



Evaluating a computer-based reading programme with children with Intellectual Disabilities: feasibility and pilot research

Roberts-Tyler, Emily; Hughes, John; Hastings, Richard

### Journal of Research in Special Educational Needs

DOI:

[10.1111/1471-3802.12458](https://doi.org/10.1111/1471-3802.12458)

Published: 01/01/2020

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

*Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):*

Roberts-Tyler, E., Hughes, J., & Hastings, R. (2020). 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>

#### Hawliau Cyffredinol / General rights

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

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

#### Take down policy

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

Running head: Feasibility of reading programme evaluation in ID

Evaluating a computer-based reading programme with children with Intellectual  
Disabilities: Feasibility and pilot research.

Emily J. Roberts-Tyler<sup>1,2</sup>, J. Carl Hughes<sup>1,2</sup> and Richard P. Hastings<sup>3</sup>

<sup>1</sup>School of Education and Human Development, Bangor University, Wales

<sup>2</sup>Collaborative Institute for Education Research, Evidence and Impact (CIEREI),  
Bangor University, Wales

<sup>3</sup>Centre for Educational Development Appraisal and Research, University of Warwick

**Emily Jehanne Roberts-Tyler:**

ORCID ID: 0000-0001-6055-0475

Twitter handle: @DrEmilyJT Tyler

**John Carl Hughes:**

ORCID ID: 0000-0001-7130-7286

Twitter handle: @ProfJCarlHughes

**Richard Patrick Hastings:** ORCID ID: 0000-0002-0495-8270

Twitter handle: @ProfRHastings

\*Corresponding Author: Emily J Roberts-Tyler, Bangor University, School of  
Education and Human Development, Normal Site, Bangor, Gwynedd, Wales, LL57  
2PZ (e-mail: e.j.tyler@bangor.ac.uk).



## **Introduction**

### **Reading instruction in ID**

Many children with Intellectual Disabilities (ID) struggle to learn to read. For example, in the USA, 67% of children with ID have considerable difficulty learning basic reading skills (ranging from early decoding skills, through to demonstrating an understanding of what they read and making inferences; NAEP, US Department of Education, 2007). Despite these difficulties internationally, information and guidelines regarding teaching reading or other academic skills for children with ID are scarce, and often inadequate (Marks, 2000; Wehmeyer, 2006).

As has been found for typically developing (TD) children, increasing evidence indicates individuals with ID might benefit from phonics-based instruction (NRP, 2000; Joseph & Seery, 2004; Whalon, Otaiba, & Delano, 2009). However, research and instruction in the ID field has predominantly focused on sight word reading (Katims, 2000) and has less frequently investigated phonics instruction (Browder, Wakeman, Spooner, Ahlgrim-Delzell, & Aldozzine, 2006; Joseph & Seery, 2004). Therefore, further research is required to investigate the effects of phonics-based programmes, and programmes incorporating evