents/guardians will be provided by the delivery team and shared with schools shortly. Schools have the option of distributing either a hard or electronic copy or providing parents a link to an internet based version depending on their preferred method of contracting parents.

The information sheet and linked privacy notice will inform parents of the nature of the project, the personal data that will be collected about their child and how this data will be processed. It will give them the opportunity to decide whether or not to consent to their child taking part in the project.

Schools should allow parents an initial two weeks to return the consent form, and should follow-up with them if they have not done so by this point to encourage them to seek further information from the delivery or evaluation teams, should they need it.

Parents/carers will be able to withdraw their child from the project at any stage even if they have provided consent by contacting the evaluation team at **HERiSS@warwick.ac.uk**

Parents/carers will be provided with full details on their rights under data protection laws and contact details for the delivery and evaluation teams in the information sheet (which the delivery team will provide to schools).

3. *The evaluation*

The evaluation of the programme is being conducted by the University of Warwick (the "**evaluation team**"). The project involves the evaluation of the Headsprout® Early Reading programme through a randomised controlled trial (RCT), along with an implementation and process evaluation. Schools will select between 5 and 15 pupils who meet the study eligibility criteria to be involved in the project. Recruited schools will be randomly chosen to either receive or not receive the intervention (a 50:50 split). Using randomly assigned groups is the best way of evaluating if an intervention has an impact on pupil outcomes, as the two groups can then be assumed to be the same, therefore any differences in outcomes found can be attributed to the intervention.

As part of the evaluation, all schools will be asked to:

- Complete a short survey at the start and end of the study, covering Teaching as Usual (TAU) practices
- Provide selected information about participating pupils (a template will be provided by the evaluation team)
- Facilitate the assessments of pupils prior to the intervention (April-July 2021) and at the end of the study (May-July 2022) All assessments will be conducted remotely
- Enable selected staff to participate in telephone interviews at the end of the study. Only some staff will be selected from a sample of schools to participate in this part of the research
- Enable selected pupils to participate in interviews (remotely conducted) at the end of the study. Only some pupils will be selected from a sample of schools to participate in this part of the research. Parents/guardians will be asked for their consent for their child to be interviewed.

4. ***Data sharing and data protection***

- For the purposes of conducting the evaluation to assess the impact of Headsprout®, Bangor University and the University of Warwick will become data controllers and processor of personal data of pupils (e.g., pupil names, data from the Headsprout® programme) obtained from schools and other sources such as the National Pupil Database.
- The University of Warwick will be a data controller for all data collected as part of the evaluation.
- At the end of the trial, the University of Warwick will share the data with EEF's data archive processor through secure data portals, where it will be encrypted and saved to secure servers. At that point EEF will become the data controller and the University of Warwick will no longer have any responsibility for it.
- The legal basis for processing personal data for this project is informed consent.
- The delivery and evaluation teams will securely delete all personal data within six months of the project finishing. The University of Warwick will retain the data from this project until Spring 2032 to permit further analysis.
  The Privacy Notice for this project is attached.
- Pupils will be asked to complete short assessments in April-July 2021 and May-July 2022. In addition, the delivery team will collect data about pupils' participation in the intervention directly from the Headsprout® programme, and this will be shared with the evaluation team.
- Further matching to NPD and other administrative data may take place during subsequent research.
- Your school's data will be treated with the strictest confidence and will be transferred securely and saved in secure locations only accessible by the delivery and evaluation teams in line with GDPR and the Data Protection Act 2018.
- We will not use names or the name of your school in any report arising from this project.

5. ***Responsibilities***

a. **The delivery team (Bangor University) will:**

- Recruit special schools in England for this study
- Facilitate the sharing of school and pupil data (e.g., names, Free School Meal eligibility) with the evaluation team, so that they are able to undertake their independent evaluation.
- Train three members of staff (including the staff who will be delivering the intervention, and at least one teacher) from schools in the Headsprout® group to deliver the intervention to pupils in their school
- Provide ongoing supervision to teachers and TAs who are delivering the intervention. This will take the form of monthly school meetings either remotely or, if possible, in-situ, with phone supervision and email support provided in between these.
- Act as the main point of contact for schools and parents for anything to do with the intervention
- Collate data on implementation throughout the study, both to inform implementation support and to provide implementation data for the study

- Provide progress updates and implementation information to headteachers and/or other members of the senior leadership team throughout the study
- Collect data from the Headsprout® programme at the end of the study about pupils' progress with the intervention. They will then share this with the University of Warwick so that they are able to undertake their independent evaluation.

b. <u>The evaluation team (University of Warwick) will:</u>

- Share a data collection template with schools and assist schools with data collection where necessary
- Collect the consent forms from the schools
- Conduct the randomisation of schools to the intervention (receiving Headsprout®) and control groups (not receiving Headsprout®)
- Act as the main point of contact for schools regarding data collection
- Coordinate with schools to arrange remote pupil testing in April-July 2021 and May-July 2022
- Carry out questionnaires with schools about their Teaching as Usual (TAU) practice in April-July 2021 and May-July 2022
- Carry out interviews by telephone with a sample of staff who delivered the intervention
- Carry out remote interviews with a sample of pupils who received the intervention
- Collect and analyse the data from the project and write up the findings
- Ensure that any parental withdrawals (once the intervention has started) are attended to as quickly as possible.
- Disseminate findings from the study - the final summary report for this project will be shared with all participating schools and will be available online on the EEF's website

c. <u>The schools will</u>:

- Name a 'Project Champion' to serve as the main point of contact for the school with the delivery and evaluation teams
- Send parents of pupils in KS1 and KS2 who are selected to take part in the project the parent information sheet, consent form, and the link to the online privacy notice
- Follow-up with parents and collect consent forms
- Provide the evaluation team with the data required to evaluate the project (ensuring accuracy of the data and removal of all parents and pupils who have withdrawn from the study)
- Allow time for staff to complete the TAU questionnaires in April-July 2021 and May-July 2022
- Allow time for each remote assessment phase and liaise with the evaluation team to find appropriate dates and times for assessments to take place
- Inform the evaluation team as soon as possible on planned assessment dates of any pupils absent that day
- Ensure the shared understanding and support of all staff for the project and personnel involved
- Inform the delivery team if the school is taking part in another EEF funded project
- **<u>Headsprout® Schools Only:</u>** Ensure staff are briefed about the programme and their role in it and support them to complete the training and delivery of the intervention (between 10 and 20 hours of teacher/TA time per week)
- **<u>Headsprout® Schools Only:</u>** Liaise with the evaluation team and assist in the arrangement of remote case studies and focus groups, if selected to take part, enabling short telephone interviews with relevant staff and providing a suitable space for remote interviews with pupils
- **<u>Headsprout® Schools Only:</u>** Allow the delivery team access to collect data from the Headsprout® programme, should this be necessary
- **<u>Control Schools Only:</u>** Schools will not access or try to access the intervention until the 2022/23 academic year

d. All parties will:

- Provide such assistance to each other as is reasonably required to enable all parties to comply with requests from parents and pupils who are involved in the project to exercise their rights under data protection legislation
- Comply with ethical guidelines and EU data protection laws including the General Data Protection Regulation and the data protection laws of the UK including the Data Protection Act 2018
- Use all reasonable endeavours to work together collaboratively and productively, in particular in relation to meeting key dates and timeframes set out in the School Information sheet.

6. <u>Funding and costs</u>

As set out in the information sheet provided to schools, a contribution to training and implementation support costs of £200 is required from schools in the **Headsprout® group**. It is not anticipated that there will be any significant implementation costs for schools, although there may be costs related to the allocation of teachers and TAs to deliver the intervention and some minor administrational costs at various points in the project.

Schools in the **control group** will receive two payments totalling £1000. The first payment of £250 will be paid following pre-test assessments, and the second payment of £750 will be paid following post-test assessments.

However, for the avoidance of doubt, any funding provided by the EEF for this research will cover the delivery and evaluation teams' staff costs and implementation costs only, and it cannot be used for any administrative costs that may be incurred by schools.

7. <u>Named contacts within delivery and evaluation teams</u>

<u>Delivery team:</u>
Dr Emily Roberts-Tyler
Email: **e.j.tyler@bangor.ac.uk**; Telephone: 01248 383962

<u>Evaluation team:</u>
Dr Louise Denne
Email: **L.Denne@warwick.ac.uk**; Telephone: 024 7652 3638

Dr Samantha Flynn
Email: **S.Flynn.1@warwick.ac.uk**; Telephone: 07823 362152

8. <u>No partnership or agency</u>

Nothing in this agreement is intended to, or shall be deemed to, establish any legal partnership or joint venture between any of the parties, constitute any party the agent of another party, or authorise any party to make or enter into any commitments for or on behalf of any other party.

9. <u>Not Binding Terms</u>

Nothing in this document will constitute or evidence a legally binding contract to create legal relations between the Parties.

We commit to participating in the Headsprout® study as detailed above [ ] (*please tick*).

We understand that if we are chosen to receive Headsprout® Early Reading, we will pay a contribution to training and implementation support costs, totalling £200. [ ] (*please tick*).

We agree that if we are chosen to be in the control group for the study, we will not use Headsprout® Early Reading during the study period. [ ] (*please tick*).

1. Signature of Head teacher of the School: ______________________________________

Date: ______________________________________

Full name: ______________________________________

Position: ______________________________________

Contact email: ______________________________________

2. Project Champion(s):

Name: ______________________________________

Job title/role: ______________________________________

Email address: ______________________________________

Name: ______________________________________

Job title/role: ______________________________________

Email address: ______________________________________

Appendix H: R analysis code

```
#==============================================#

# HERiSS study analysis script

#==============================================#


# Paul Thompson - 14-09-2021

#  edits - PT 13-10-2022 - added data wrangling parts.



# Load required R packages

library(lme4)

library(lmerTest)

library(glmmTMB)

library(tidyverse)

library(mediation)

library(sjstats)

library(naniar)

library(haven)

library(stringr)

library(dplyr)

library(sjPlot)

library(performance)

library(mice)

library(miceadds)

library(mitml)

library(countimp)

library(broom.mixed)

library(eefAnalytics)

library(readxl)
```

```
library(ivpack)

library(ivreg)

library(robustlmm)

library(robustlm)

library(r2glmm)



#library(devtools)

#devtools::install_git(url = "https://github.com/kkleinke/countimp", branch = "master")


#==============================================================================#


# Load HERiSS study data


HERiSS_BL_data<-read_sav("M:/EA/CEDAR/EEF - Headsprout/Data files/Baseline data/CW data cleaning/HERiSS Data Collection (Baseline)_Clean_04_11_2022_PT.sav")


HERiSS_PT_data<-read_sav("M:/EA/CEDAR/EEF - Headsprout/Data files/Post-testing data/CW data cleaning/HERiSS Data Collection (Post-testing)_August 26, 2022_12.10.22 CW PT 08_11_2022_clean.sav")


#remove SPSS label formatting (unnecessary for analysis purposes)

HERiSS_BL_data<-zap_formats(zap_labels(HERiSS_BL_data))

HERiSS_PT_data<-zap_formats(zap_labels(HERiSS_PT_data))


# replace all 999 values (SPSS missing data codes) with NA (as recognised by R)

HERiSS_PT_data <- HERiSS_PT_data %>% replace_with_na_all(condition = ~.x == 999)

HERiSS_BL_data <- HERiSS_BL_data %>% replace_with_na_all(condition = ~.x == 999)


#================================================#

#Further data cleaning - Baseline - PT 13/09/2022


#relates to any blank data to be coded as missing NA.

HERiSS_BL_data$School_Year <- str_replace_all(HERiSS_BL_data$School_Year, fixed(" "), "")
```

```
HERiSS_BL_data <- HERiSS_BL_data %>%
  mutate(across(where(is.character), ~na_if(., "")))


#all measurements were reordered between may and July, so always 'end of year'


#Post test data column headers need renaming to be clear in the analyses.

HERiSS_PT_data <- HERiSS_PT_data%>% dplyr::rename(., Pupil_ID = Q23, LNF_12mth=Q12,PSF_12mth=Q13,NWF_CLS_12mth=Q14,NWF_WRC_12mth=Q15,WRF_12mth=Q16,ORF_Correct_12mth=Q17,ORF_Errors_12mth=Q18,RSCS1_12mth=Q25,RSCS2_12mth=Q26,RSCS3_12mth=Q27,RSCS4_12mth=Q28,RSCS5_12mth=Q29,RSCS6_12mth=Q30)


#================================================#


#recode RSCS data, so that 0 and 1, all 3s are coded as NA.


HERiSS_BL_data <- HERiSS_BL_data %>% dplyr::mutate(RSCS1=dplyr::recode(RSCS1,`1` = 1L, `2` = 2L,`3`=3L),
                             RSCS2=dplyr::recode(RSCS2,`1` = 2L, `2` = 1L,`3`=3L),
                             RSCS3=dplyr::recode(RSCS3,`1` = 2L, `2` = 1L,`3`=3L),
                             RSCS4=dplyr::recode(RSCS4,`1` = 2L, `2` = 1L,`3`=3L),
                             RSCS5=dplyr::recode(RSCS5,`1` = 1L, `2` = 2L,`3`=3L),
                             RSCS6=dplyr::recode(RSCS6,`1` = 1L, `2` = 2L,`3`=3L))


HERiSS_BL_data <- HERiSS_BL_data %>% dplyr::mutate(RSCS1=na_if(RSCS1, 3),
                             RSCS2=na_if(RSCS2, 3),
                             RSCS3=na_if(RSCS3, 3),
                             RSCS4=na_if(RSCS4, 3),
                             RSCS5=na_if(RSCS5, 3),
                             RSCS6=na_if(RSCS6, 3))


HERiSS_PT_data <- HERiSS_PT_data %>% dplyr::mutate(RSCS1_12mth=dplyr::recode(RSCS1_12mth,`1` = 1L, `2` = 2L,`3`=3L),
                             RSCS2_12mth=dplyr::recode(RSCS2_12mth,`1` = 2L, `2` = 1L,`3`=3L),
                             RSCS3_12mth=dplyr::recode(RSCS3_12mth,`1` = 2L, `2` = 1L,`3`=3L),
```

```
RSCS4_12mth=dplyr::recode(RSCS4_12mth,`1` = 2L, `2` = 1L,`3`=3L),

RSCS5_12mth=dplyr::recode(RSCS5_12mth,`1` = 1L, `2` = 2L,`3`=3L),

RSCS6_12mth=dplyr::recode(RSCS6_12mth,`1` = 1L, `2` = 2L,`3`=3L))


HERiSS_PT_data <- HERiSS_PT_data %>% dplyr::mutate(RSCS1_12mth=na_if(RSCS1_12mth, 3),

RSCS2_12mth=na_if(RSCS2_12mth, 3),

RSCS3_12mth=na_if(RSCS3_12mth, 3),

RSCS4_12mth=na_if(RSCS4_12mth, 3),

RSCS5_12mth=na_if(RSCS5_12mth, 3),

RSCS6_12mth=na_if(RSCS6_12mth, 3))


#=================================================#


# BASELINE - Composite score calculation - 13/09/2022


HERiSS_BL_data <- HERiSS_BL_data %>%

  mutate(DIBELS_Comp_BL = (((((ifelse(is.na(LNF),0,LNF)*10.72)+(ifelse(is.na(PSF),0,PSF)*2.13)+(ifelse(is.na(NWF_CLS),0,NWF_CLS)*23.13)+(ifelse(is.na(NWF_WRC),0,NWF_WRC)*7.79)+(ifelse(is.na(WRF),0,WRF)*13.51)+(ifelse(is.na(ORF_Correct),0,ORF_Correct)*25.36)+(ifelse(is.na(ORF_Accuracy),0,ORF_Accuracy)*0.25))-3371)/2251)*40)+ 440)



HERiSS_BL_data <- HERiSS_BL_data %>%

  mutate(DIBELS_Comp_BL = ifelse(is.na(LNF) & is.na(PSF) & is.na(NWF_CLS) & is.na(NWF_WRC) & is.na(WRF) & is.na(ORF_Correct) & is.na(ORF_Accuracy),NA,DIBELS_Comp_BL))


HERiSS_BL_data <- HERiSS_BL_data %>%

  mutate(RSCS_total_BL = rowSums(.[32:37],na.rm=TRUE))


HERiSS_BL_data <- HERiSS_BL_data %>%

  mutate(RSCS_total_BL = ifelse(is.na(RSCS1) & is.na(RSCS2) & is.na(RSCS3) & is.na(RSCS4) & is.na(RSCS5) & is.na(RSCS6),NA,RSCS_total_BL))
```

```
#================================================#

# Post Testing - Composite score calculation - 13/09/2022

# initially calculate the ORF accuracy from the ORF 'correct' and 'Errors'

HERiSS_PT_data <- HERiSS_PT_data %>% mutate(ORF_Accuracy_12mth = ifelse(ORF_Correct_12mth==0,0,((ORF_Correct_12mth/(ORF_Errors_12mth+ORF_Correct_12mth)*100))))

#calculate composite at post test

HERiSS_PT_data <- HERiSS_PT_data %>%

  mutate(DIBELS_Comp = (((((ifelse(is.na(LNF_12mth),0,LNF_12mth)*10.72)+(ifelse(is.na(PSF_12mth),0,PSF_12mth)*2.13)+(ifelse(is.na(NWF_CLS_12mth),0,NWF_CLS_12mth)*23.13)+(ifelse(is.na(NWF_WRC_12mth),0,NWF_WRC_12mth)*7.79)+(ifelse(is.na(WRF_12mth),0,WRF_12mth)*13.51)+(ifelse(is.na(ORF_Correct_12mth),0,ORF_Correct_12mth)*25.36)+(ifelse(is.na(ORF_Accuracy_12mth),0,ORF_Accuracy_12mth)*0.25))-3371)/2251)*40)+ 440)

#adjust so that any with all NAs are coded as completely missing, not zero for composite.

HERiSS_PT_data <- HERiSS_PT_data %>%

  mutate(DIBELS_Comp = ifelse(is.na(LNF_12mth) & is.na(PSF_12mth) & is.na(NWF_CLS_12mth) & is.na(NWF_WRC_12mth) & is.na(WRF_12mth) & is.na(ORF_Correct_12mth) & is.na(ORF_Accuracy_12mth),NA,DIBELS_Comp))

HERiSS_PT_data <- HERiSS_PT_data %>%

  mutate(RSCS_total = rowSums(.[27:32],na.rm=TRUE))

HERiSS_PT_data <- HERiSS_PT_data %>%

  mutate(RSCS_total = ifelse(is.na(RSCS1_12mth) & is.na(RSCS2_12mth) & is.na(RSCS3_12mth) & is.na(RSCS4_12mth) & is.na(RSCS5_12mth) & is.na(RSCS6_12mth),NA,RSCS_total))

#================================================#
```

```
#select correct variables for analysis in the baseline data

HERiSS_BL_data <- HERiSS_BL_data %>% dplyr::select(School_ID, Pupil_ID, School_Year, FSM, EAL, School_Size, Condition, SEND_SpLD, SEND_MLD, SEND_SLD, SEND_PMLD, SEND_SLCN, SEND_ASD, SEND_Other, SEND_PD, SEND_SEMH, LNF, PSF, NWF_CLS, NWF_WRC, WRF, ORF_Correct, ORF_Errors, ORF_Accuracy, DIBELS_Comp_BL,RSCS1,RSCS2,RSCS3,RSCS4,RSCS5,RSCS6,RSCS_total_BL)


#select correct variables for analysis in the post-test data

HERiSS_PT_data <- HERiSS_PT_data %>% dplyr::select(Pupil_ID, LNF_12mth, PSF_12mth, NWF_CLS_12mth, NWF_WRC_12mth, WRF_12mth, ORF_Correct_12mth, ORF_Errors_12mth, ORF_Accuracy_12mth, DIBELS_Comp,RSCS1_12mth,RSCS2_12mth,RSCS3_12mth,RSCS4_12mth,RSCS5_12mth,RSCS6_12mth,RSCS_total)


#For all send variables ensure that blanks are coded as zero to correspond to correct dummy coding.

HERiSS_BL_data <- HERiSS_BL_data %>%

  mutate_at(vars(matches("SEND")), ~coalesce(.,0))


for(i in 1:dim(HERiSS_BL_data)[1])

{

  HERiSS_BL_data[i, 8:16][!!rowSums(HERiSS_BL_data[i, 8:16])==0,] <- NA

}


#==================================================#


#Merge data sets together using the pupil ID (baseline and postest into one dataset)


HERiSS_data = HERiSS_BL_data %>% full_join(HERiSS_PT_data,by="Pupil_ID")



#Load additional functions for running script

#source("HERiSS_additional_functions.R")


#==================================================#

# check missing data
```

```
#==================================================#

vis_miss(HERiSS_data)

#==================================================#
#model setups
#==================================================#

#ensure dummy coding is correct (condition and school size covariate)

HERiSS_data <- HERiSS_data %>% dplyr::mutate(Condition=dplyr::recode(Condition,`1` = 1L, `2` = 0L))

#added after unblinding occurred. Analysed blind initially.
HERiSS_data$Condition <- factor(HERiSS_data$Condition,levels=c(0,1),labels=c("Control","Intervention"))



HERiSS_data <- HERiSS_data %>% dplyr::mutate(School_Size=dplyr::recode(School_Size,`1` = 1L, `2` = 0L))

HERiSS_data$School_Size <- factor(HERiSS_data$School_Size,levels=c(0,1),labels=c("Small","Large"))

HERiSS_data <- HERiSS_data %>% dplyr::mutate(FSM=dplyr::recode(FSM,`1` = 1L, `2` = 0L))

HERiSS_data$FSM <- factor(HERiSS_data$FSM,levels=c(0,1),labels=c("No","Yes"))

HERiSS_data <- HERiSS_data %>% dplyr::mutate(EAL=dplyr::recode(EAL,`1` = 1L, `2` = 0L))

HERiSS_data$EAL <- factor(HERiSS_data$EAL,levels=c(0,1),labels=c("No","Yes"))



####DIBELS composite total score at 12 months####
```

```
#unconditional model

HERiSS_m0a_unc<-lmer(DIBELS_Comp~1+(1|School_ID),data=HERiSS_data)

#SAP analysis

HERiSS_m0a<-lmer(DIBELS_Comp~DIBELS_Comp_BL+School_Size+Condition+(1|School_ID),data=HERiSS_data)

#robust alternative

HERiSS_m3a<-
rlmer(DIBELS_Comp~DIBELS_Comp_BL+School_Size+Condition+(1|School_ID),data=HERiSS_data)

#output to table (publication ready)

sjPlot::tab_model(HERiSS_m0a,HERiSS_m3a,pred.labels = c("Intercept", "DIBELS Baseline", "School Size", "Intervention"),
        dv.labels = c("Primary analysis(two-level)","Robust primary analysis"),
        string.pred = "Coefficient",
        string.se = "S.E",
        string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=TRUE)

#YEF requested format

sjPlot::tab_model(HERiSS_m0a,HERiSS_m3a,pred.labels = c("Intercept", "DIBELS Baseline", "School Size", "Intervention"),
        dv.labels = c("Primary analysis(two-level)","Robust primary analysis"))

#ICC

icc(HERiSS_m0a_unc,ci = 0.95,iterations = 100)

icc(HERiSS_m0a,ci = 0.95,iterations = 100)

icc(HERiSS_m3a,ci = 0.95,iterations = 100)

#r^2
```

```
r2beta(HERiSS_m0a)

#====================================================================================
===#

#Additional data wrangling for longitudinal plot and alternative primary model (not specified in SAP)

HERiSS_long<-HERiSS_data %>%
dplyr::select(Pupil_ID,School_ID,Condition,School_Size,DIBELS_Comp,DIBELS_Comp_BL) %>%
  pivot_longer(!c(Pupil_ID,School_ID,Condition,School_Size),names_to='Time',values_to='DIBELS_comp')

HERiSS_long$Time <-factor(HERiSS_long$Time,levels=c("DIBELS_Comp_BL",
"DIBELS_Comp"),labels=c("Baseline","Post test"))

ggplot(HERiSS_long,aes(y=DIBELS_comp,x=Time,colour=Condition))+geom_line(aes(group = interaction(Pupil_ID,
Condition)),alpha=0.3)+ geom_smooth(aes(group=Condition),se=FALSE,size=1.5)+theme_bw()

#edited plot based on reviewers suggestions:

ggplot(HERiSS_long,aes(y=DIBELS_comp,x=Time,colour=Condition))+
geom_smooth(aes(group=Condition),se=FALSE,size=1.5)+theme_bw()

#====================================================================================
===#

#Alternative model
HERiSS_prim_extra <-
lmer(DIBELS_comp~School_Size+Time*Condition+(1|School_ID)+(1|Pupil_ID),data=HERiSS_long)

sjPlot::tab_model(HERiSS_prim_extra,pred.labels = c("Intercept", "School Size", "Time", "Intervention", "Time x
Intervention"),
          dv.labels = c("Supplementary analysis(two-level)"),
          string.pred = "Coefficient",
          string.se = "S.E",
          string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=TRUE)
```

```
#====================================================================================
===#

# To get Hedges' g for primary analysis

library(eefAnalytics)


# TO obtain Effect size - hedge's g (R package eefAnalytics) -2 level model

HERiSS_data2<-as.data.frame(HERiSS_data)

HERiSS_data2$School_ID <- as.numeric(as.factor(HERiSS_data2$School_ID))


output1 <-
crtFREQ(DIBELS_Comp~DIBELS_Comp_BL+School_Size+Condition,random="School_ID",intervention="Condition",
data=HERiSS_data2)


summary(output1)


output1$ES #extract hedges g.


#obtain the R2 values.


HERiSS_test<-HERiSS_data


HERiSS_test$Condition<-as.numeric(HERiSS_test$Condition)

HERiSS_test$School_Size<-as.numeric(HERiSS_test$School_Size)

HERiSS_test$School_ID<-as.factor(HERiSS_test$School_ID)


HERiSS_test$DIBELS_Comp_BL.cwc <- misty::center(HERiSS_test$DIBELS_Comp_BL, type = "CWC", cluster =
HERiSS_test$School_ID)


test.unc<-lmer(DIBELS_Comp~1+(1|School_ID),data=HERiSS_test)

test<-lmer(DIBELS_Comp~DIBELS_Comp_BL.cwc+School_Size+Condition+(1|School_ID),data=HERiSS_test)


r2mlm(test)
```

```
#=====================================================================================
====#

# To get Hedges' g for additonal covariates analysis


# To obtain Effect size - hedge's g (R package eefAnalytics) -2 level model


output2 <-
crtFREQ(DIBELS_Comp~DIBELS_Comp_BL+Condition+School_Size+FSM+SEND_SpLD+SEND_MLD+SEND_SLD
+SEND_SLCN+SEND_ASD+SEND_Other+SEND_PD+SEND_SEMH+EAL+

                  +
School_Size:Condition+FSM:Condition+SEND_SpLD:Condition+SEND_MLD:Condition+SEND_SLD:Condition+SEN
D_SLCN:Condition+SEND_ASD:Condition+SEND_Other:Condition+SEND_PD:Condition+SEND_SEMH:Condition+E
AL:Condition,random="School_ID",intervention="Condition",data=HERiSS_data2)

summary(output2)


output2$ES #extract hedges g.


#=====================================================================================
====#




#hedges g (robust model)


# values from model imputed into formula: g = estimate for fixed effect / (sqrt of sum of variances of random effects)


round(0.46/sqrt(66.72),2)


round((0.46+1.96*1.04440)/sqrt(66.72),2)


round((0.46-1.96*1.04440)/sqrt(66.72),2)


#=========================================================#

#Sensitivity analysis for primary analysis 2-levl vs 3-level model
```

```
# PLEASE NOTE: this provides a singularity due to tiny/non existent variance due to the extra level in model.

#HERiSS_m0b<-lmer(DIBELS_Comp~DIBELS_Comp_BL+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data)


#HERiSS_m0b <-update(HERiSS_m0a, . ~ . -(1|School_ID) +(1|School_ID/School_Year))


#anova(HERiSS_m0a,HERiSS_m0b)

#rand(HERiSS_m0b) #test random effects


#compare_performance(HERiSS_m0a,HERiSS_m0b)


#Results from this analysis cannot be reported as the model does not converge and will be biased.


#============================================================#

# FSM subgroup analysis


#Subgroup model - FSM pupils only

HERiSS_m0a_FSM_subset<-rlmer(DIBELS_Comp~DIBELS_Comp_BL+School_Size+Condition+(1|School_ID),data=HERiSS_data[HERiSS_data$FSM=="Yes",])


#Subgroup model 2 - all pupils but FSM as moderator.

HERiSS_m0_FSM_int<-rlmer(DIBELS_Comp~DIBELS_Comp_BL+School_Size+Condition*FSM+(1|School_ID),data=HERiSS_data)


sjPlot::tab_model(HERiSS_m0a,HERiSS_m0a_FSM_subset,HERiSS_m0_FSM_int,pred.labels = c("Intercept", "DIBELS Baseline", "School Size", "Intervention", "FSM", "Intervention x FSM"),

          dv.labels = c("Primary analysis","FSM individuals only","FSM interaction"),

          string.pred = "Coeffcient",

          string.se = "S.E",

          string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=FALSE)


sjPlot::tab_model(HERiSS_m0a,HERiSS_m0a_FSM_subset,HERiSS_m0_FSM_int,pred.labels = c("Intercept", "DIBELS Baseline", "School Size", "Intervention", "FSM", "Intervention x FSM"),
```

```
dv.labels = c("Primary analysis","FSM individuals only","FSM interaction"))


#obtain the R2 values.

r2beta(lmer(DIBELS_Comp~DIBELS_Comp_BL+School_Size+Condition+(1|School_ID),data=HERiSS_data[HERiSS_data$FSM=="Yes",]))


#==========================================================#


#hedges g (robust model) - FSM subset


# values from model inputted into formula: g = estimate for fixed effect / (sqrt of sum of variances of random effects)


round(0.67266/sqrt(6.178+51.797),2)

round((0.67266+1.96*1.50646)/sqrt(6.178+51.797),2)

round((0.67266-1.96*1.50646)/sqrt(6.178+51.797),2)


#hedges g (robust model) - FSM interaction


# values from model inputted into formula: g = estimate for fixed effect / (sqrt of sum of variances of random effects)


round(-0.39816/sqrt(66.07),2)

round((-0.39816+1.96*2.06816)/sqrt(66.07),2)

round((-0.39816-1.96*2.06816)/sqrt(66.07),2)


#==========================================================#

#Additional covariates


#Omit SEND categories: SEND_MSI, SEND_VI,SEND_HI, due to tiny numbers or zero individuals in data.
```

```
HERiSS_m1a<-
lmer(DIBELS_Comp~DIBELS_Comp_BL+Condition+School_Size+FSM+SEND_SpLD+SEND_MLD+SEND_SLD+SE
ND_SLCN+SEND_ASD+SEND_Other+SEND_PD+SEND_SEMH+EAL+

School_Size:Condition+FSM:Condition+SEND_SpLD:Condition+SEND_MLD:Condition+SEND_SLD:Condition+SEN
D_SLCN:Condition+SEND_ASD:Condition+SEND_Other:Condition+SEND_PD:Condition+SEND_SEMH:Condition+E
AL:Condition+(1|School_ID),data=HERiSS_data)


HERiSS_m1b<-
rlmer(DIBELS_Comp~DIBELS_Comp_BL+Condition+School_Size+FSM+SEND_SpLD+SEND_MLD+SEND_SLD+S
END_SLCN+SEND_ASD+SEND_Other+SEND_PD+SEND_SEMH+EAL+

School_Size:Condition+FSM:Condition+SEND_SpLD:Condition+SEND_MLD:Condition+SEND_SLD:Condition+SEN
D_SLCN:Condition+SEND_ASD:Condition+SEND_Other:Condition+SEND_PD:Condition+SEND_SEMH:Condition+E
AL:Condition+(1|School_ID),data=HERiSS_data)


table(sjPlot::tab_model(HERiSS_m1a,HERiSS_m1b,
            dv.labels = c("Additional Primary analysis","Additional Robust analysis"),
            string.pred = "Coeffcient",
            string.se = "S.E",
            string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=FALSE)

   sjPlot::tab_model(HERiSS_m1a,HERiSS_m1b,
            dv.labels = c("Additional Primary analysis","Additional Robust analysis"))

   summary(HERiSS_m1b)

#========================================#
# ICC
   icc(HERiSS_m1a,ci = 0.95,iterations = 100)
#========================================#


   #hedges g (std model) - additional primary analysis
```

```
# values from model inputted into formula: g = estimate for fixed effect / (sqrt of sum of variances of random effects)

round( 12.03071/sqrt(6.658+101.571),2)
round(( 12.80306+1.96*9.55286)/sqrt(6.658+101.571),2)
round(( 12.80306-1.96*9.55286)/sqrt(6.658+101.571),2)

#hedges g (std model) - additional primary analysis - SEND SEMH

# values from model inputted into formula: g = estimate for fixed effect / (sqrt of sum of variances of random effects)

round( 11.49681/sqrt(6.658+101.571),2)
round(( 11.49681+1.96*5.06336)/sqrt(6.658+101.571),2)
round(( 11.49681-1.96*5.06336)/sqrt(6.658+101.571),2)

#hedges g (std model) - additional primary analysis - intervention*SEMH

# values from model inputed into formula: g = estimate for fixed effect / (sqrt of sum of variances of random effects)

round( -14.23427/sqrt(6.658+101.571),2)
round(( -14.23427+1.96*7.60883)/sqrt(6.658+101.571),2)
round(( -14.23427-1.96*7.60883)/sqrt(6.658+101.571),2)


#hedges g (robust model) - additional primary analysis - MLD*intervention

# values from model inputed into formula: g = estimate for fixed effect / (sqrt of sum of variances of random effects)

round( -9.93148/sqrt(62.07),2)
round(( -9.93148+1.96*4.18346)/sqrt(62.07),2)
round(( -9.93148-1.96*4.18346)/sqrt(62.07),2)
```

```
#hedges g (robust model) - additional primary analysis - ASD*intervention

# values from model inputed into formula: g = estimate for fixed effect / (sqrt of sum of variances of random effects)

round( -7.51559/sqrt(62.07),2)
round(( -7.51559+1.96*2.96832)/sqrt(62.07),2)
round(( -7.51559-1.96*2.96832)/sqrt(62.07),2)

#hedges g (robust model) - additional primary analysis - SEMH*intervention

# values from model inputed into formula: g = estimate for fixed effect / (sqrt of sum of variances of random effects)

round( -12.35944/sqrt(62.07),2)
round(( -12.35944+1.96*5.90737)/sqrt(62.07),2)
round(( -12.35944-1.96*5.90737)/sqrt(62.07),2)


######### get p values

coefs <- data.frame(coef(summary(HERiSS_m1a)))

# get coefficients from robust model to extract t-values
coefs.robust <- coef(summary(HERiSS_m1b))

# calculate p-values based on robust t-values and non-robust approx. DFs
p.values <- 2*pt(abs(coefs.robust[,3]), coefs$df, lower=FALSE)
p.values["ConditionIntervention"]

#============================================================#
#Response to reviewers - overparameterised model (agree), so have omitted SEND variable as all children in special schools fall under this criteria.
```

```
HERiSS_m1c<-rlmer(DIBELS_Comp~DIBELS_Comp_BL+Condition+School_Size+FSM+EAL+School_Size:Condition+FSM:Condition+EAL:Condition+(1|School_ID),data=HERiSS_data)



sjPlot::tab_model(HERiSS_m1a,HERiSS_m1b, HERiSS_m1c,
          dv.labels = c("Additional Primary analysis","Additional Robust analysis","Additional Robust Analysis (no SEND)"))


#=============================================================#

coefs <- data.frame(coef(summary(lmer(DIBELS_Comp~DIBELS_Comp_BL+Condition+School_Size+FSM+EAL+School_Size:Condition+FSM:Condition+EAL:Condition+(1|School_ID),data=HERiSS_data))))


# get coefficients from robust model to extract t-values
coefs.robust <- coef(summary(HERiSS_m1c))


# calculate p-values based on robust t-values and non-robust approx. DFs
p.values <- 2*pt(abs(coefs.robust[,3]), coefs$df, lower=FALSE)
p.values["ConditionIntervention"]
#=============================================================#


round( 3.25436/sqrt(66.53),2)
round(( 3.25436+1.96*3.88265)/sqrt(66.53),2)
round(( 3.25436-1.96*3.88265)/sqrt(66.53),2)



#===============================================#


###SECONDARY OUTCOMES###


#PLEASE NOTE, as per SAP, all subsequent analyses are based on 2-level model following primary outcome sensitivity analysis.
```

```
#Letter naming fluency

#two-level glmm (poisson)
HERiSS_m_LNF0<-glmer(LNF_12mth~LNF+School_Size+Condition+(1|School_ID),data=HERiSS_data,family=poisson)

#three-level glmm (poisson)
#HERiSS_m_LNF0<-glmer(LNF_12mth~LNF+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data,family=poisson)

#checks for overdispersion and zero inflation as required.
check_overdispersion(HERiSS_m_LNF0)
check_zeroinflation(HERiSS_m_LNF0)

summary(HERiSS_m_LNF0)

#if overdispersed, use:

#two-level glmm (zero inflated negative binomial)
HERiSS_m_LNF0b<-glmmTMB(LNF_12mth~LNF+School_Size+Condition+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data)
#HERiSS_m_LNF0b<-glmer.nb(LNF_12mth~LNF+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data)

# FSM subgroup analysis (based on two-level glmm (zero inflated negative binomial)

#FSM only group
HERiSS_m_LNF0_FSM_subset<-glmmTMB(LNF_12mth~LNF+School_Size+Condition+(1|School_ID),data=HERiSS_data[HERiSS_data$FSM==1,],family="nbinom2",ziformula=~1)

#All pupils with FSM as moderator
```

```
HERiSS_m_LNF0_FSM_int<-
glmmTMB(LNF_12mth~LNF+School_Size+Condition*FSM+(1|School_ID),data=HERiSS_data,family="nbinom2",zifor
mula=~1)


#results table for all models compared.

sjPlot::tab_model(HERiSS_m_LNF0,HERiSS_m_LNF0b,HERiSS_m_LNF0_FSM_subset,HERiSS_m_LNF0_FSM_int
,
        dv.labels = c("Primary LNF analysis","Overdispersed / zeroinflated","FSM individuals only","FSM
interaction"),
        string.pred = "Coefficient",
        string.se = "S.E",
        string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=FALSE)


sjPlot::tab_model(HERiSS_m_LNF0,HERiSS_m_LNF0b,HERiSS_m_LNF0_FSM_subset,HERiSS_m_LNF0_FSM_int
,
        dv.labels = c("Primary LNF analysis","Overdispersed / zeroinflated","FSM individuals only","FSM
interaction"))


#========================================#
# ICC


#poisson - null model
icc(glmer(LNF_12mth~1+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)


#poisson - conditional
icc(glmer(LNF_12mth~LNF+School_Size+Condition+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)


#zinb - null
icc(glmmTMB(LNF_12mth~1+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))


#zinb - conditional

icc(glmmTMB(LNF_12mth~LNF+School_Size+Condition+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS
_data))
```

```
#=====================================#

#===============================#

#Phonemic segmentation fluency

#two-level glmm (poisson)
HERiSS_m_PSF0<-
glmer(PSF_12mth~PSF+Condition+School_Size+(1|School_ID),data=HERiSS_data,family=poisson)

#three-level glmm (poisson)
#HERiSS_m_PSF0<-
glmer(PSF_12mth~PSF+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data,family=poisson)

#check for zero-inflation and overdispersion
check_overdispersion(HERiSS_m_PSF0)
check_zeroinflation(HERiSS_m_PSF0)

summary(HERiSS_m_PSF0)

#if overdispersed, use:

#two-level glmm (zero inflated negative binomial)
HERiSS_m_PSF0b<-
glmmTMB(PSF_12mth~PSF+Condition+School_Size+(1|School_ID),data=HERiSS_data,family="nbinom2",ziformula=
~1)
#HERiSS_m_PSF0b<-
glmer.nb(PSF_12mth~PSF+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data)
```

```
summary(HERiSS_m_PSF0)

# FSM subgroup analysis

#FSM pupils only.

HERiSS_m_PSF0_FSM_subset<-
glmmTMB(PSF_12mth~PSF+Condition+School_Size+(1|School_ID),data=HERiSS_data[HERiSS_data$FSM==1,],family="nbinom2",ziformula=~1)

summary(HERiSS_m_PSF0_FSM_subset)

#all pupils with FSM as moderator.

HERiSS_m_PSF0_FSM_int<-
glmmTMB(PSF_12mth~PSF+Condition*FSM+School_Size+(1|School_ID),data=HERiSS_data,family="nbinom2",ziformula=~1)

summary(HERiSS_m_PSF0_FSM_int)

#results table for model comparison.

sjPlot::tab_model(HERiSS_m_PSF0,HERiSS_m_PSF0b,HERiSS_m_PSF0_FSM_subset,HERiSS_m_PSF0_FSM_int
,
          dv.labels = c("Primary PSF analysis","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"),
          string.pred = "Coeffcient",
          string.se = "S.E",
          string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=FALSE)

sjPlot::tab_model(HERiSS_m_PSF0,HERiSS_m_PSF0b,HERiSS_m_PSF0_FSM_subset,HERiSS_m_PSF0_FSM_int
,
          dv.labels = c("Primary PSF analysis","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"))

#========================================#
```

```
# ICC

#poisson - null model
icc(glmer(PSF_12mth~1+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)

#poisson - conditional
icc(glmer(PSF_12mth~PSF+School_Size+Condition+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)

#zinb - null
icc(glmmTMB(PSF_12mth~1+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))

#zinb - conditional
icc(glmmTMB(PSF_12mth~PSF+School_Size+Condition+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))


#==================================#

#nonsense word fluency - Correct Letter Sounds (CLS)

#two-level glmm (poisson)
HERiSS_m_NWF_CLS0<-glmer(NWF_CLS_12mth~NWF_CLS+Condition+School_Size+(1|School_ID),data=HERiSS_data,family=poisson)

#three-level glmm (poisson)
#HERiSS_m_NWF_CLS0<-glmer(NWF_CLS_12mth~NWF_CLS+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data,family=poisson)

#checks for zero-inflation and overdispersion.
check_overdispersion(HERiSS_m_NWF_CLS0)
check_zeroinflation(HERiSS_m_NWF_CLS0)
```

```
summary(HERiSS_m_NWF_CLS0)

#if overdispersed, use:

#two-level glmm (zero-inflated negative binomial)

HERiSS_m_NWF_CLS0b<-
glmmTMB(NWF_CLS_12mth~NWF_CLS+Condition+School_Size+(1|School_ID),data=HERiSS_data,family="nbinom
2",ziformula=~1)

#HERiSS_m_NWF_CLS0b<-
glmer.nb(NWF_CLS_12mth~NWF_CLS+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data)

summary(HERiSS_m_NWF_CLS0b)

# FSM subgroup analysis

HERiSS_m_NWF_CLS0_FSM_subset<-
glmmTMB(NWF_CLS_12mth~NWF_CLS+Condition+School_Size+(1|School_ID),data=HERiSS_data[HERiSS_data$
FSM==1,],family="nbinom2",ziformula=~1)

HERiSS_m_NWF_CLS0_FSM_int<-
glmmTMB(NWF_CLS_12mth~NWF_CLS+Condition*FSM+School_Size+(1|School_ID),data=HERiSS_data,family="n
binom2",ziformula=~1)

sjPlot::tab_model(HERiSS_m_NWF_CLS0,HERiSS_m_NWF_CLS0b,HERiSS_m_NWF_CLS0_FSM_subset,HERiSS
_m_NWF_CLS0_FSM_int,

        dv.labels = c("Primary NWF-CLS analysis","Overdispersed / zeroinflated","FSM individuals only","FSM
interaction"),

        string.pred = "Coeffcient",

        string.se = "S.E",

        string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=TRUE)

sjPlot::tab_model(HERiSS_m_NWF_CLS0,HERiSS_m_NWF_CLS0b,HERiSS_m_NWF_CLS0_FSM_subset,HERiSS
_m_NWF_CLS0_FSM_int,
```

```
                    dv.labels = c("Primary NWF-CLS analysis","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"))


    #========================================#

    # ICC


    #poisson - null model

    icc(glmer(NWF_CLS_12mth~1+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)


    #poisson - conditional

icc(glmer(NWF_CLS_12mth~NWF_WRC+School_Size+Condition+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)


    #zinb - null

    icc(glmmTMB(NWF_CLS_12mth~1+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))


    #zinb - conditional

icc(glmmTMB(NWF_CLS_12mth~NWF_CLS+School_Size+Condition*FSM+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))


    #zinb - conditional(fsm only)

icc(glmmTMB(NWF_CLS_12mth~NWF_CLS+School_Size+Condition+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data[HERiSS_data$FSM=="Yes",]))


    #=================================#


    #nonsense word fluency - words read correctly (WRC)


    #two-level glmm (poisson)

    HERiSS_m_NWF_WRC0<-glmer(NWF_WRC_12mth~NWF_WRC+Condition+School_Size+(1|School_ID),data=HERiSS_data,family=poisson)
```

```
#three-level glmm (poisson)

#HERiSS_m_NWF_WRC0<-glmer(NWF_WRC_12mth~NWF_WRC+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data,family=poisson)



check_overdispersion(HERiSS_m_NWF_WRC0)

check_zeroinflation(HERiSS_m_NWF_WRC0)


summary(HERiSS_m_NWF_WRC0)


#if overdispersed, use:


#two-level glmm (zero-inflated negative binomial)

HERiSS_m_NWF_WRC0b<-glmmTMB(NWF_WRC_12mth~NWF_WRC+Condition+School_Size+(1|School_ID),data=HERiSS_data,family="nbinom2",ziformula=~1)

#HERiSS_m_NWF_WRC0b<-glmer.nb(NWF_WRC_12mth~NWF_WRC+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data)


# FSM subgroup analysis


HERiSS_m_NWF_WRC0_FSM_subset<-glmmTMB(NWF_WRC_12mth~NWF_WRC+Condition+School_Size+(1|School_ID),data=HERiSS_data[HERiSS_data$FSM==1,],family="nbinom2",ziformula=~1)


HERiSS_m_NWF_WRC0_FSM_int<-glmmTMB(NWF_WRC_12mth~NWF_WRC+Condition*FSM+School_Size+(1|School_ID),data=HERiSS_data,family="nbinom2",ziformula=~1)


#see here for explanation on condition effect size direction.

with(HERiSS_data,boxplot((NWF_WRC_12mth-NWF_WRC)~Condition))



sjPlot::tab_model(HERiSS_m_NWF_WRC0,HERiSS_m_NWF_WRC0b,HERiSS_m_NWF_WRC0_FSM_subset,HERiSS_m_NWF_WRC0_FSM_int,
```

```
dv.labels = c("Primary NWF-WRC analysis","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"),

string.pred = "Coefficient",

string.se = "S.E",

string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=FALSE)



sjPlot::tab_model(HERiSS_m_NWF_WRC0,HERiSS_m_NWF_WRC0b,HERiSS_m_NWF_WRC0_FSM_subset,HERiSS_m_NWF_WRC0_FSM_int,

dv.labels = c("Primary NWF-WRC analysis","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"))


#=========================================#

# ICC


#poisson - null model

icc(glmer(NWF_WRC_12mth~1+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)


#poisson - conditional

icc(glmer(NWF_WRC_12mth~NWF_WRC+School_Size+Condition+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)


#zinb - null

#icc(glmmTMB(NWF_WRC_12mth~1+(1|School_ID),family="nbinom2",ziformula=~0,data=HERiSS_data))


#zinb - conditional

#icc(glmmTMB(NWF_WRC_12mth~NWF_WRC+School_Size+Condition+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))


#==================================#


#Oral reading fluency - Words Correct
```

```
#two-level glmm (poisson)

HERiSS_m_ORF0<-
glmer(ORF_Correct_12mth~ORF_Correct+Condition+School_Size+(1|School_ID),data=HERiSS_data,family=poisson
)


#three-level glmm (poisson)

#HERiSS_m_ORF_Correct0<-
glmer(ORF_Correct_12mth~ORF_Correct+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data,f
amily=poisson)

check_overdispersion(HERiSS_m_ORF0)

check_zeroinflation(HERiSS_m_ORF0)


summary(HERiSS_m_ORF0)


#if overdispersed, use:


#two-level glmm (zero-inflated negative binomial)

HERiSS_m_ORF0b<-
glmmTMB(ORF_Correct_12mth~ORF_Correct+Condition+School_Size+(1|School_ID),data=HERiSS_data,family="nb
inom2",ziformula=~1)

#HERiSS_m_ORF_Correct0b<-
glmer.nb(ORF_Correct_12mth~ORF__Correct+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_d
ata)


# FSM subgroup analysis


HERiSS_m_ORF0_FSM_subset<-
glmmTMB(ORF_Correct_12mth~ORF_Correct+Condition+School_Size+(1|School_ID),data=HERiSS_data[HERiSS_
data$FSM==1,],family="nbinom2",ziformula=~1)


HERiSS_m_ORF0_FSM_int<-
glmmTMB(ORF_Correct_12mth~ORF_Correct+Condition*FSM+School_Size+(1|School_ID),data=HERiSS_data,famil
y="nbinom2",ziformula=~1)



sjPlot::tab_model(HERiSS_m_ORF0,HERiSS_m_ORF0b,HERiSS_m_ORF0_FSM_subset,HERiSS_m_ORF0_FSM_i
nt,
```

```
            dv.labels = c("Primary ORF analysis","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"),

            string.pred = "Coeffcient",

            string.se = "S.E",

            string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=FALSE)



sjPlot::tab_model(HERiSS_m_ORF0,HERiSS_m_ORF0b,HERiSS_m_ORF0_FSM_subset,HERiSS_m_ORF0_FSM_int,

            dv.labels = c("Primary ORF analysis","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"))


   #========================================#

   # ICC


   #poisson - null model

   icc(glmer(ORF_Correct_12mth~1+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)


   #poisson - conditional

icc(glmer(ORF_Correct_12mth~ORF_Correct+School_Size+Condition+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)


   #zinb - null

   icc(glmmTMB(ORF_Correct_12mth~1+(1|School_ID),family="nbinom2",ziformula=~0,data=HERiSS_data))


   #zinb - conditional

icc(glmmTMB(ORF_Correct_12mth~ORF_Correct+School_Size+Condition+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))


   #zinb - conditional

icc(glmmTMB(ORF_Correct_12mth~ORF_Correct+School_Size+Condition*FSM+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))
```

```
#==================================#

#Oral reading fluency - Errors

#two-level glmm (poisson)
HERiSS_m_ORF_Errors0<-glmer(ORF_Errors_12mth~ORF_Errors+Condition+School_Size+(1|School_ID),data=HERiSS_data,family=poisson)

#three-level glmm (poisson)
#HERiSS_m_ORF_Errors0<-glmer(ORF_Errors_12mth~ORF_Errors+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data,family=poisson)

# check for zero-inflation and overdispersion
check_overdispersion(HERiSS_m_ORF_Errors0)
check_zeroinflation(HERiSS_m_ORF_Errors0)

summary(HERiSS_m_ORF_Errors0)

#if overdispersed, use:

#two-level glmm (zero-inflated negative binomial)
HERiSS_m_ORF_Errors0b<-glmmTMB(ORF_Errors_12mth~ORF_Errors+Condition+School_Size+(1|School_ID),data=HERiSS_data,family="nbinom2",ziformula=~1)
#HERiSS_m_ORF_Errors0b<-glmer.nb(ORF_Errors_12mth~ORF_Errors+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data)

# FSM subgroup analysis

HERiSS_m_ORF_Errors0_FSM_subset<-glmmTMB(ORF_Errors_12mth~ORF_Errors+Condition+School_Size+(1|School_ID),data=HERiSS_data[HERiSS_data$FSM==1,],family="nbinom2",ziformula=~1)
```

```
HERiSS_m_ORF_Errors0_FSM_int<-
glmmTMB(ORF_Errors_12mth~ORF_Errors+Condition*FSM+School_Size+(1|School_ID),data=HERiSS_data,family="nbinom2",ziformula=~1)

with(HERiSS_data,boxplot((ORF_Errors_12mth-ORF_Errors)~Condition))

sjPlot::tab_model(HERiSS_m_ORF_Errors0,HERiSS_m_ORF_Errors0b,HERiSS_m_ORF_Errors0_FSM_subset,HERiSS_m_ORF_Errors0_FSM_int,
        dv.labels = c("Primary ORF analysis (errors)","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"),
        string.pred = "Coeffcient",
        string.se = "S.E",
        string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=FALSE)

sjPlot::tab_model(HERiSS_m_ORF_Errors0,HERiSS_m_ORF_Errors0b,HERiSS_m_ORF_Errors0_FSM_subset,HERiSS_m_ORF_Errors0_FSM_int,
        dv.labels = c("Primary ORF analysis (errors)","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"))

#======================================#
# ICC

#poisson - null model
icc(glmer(ORF_Errors_12mth~1+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)

#poisson - conditional
icc(glmer(ORF_Errors_12mth~ORF_Errors+School_Size+Condition+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)

#zinb - null
icc(glmmTMB(ORF_Errors_12mth~1+(1|School_ID),family="nbinom2",ziformula=~0,data=HERiSS_data))

#zinb - conditional
```

```
icc(glmmTMB(ORF_Errors_12mth~ORF_Correct+School_Size+Condition+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))


    #zinb - conditional

icc(glmmTMB(ORF_Errors_12mth~ORF_Correct+School_Size+Condition*FSM+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))



    #=================================#


    #Word reading fluency


    #two-level glmm (poisson)

    HERiSS_m_WRF0<-
glmer(WRF_12mth~WRF+Condition+School_Size+(1|School_ID),data=HERiSS_data,family=poisson)


    #three-level glmm (poisson)

    #HERiSS_m_WRF0<-
glmer(WRF_12mth~WRF+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data,family=poisson)


    #check for zero-inflation and overdispersion

    check_overdispersion(HERiSS_m_WRF0)

    check_zeroinflation(HERiSS_m_WRF0)


    summary(HERiSS_m_WRF0)


    #if overdispersed, use:


    #two-level glmm (zero-inflated negative binomial)

    HERiSS_m_WRF0b<-
glmmTMB(WRF_12mth~WRF+Condition+School_Size+(1|School_ID),data=HERiSS_data,family="nbinom2",ziformula=~1)

    #HERiSS_m_WRF0b<-
glmer.nb(WRF_12mth~WRF+Condition+School_Size+(1|School_ID/School_Year),data=HERiSS_data)
```

```
# FSM subgroup analysis

HERiSS_m_WRF0_FSM_subset<-glmmTMB(WRF_12mth~WRF+Condition+School_Size+(1|School_ID),data=HERiSS_data[HERiSS_data$FSM==1,],family="nbinom2",ziformula=~1)

HERiSS_m_WRF0_FSM_int<-glmmTMB(WRF_12mth~WRF+Condition*FSM+School_Size+(1|School_ID),data=HERiSS_data,family="nbinom2",ziformula=~1)

sjPlot::tab_model(HERiSS_m_WRF0,HERiSS_m_WRF0b,HERiSS_m_WRF0_FSM_subset,HERiSS_m_WRF0_FSM_int,
        dv.labels = c("Primary WRF analysis","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"),
        string.pred = "Coeffcient",
        string.se = "S.E",
        string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=FALSE)

sjPlot::tab_model(HERiSS_m_WRF0,HERiSS_m_WRF0b,HERiSS_m_WRF0_FSM_subset,HERiSS_m_WRF0_FSM_int,
        dv.labels = c("Primary WRF analysis","Overdispersed / zeroinflated","FSM individuals only","FSM interaction"))

#========================================#
# ICC

#poisson - null model
icc(glmer(WRF_12mth~1+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)

#poisson - conditional
icc(glmer(WRF_12mth~WRF+School_Size+Condition+(1|School_ID),data=HERiSS_data,family=poisson),ci=0.95)

#zinb - null
```

```
icc(glmmTMB(WRF_12mth~1+(1|School_ID),family="nbinom2",ziformula=~0,data=HERiSS_data))

#zinb - conditional

icc(glmmTMB(WRF_12mth~WRF+School_Size+Condition+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))

#zinb - condition (fsm only)

icc(glmmTMB(WRF_12mth~WRF+School_Size+Condition+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data[HERiSS_data$FSM=="Yes",]))

#zinb - conditional

icc(glmmTMB(WRF_12mth~WRF+School_Size+Condition*FSM+(1|School_ID),family="nbinom2",ziformula=~1,data=HERiSS_data))

#==================================#

#Reading self concept

#two-level lmm

HERiSS_m_RSCS0<-lmer(RSCS_total~RSCS_total_BL+Condition+School_Size+(1|School_ID),data=HERiSS_data)

HERiSS_m_RSCS02<-rlmer(RSCS_total~RSCS_total_BL+Condition+School_Size+(1|School_ID),data=HERiSS_data)

summary(HERiSS_m_RSCS0)

# FSM subgroup analysis

library(robustlm)
```

```
HERiSS_m_RSCS0_FSM_subset<-
rlm(RSCS_total~RSCS_total_BL+Condition+School_Size,data=HERiSS_data[HERiSS_data$FSM==1,])


HERiSS_m_RSCS0_FSM_int<-
rlmer(RSCS_total~RSCS_total_BL+Condition*FSM+School_Size+(1|School_ID),data=HERiSS_data)


sjPlot::tab_model(HERiSS_m_RSCS0,HERiSS_m_RSCS02,HERiSS_m_RSCS0_FSM_subset,HERiSS_m_RSCS0_
FSM_int,
        dv.labels = c("Primary RSCS analysis","Primary RSCS (robust)","FSM individuals only","FSM interaction
(robust)"),
        string.pred = "Coefficient",
        string.se = "S.E",
        string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=TRUE)


sjPlot::tab_model(HERiSS_m_RSCS0,HERiSS_m_RSCS02,HERiSS_m_RSCS0_FSM_subset,HERiSS_m_RSCS0_
FSM_int,
        dv.labels = c("Primary RSCS analysis","Primary RSCS (robust)","FSM individuals only","FSM interaction
(robust)"))


HERiSS_data2<-as.data.frame(HERiSS_data)

HERiSS_data2$School_ID <- as.numeric(as.factor(HERiSS_data2$School_ID))


output2 <-
crtFREQ(RSCS_total~RSCS_total_BL+School_Size+Condition,random="School_ID",intervention="Condition",data=H
ERiSS_data2)


summary(output2)


#==============================================#


#ICC


icc(lmer(RSCS_total~RSCS_total_BL+Condition+School_Size+(1|School_ID),data=HERiSS_data),ci=0.95)
```

```
#hedges g (robust model) -

# values from model inputed into formula: g = estimate for fixed effect / (sqrt of sum of variances of random effects)

round( 0.21253/sqrt(0.2041+2.4061),2)
round(( 0.21253+1.96*0.05690)/sqrt(0.2041+2.4061),2)
round(( 0.21253-1.96*0.05690)/sqrt(0.2041+2.4061),2)


######### get p values

coefs <- data.frame(coef(summary(HERiSS_m_RSCS0)))

# get coefficients from robust model to extract t-values
coefs.robust <- coef(summary(HERiSS_m_RSCS02))

# calculate p-values based on robust t-values and non-robust approx. DFs
p.values <- 2*pt(abs(coefs.robust[,3]), coefs$df, lower=FALSE)
p.values["ConditionIntervention"]


##EXTRA RSCS

HERiSS_long2<-HERiSS_data %>%
dplyr::select(Pupil_ID,School_ID,Condition,School_Size,RSCS_total,RSCS_total_BL) %>%
  pivot_longer(!c(Pupil_ID,School_ID,Condition,School_Size),names_to='Time',values_to='RSCS')

HERiSS_long2$Time <-factor(HERiSS_long2$Time,levels=c("RSCS_total_BL",
"RSCS_total"),labels=c("Baseline","Post test"))
```

```
ggplot(HERiSS_long2,aes(y=RSCS,x=Time,colour=Condition))+geom_line(aes(group = interaction(Pupil_ID, Condition)),alpha=0.3)+ geom_smooth(aes(group=Condition),se=FALSE,size=1.5)+theme_bw()

#updated to incorporate reviewer suggestions:

ggplot(HERiSS_long2,aes(y=RSCS,x=Time,colour=Condition))+
geom_smooth(aes(group=Condition),se=FALSE,size=1.5)+theme_bw()+ylim(0,12.5)

#===================================================================================
===#

#Alternative model

HERiSS_prim_extra2 <-
lmer(RSCS~School_Size+Time*Condition+(1|School_ID)+(1|Pupil_ID),data=HERiSS_long2)

sjPlot::tab_model(HERiSS_prim_extra2,pred.labels = c("Intercept", "School Size", "Time", "Intervention", "Time x Intervention"),
          dv.labels = c("Supplementary analysis(two-level)"),
          string.pred = "Coefficient",
          string.se = "S.E",
          string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=TRUE)

#===================================================================================
==========#

#===================================================================================
==========#

#Instrumental variables approach

#===================================#
```

```
#In this approach, we do not use linear mixed effects models, instead reporting cluster robust SEs.

compliance_data <- read_excel("HERiSS_compliance_PT_OCT2022.xlsx")

HERiSS_compliance<-left_join(HERiSS_data, compliance_data,by="School_ID") %>% as.data.frame()

HERiSS_compliance <- HERiSS_compliance %>% mutate_at(vars(full_comp:Partial_B), ~replace(., is.na(.), 0))


#Full compliance

m_iv1 <- ivreg(DIBELS_Comp ~ DIBELS_Comp_BL+School_Size | Condition | full_comp, data = HERiSS_compliance)

summary(m_iv1,diagnostics = TRUE)

#Partial compliance A

m_iv2 <- ivreg(DIBELS_Comp ~ DIBELS_Comp_BL+School_Size | Condition | Partial_A, data = HERiSS_compliance)

summary(m_iv2,diagnostics = TRUE)

#Partial compliance B

m_iv3 <- ivreg(DIBELS_Comp ~ DIBELS_Comp_BL+School_Size | Condition |Partial_B, data = HERiSS_compliance)

summary(m_iv3,diagnostics = TRUE)


#results table.
```

```
sjPlot::tab_model(HERiSS_m0a, m_iv1, m_iv2, m_iv3,
        dv.labels = c("Primary", "IV full","IV part A","IV part B"),
        string.pred = "Coefficient",
        string.se = "S.E",
        string.p = "P-Value", p.style = "stars",show.se=TRUE,show.ci=FALSE)


sjPlot::tab_model(HERiSS_m0a, m_iv1, m_iv2, m_iv3,
        dv.labels = c("Primary", "IV full","IV part A","IV part B"))


#hedges g (robust model) - partial B compliance

# values from model inputed into formula: d = estimate for effect / SE*SQRT(N)
dim(model.frame(m_iv3))# first number gives the N in the model

round(2.08777/(sqrt(276)*1.69362),2)
round((2.08777+1.96*1.69362)/(sqrt(276)*1.69362),2)
round((2.08777-1.96*1.69362)/(sqrt(276)*1.69362),2)

#reviewer comment, use simple hedges g adjustment (Borenstein, Hedges, Higgins, & Rothstein, 2011,p27)

d<-round(2.08777/(sqrt(276)*1.69362),2)
df<- 276

g<-d∗(1−(3/(4*df)−1))


d_up<-round((2.08777+1.96*1.69362)/(sqrt(276)*1.69362),2)
d_low<-round((2.08777-1.96*1.69362)/(sqrt(276)*1.69362),2)

g_up<-d_up∗(1−(3/(4*df)−1))
```

```
g_low<-d_low∗(1−(3/(4*df)−1))


g_up

g_low


#hedges g (robust model) - partial A compliance


# values from model inputed into formula:d = estimate for effect / SE*SQRT(N)

dim(model.frame(m_iv2))# first number gives the N in the model


round(0.48098/(sqrt(276)*2.30828),2)

round((0.48098+1.96*2.30828)/(sqrt(276)*2.30828),2)

round((0.48098-1.96*2.30828)/(sqrt(276)*2.30828),2)


#reviewer comment, use simple hedges g adjustment (Borenstein, Hedges, Higgins, & Rothstein, 2011,p27)


d<-round(0.48098/(sqrt(276)*2.30828),2)

df<- 276


g<-d∗(1−(3/(4*df)−1))


d_up<-round((0.48098+1.96*2.30828)/(sqrt(276)*2.30828),2)

d_low<-round((0.48098-1.96*2.30828)/(sqrt(276)*2.30828),2)


g_up<-d_up∗(1−(3/(4*df)−1))

g_low<-d_low∗(1−(3/(4*df)−1))


g_up

g_low

g
```

```
#hedges g (robust model) - full compliance

# values from model inputed into formula: d = estimate for effect / SE*SQRT(N)
dim(model.frame(m_iv1))# first number gives the N in the model

round(1.06574/(sqrt(276)*2.59235),2)
round((1.06574+1.96*2.59235)/(sqrt(276)*2.59235),2)
round((1.06574-1.96*2.59235)/(sqrt(276)*2.59235),2)

d<-round(1.06574/(sqrt(276)*2.59235),2)
df<- 276

g<-d∗(1−(3/(4*df)−1))


d_up<-round((1.06574+1.96*2.59235)/(sqrt(276)*2.59235),2)
d_low<-round((1.06574-1.96*2.59235)/(sqrt(276)*2.59235),2)

g_up<-d_up∗(1−(3/(4*df)−1))
g_low<-d_low∗(1−(3/(4*df)−1))

g_up
g_low
g


#================================================================================
==============#
# Impute missing data

#================================================================================
==============#
# PT = 27/09/2022
```

```
# modelling the binary response of missingness in the primary outcome

HERiSS_data_bin<-HERiSS_data

#create a binary outcome variable (missing or not on the DIBELS Comp)
HERiSS_data_bin$DIBELS_Comp_miss <- ifelse(is.na(HERiSS_data_bin$DIBELS_Comp)==TRUE,1,0)

#Fit a logistic regression with covariates (as per SAP)
DIBELS_miss_mod<-glm(DIBELS_Comp_miss ~
DIBELS_Comp_BL+FSM+SEND_SpLD+SEND_MLD+SEND_SLD+SEND_PMLD+SEND_SLCN+SEND_ASD+SEND
_Other+SEND_PD+SEND_SEMH+EAL,data=HERiSS_data_bin, family = "binomial")

DIBELS_miss_mod<-glm(DIBELS_Comp_miss ~
DIBELS_Comp_BL+FSM+SEND_SpLD+SEND_MLD+SEND_SLD+SEND_SLCN+SEND_ASD+SEND_Other+SEND
_PD+EAL,data=HERiSS_data_bin, family = "binomial")

summary(DIBELS_miss_mod)

sjPlot::tab_model(DIBELS_miss_mod)

#=========================================================#

set.seed(2021) # for reproducibility

#select variables for imputation data
HERiSS_imp_sub <- HERiSS_data %>% dplyr::select(School_ID, Pupil_ID, School_Year, FSM, EAL, School_Size,
Condition,
                                   LNF, PSF, NWF_CLS, NWF_WRC, WRF, ORF_Correct, ORF_Errors,
```

LNF_12mth, PSF_12mth, NWF_CLS_12mth, NWF_WRC_12mth, WRF_12mth, ORF_Correct_12mth, ORF_Errors_12mth)

#rename to comply with imputation functions

HERiSS_imp_sub <- HERiSS_imp_sub %>% rename(.,"SchoolID"="School_ID", "PupilID"="Pupil_ID", "SchoolYear"="School_Year", "SchoolSize"="School_Size",

"NWFCLS"="NWF_CLS","NWFWRC"="NWF_WRC","ORFCorrect"="ORF_Correct","ORFErrors"="ORF_Errors", "LNF12mth"="LNF_12mth", "PSF12mth"="PSF_12mth",

"NWFCLS12mth"="NWF_CLS_12mth","NWFWRC12mth"="NWF_WRC_12mth","WRF12mth"="WRF_12mth","ORFCorrect12mth"="ORF_Correct_12mth",

"ORFErrors12mth"="ORF_Errors_12mth")

predictor matrix and imputation method (defaults)

predMatrix <- make.predictorMatrix(data = HERiSS_imp_sub)

impMethod <- make.method(data = HERiSS_imp_sub)

#==#

#Set up the imputation method for each variable.

method for lower-level variables (x, y, and z)

impMethod[c("LNF", "PSF", "WRF", "ORFCorrect", "ORFErrors","LNF12mth", "PSF12mth", "WRF12mth", "ORFCorrect12mth", "ORFErrors12mth")] <- "nb" #"gamlssPO"

impMethod[c("NWFCLS", "NWFWRC", "NWFCLS12mth", "NWFWRC12mth")] <- "zinb" #"gamlssPO"

method for variables at top level (w)

impMethod[c("SchoolSize","Condition","EAL","FSM")] <- "pmm"

impMethod[c("SchoolYear")] <- "pmm"

remove indicator variables from predictor matrix

predMatrix[, "SchoolID"] <- 0

predMatrix[, "PupilID"] <- 0

predMatrix[, "Condition"] <- 0

#==#

#set which variables are used in the imputation. We use any other DIBELS component measure as the same time point and the SEND ASD. Exception is if correlations are too high among variables as this indices collinearity in the predictors and the imputation models fail.

predMatrix["SchoolYear",] <- c(0,0, 0,1,1,1, 0, 1,1,1,1,1,1,1, 0,0,0,0,0,0,0)

predMatrix["FSM",] <- c(0,0, 1,0,1,1, 0, 1,1,1,1,1,1,1, 0,0,0,0,0,0,0)

predMatrix["EAL",] <- c(0,0, 1,1,0,1, 0, 1,1,1,1,1,1,1, 0,0,0,0,0,0,0)

predMatrix["SchoolSize",] <- c(0,0, 1,1,1,1, 0, 1,1,1,1,1,1,1, 0,0,0,0,0,0,0)

predMatrix["LNF",] <- c(0,0, 1,1,1,1, 0, 0,1,1,1,1,0,0, 0,0,0,0,0,0,0)

predMatrix["PSF",] <- c(0,0, 1,1,1,1, 0, 1,0,1,1,1,0,0, 0,0,0,0,0,0,0)

predMatrix["NWFCLS",] <- c(0,0, 1,1,1,1, 0, 1,1,0,1,1,0,0, 0,0,0,0,0,0,0)

predMatrix["NWFWRC",] <- c(0,0, 1,1,1,1, 0, 1,1,1,0,1,0,0, 0,0,0,0,0,0,0)

predMatrix["WRF",] <- c(0,0, 1,1,1,1, 0, 1,1,1,1,0,0,0, 0,0,0,0,0,0,0)

predMatrix["ORFCorrect",] <- c(0,0, 1,1,1,1, 0, 1,1,0,1,0,1,0, 0,0,0,0,0,0,0)

predMatrix["ORFErrors",] <- c(0,0, 1,1,1,1, 0, 1,1,0,1,1,0,0, 0,0,0,0,0,0,0)

predMatrix["LNF12mth",] <- c(0,0, 1,1,1,1, 0, 0,0,0,0,0,0,0, 0,1,1,1,1,0,0)

predMatrix["PSF12mth",] <- c(0,0, 1,1,1,1, 0, 0,0,0,0,0,0,0, 1,0,1,1,1,0,0)

predMatrix["NWFCLS12mth",] <- c(0,0, 1,1,1,1, 0, 0,0,0,0,0,0,0, 1,1,0,1,1,0,0)

predMatrix["NWFWRC12mth",] <- c(0,0, 1,1,1,1, 0, 0,0,0,0,0,0,0, 1,1,1,0,1,0,0)

predMatrix["WRF12mth",] <- c(0,0, 1,1,1,1, 0, 0,0,0,0,0,0,0, 1,1,1,1,0,0,0)

predMatrix["ORFCorrect12mth",] <- c(0,0, 1,1,1,1, 0, 0,0,0,0,0,0,0, 0,0,0,1,1,0,0)

predMatrix["ORFErrors12mth",] <- c(0,0, 1,1,1,1, 0, 0,0,0,0,0,0,0, 1,1,0,1,1,0,0)

===

#RUN THE IMPUTATION MODELs (20 imputed datasets generated and pooled)

imp_new <- countimp(HERiSS_imp_sub, method = impMethod, predictorMatrix = predMatrix, maxit = 100, m = 20,print=FALSE)

```
plot(imp_new) # add as a appendix to satisfy reviewer comment regarding convergence.

# ========================================================================== #


# Density plot to check the approximate imputations follow the same distribution as the original data.


imp2new <- complete(imp_new,"long", include=TRUE)

library(patchwork)


# Add a variable for the plot legend

imp2new$Imputed<-ifelse(imp2new$".imp"==0,"Observed","Imputed")


# Plot. Be sure to use stat_density instead of geom_density in order

#  to prevent what you call "unwanted horizontal and vertical lines"

p1 <- ggplot(imp2new, aes(x=LNF, group=.imp, colour=Imputed)) +

  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("LNF")


p2 <- ggplot(imp2new, aes(x=PSF, group=.imp, colour=Imputed)) +

  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("PSF")


p3 <- ggplot(imp2new, aes(x=NWFCLS, group=.imp, colour=Imputed)) +

  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("NWF-CLS")


p4 <- ggplot(imp2new, aes(x=NWFWRC, group=.imp, colour=Imputed)) +

  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("NWF-WRC")


p5 <- ggplot(imp2new, aes(x=WRF, group=.imp, colour=Imputed)) +
```

```
  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("WRF")


 p6 <- ggplot(imp2new, aes(x=ORFCorrect, group=.imp, colour=Imputed)) +

  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("ORF Correct")


 p7 <- ggplot(imp2new, aes(x=ORFErrors, group=.imp, colour=Imputed)) +

  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("ORF_Errors")


 p8 <- ggplot(imp2new, aes(x=LNF12mth, group=.imp, colour=Imputed)) +

  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("LNF 12mths")


 p9 <- ggplot(imp2new, aes(x=PSF12mth, group=.imp, colour=Imputed)) +

  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("PSF 12mth")


 p10 <- ggplot(imp2new, aes(x=NWFCLS12mth, group=.imp, colour=Imputed)) +

  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("NWF-CLS 12mth")


 p11 <- ggplot(imp2new, aes(x=NWFWRC12mth, group=.imp, colour=Imputed)) +

  stat_density(geom = "path",position = "identity")+

  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("NWF-WRC 12mth")


 p12 <- ggplot(imp2new, aes(x=WRF12mth, group=.imp, colour=Imputed)) +
```

```
  stat_density(geom = "path",position = "identity")+
  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("WRF 12mth")

p13 <- ggplot(imp2new, aes(x=ORFCorrect12mth, group=.imp, colour=Imputed)) +
  stat_density(geom = "path",position = "identity")+
  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("ORF Correct 12mth")

p14 <- ggplot(imp2new, aes(x=ORFErrors12mth, group=.imp, colour=Imputed)) +
  stat_density(geom = "path",position = "identity")+
  stat_density(data=imp2new[imp2new$".imp"==0,],geom = "path",position = "identity",colour="blue") + theme_bw()
+ggtitle("ORF_Errors 12mth")

p1 + p2 + p3 + p4 + p5 + p6 + p7 + plot_layout(guides = 'collect')

p8 + p9 + p10 + p11 + p12 + p13 + p14 + plot_layout(guides = 'collect')

# ======================================================================= #

# create list of imputed data sets (wrangles data into a workable format)
implist2 <- mids2mitml.list(imp_new)

########
#We only impute the component measures then calculate the composite scores, rather than impute the composite imputations directly for better accuracy.

#add composite calculations for each dataset using purrr:map#

implist2 <- imap(implist2, ~ .x %>% mutate(ORFAccuracy = ifelse(ORFCorrect==0,0,((ORFCorrect/(ORFErrors+ORFCorrect)*100)))))
```

```
implist2 <- imap(implist2, ~ .x %>% mutate(ORFAccuracy12mth = ifelse(ORFCorrect12mth==0,0,((ORFCorrect12mth/(ORFErrors12mth+ORFCorrect12mth)*100)))))

implist2<-imap(implist2, ~ .x %>% mutate(DIBELS_Comp_BL = (((((ifelse(is.na(LNF),0,LNF)*10.72)+(ifelse(is.na(PSF),0,PSF)*2.13)+(ifelse(is.na(NWFCLS),0,NWFCLS)*23.13)+(ifelse(is.na(NWFWRC),0,NWFWRC)*7.79)+(ifelse(is.na(WRF),0,WRF)*13.51)+(ifelse(is.na(ORFCorrect),0,ORFCorrect)*25.36)+(ifelse(is.na(ORFAccuracy),0,ORFAccuracy)*0.25))-3371)/2251)*40)+ 440))

implist2<-imap(implist2, ~ .x %>% mutate(DIBELS_Comp = (((((ifelse(is.na(LNF12mth),0,LNF12mth)*10.72)+(ifelse(is.na(PSF12mth),0,PSF12mth)*2.13)+(ifelse(is.na(NWFCLS12mth),0,NWFCLS12mth)*23.13)+(ifelse(is.na(NWFWRC12mth),0,NWFWRC12mth)*7.79)+(ifelse(is.na(WRF12mth),0,WRF12mth)*13.51)+(ifelse(is.na(ORFCorrect12mth),0,ORFCorrect12mth)*25.36)+(ifelse(is.na(ORFAccuracy12mth),0,ORFAccuracy12mth)*0.25))-3371)/2251)*40)+ 440))

########

# fit model and pool results.

pooled <- implist2 %>%

  Map(f = lmer, MoreArgs = list(f=DIBELS_Comp~DIBELS_Comp_BL+SchoolSize+Condition+(1|SchoolID))) %>%

  testEstimates(.,extra.pars=TRUE)

confint(pooled)

#Manually combine results into table for report as pooled output is not compliant with SJPlot package.

#==================================================================================
=#

#Hedges g for pooled results

round((-1.054)/sqrt(124.870+15.697),2)

round((-1.054+1.96*1.661)/sqrt(124.870+15.697),2)

round((-1.054-1.96*1.661)/sqrt(8.993+151.927),2)
```

```
#confidence intervals for pooled results.

confint(pooled)

#=================================================================================
=#


#08/11/2022 - demographic summaries


HERiSS_data %>% count(Condition)


HERiSS_data %>% count(Condition,School_Size)

HERiSS_data %>% count(Condition,FSM)

HERiSS_data %>% count(Condition,EAL)

HERiSS_data %>% count(Condition,SEND_ASD)

HERiSS_data %>% count(Condition,SEND_MLD)

HERiSS_data %>% count(Condition,SEND_SLD)

HERiSS_data %>% count(Condition,SEND_PMLD)

HERiSS_data %>% count(Condition,SEND_SLCN)

HERiSS_data %>% count(Condition,SEND_SpLD)

HERiSS_data %>% count(Condition,SEND_Other)

HERiSS_data %>% count(Condition,SEND_PD)

HERiSS_data %>% count(Condition,SEND_SEMH)

HERiSS_data %>% count(Condition,is.na(DIBELS_Comp_BL))

HERiSS_data %>% count(Condition,is.na(RSCS_total_BL))



HERiSS_data %>% count(Condition,is.na(DIBELS_Comp))


#% missing


#intervention
```

```
(66/181)*100

#control

(28/201)*100


# DIBELS_comp_BL summaries

#hedges g
round(fixef(HERiSS_m0a)["DIBELS_Comp_BL"]/sqrt(sum(as.data.frame(VarCorr(HERiSS_m0a))$vcov)),3)

#baseline characteristics
HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(DIBELS_Comp_BL,na.rm=T),sd=sd(DIBELS_Comp_BL,na.rm=T))


#==========================================================================#

# LNF summaries


#baseline characteristics
HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(LNF,na.rm=T),sd=sd(LNF,na.rm=T))
HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(PSF,na.rm=T),sd=sd(PSF,na.rm=T))
HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(NWF_CLS,na.rm=T),sd=sd(NWF_CLS,na.rm=T))
HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(NWF_WRC,na.rm=T),sd=sd(NWF_WRC,na.rm=T))
HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(WRF,na.rm=T),sd=sd(WRF,na.rm=T))
HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(ORF_Correct,na.rm=T),sd=sd(ORF_Correct,na.rm=T))
```

```
HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(ORF_Errors,na.rm=T),sd=sd(ORF_Errors,na.rm=T))

HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(ORF_Accuracy,na.rm=T),sd=sd(ORF_Accuracy,na.rm=T))

HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(RSCS_total_BL,na.rm=T),sd=sd(RSCS_total_BL,na.rm=T))

HERiSS_data %>% group_by(Condition) %>% summarise(mean=mean(DIBELS_Comp,na.rm=T),sd=sd(ORF_Accuracy,na.rm=T))



#post test characteristics

library(emmeans)

emm_options(opt.digits = FALSE)

#

emmeans(HERiSS_m0a, specs = pairwise ~ Condition)

emmeans(HERiSS_m_LNF0, specs = pairwise ~ Condition)

emmeans(HERiSS_m_PSF0, specs = pairwise ~ Condition)

emmeans(HERiSS_m_NWF_CLS0, specs = pairwise ~ Condition)

emmeans(HERiSS_m_NWF_WRC0, specs = pairwise ~ Condition)

emmeans(HERiSS_m_ORF0, specs = pairwise ~ Condition)

emmeans(HERiSS_m_ORF_Errors0, specs = pairwise ~ Condition)

emmeans(HERiSS_m_RSCS0, specs = pairwise ~ Condition)



HERiSS_data %>% count(Condition,is.na(LNF_12mth))

HERiSS_data %>% count(Condition,is.na(PSF_12mth))

HERiSS_data %>% count(Condition,is.na(NWF_CLS_12mth))

HERiSS_data %>% count(Condition,is.na(NWF_WRC_12mth))

HERiSS_data %>% count(Condition,is.na(ORF_Correct_12mth))

HERiSS_data %>% count(Condition,is.na(ORF_Errors_12mth))

HERiSS_data %>% count(Condition,is.na(WRF_12mth))

HERiSS_data %>% count(Condition,is.na(RSCS_total))
```

```
#====================================================================================#

# histogram of baseline variables.

library(patchwork)

#DIBELS baseline

#ggplot(HERiSS_data,aes(DIBELS_Comp_BL))+geom_histogram(colour='darkgreen',fill="green",alpha=0.7)+theme_bw()+xlab("DIBELS composite")

p1=ggplot(HERiSS_data,aes(DIBELS_Comp_BL))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS composite")

#LNF

p2=ggplot(HERiSS_data,aes(LNF))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS LNF")

#PSF

p3=ggplot(HERiSS_data,aes(PSF))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS PSF")

#NWF_CLS

p4=ggplot(HERiSS_data,aes(NWF_CLS))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS NWF-CLS")

#NWF_WRC

p5=ggplot(HERiSS_data,aes(NWF_WRC))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS NWF_WRC")

#WRF

p6=ggplot(HERiSS_data,aes(WRF))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS WRF")
```

```
#ORF_Correct

p7=ggplot(HERiSS_data,aes(ORF_Correct))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS ORF_Correct")

#ORF_Errors

p8=ggplot(HERiSS_data,aes(ORF_Errors))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS ORF_Errors")

#ORF_Accuracy

p9=ggplot(HERiSS_data,aes(ORF_Accuracy))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS ORF_Accuracy")

#RSCS

p10=ggplot(HERiSS_data,aes(RSCS_total_BL))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("Reading Self Concept")

p1+p2+p3+p4+p5+p6+p7+p8+p9+p10+ guide_area()+ plot_layout(ncol = 3, guides = 'collect')

#================================================================================#

# histogram of outcome variables.

library(patchwork)

#DIBELS baseline

#ggplot(HERiSS_data,aes(DIBELS_Comp_BL))+geom_histogram(colour='darkgreen',fill="green",alpha=0.7)+theme_bw()+xlab("DIBELS composite")

q1=ggplot(HERiSS_data,aes(DIBELS_Comp))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS composite (12 mth)")
```

```
#LNF

q2=ggplot(HERiSS_data,aes(LNF_12mth))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS LNF (12 mth)")

#PSF

q3=ggplot(HERiSS_data,aes(PSF_12mth))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS PSF (12 mth)")

#NWF_CLS

q4=ggplot(HERiSS_data,aes(NWF_CLS_12mth))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS NWF-CLS (12 mth)")

#NWF_WRC

q5=ggplot(HERiSS_data,aes(NWF_WRC_12mth))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS NWF_WRC (12 mth)")

#WRF

q6=ggplot(HERiSS_data,aes(WRF_12mth))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS WRF (12 mth)")

#ORF_Correct

q7=ggplot(HERiSS_data,aes(ORF_Correct_12mth))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS ORF_Correct (12 mth)")

#ORF_Errors

q8=ggplot(HERiSS_data,aes(ORF_Errors_12mth))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("DIBELS ORF_Errors (12 mth)")

#RSCS
```

```
q9=ggplot(HERiSS_data,aes(RSCS_total))+geom_histogram(aes(colour=Condition,fill=Condition),alpha=0.7)+theme_bw()+xlab("Reading Self Concept (12 mth)")

q1+theme(legend.position = "bottom")

q2+q3+q4+q5+q6+q7+q8+q9+guide_area()+ plot_layout(ncol = 3, guides = 'collect')

##################################################

#post test data - consort figures

HERiSS_PT_data2 <- left_join(HERiSS_PT_data,HERiSS_data[,c("Pupil_ID","Condition","School_ID","School_Size")],by="Pupil_ID")

HERiSS_PT_data2 %>% count(Condition)

#N schools in each arm
length(unique(HERiSS_PT_data2$School_ID[HERiSS_PT_data2$Condition=="Control"]))

length(unique(HERiSS_PT_data2$School_ID[HERiSS_PT_data2$Condition=="Intervention"]))

#n schools by size in each arm
length(unique(HERiSS_PT_data2$School_ID[HERiSS_PT_data2$Condition=="Control"&HERiSS_PT_data2$School_Size=="Small"]))

length(unique(HERiSS_PT_data2$School_ID[HERiSS_PT_data2$Condition=="Control"&HERiSS_PT_data2$School_Size=="Large"]))
```

```
length(unique(HERiSS_PT_data2$School_ID[HERiSS_PT_data2$Condition=="Intervention"&HERiSS_PT_data2$School_Size=="Small"]))

length(unique(HERiSS_PT_data2$School_ID[HERiSS_PT_data2$Condition=="Intervention"&HERiSS_PT_data2$School_Size=="Large"]))


##################################################


#Attrition numbers (numbers analysed)


HERiSS_data %>% filter(is.na(DIBELS_Comp_BL)==FALSE) %>% filter(is.na(DIBELS_Comp)==FALSE) %>%count(Condition,FSM)


test<-HERiSS_data %>% filter(is.na(DIBELS_Comp_BL)==FALSE) %>% filter(is.na(DIBELS_Comp)==FALSE)

length(unique(test$School_ID))

#46

length(unique(test$School_ID[test$Condition=="Intervention"]))

#20


#therefore 26 control by default.
```

Education Endowment Foundation

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn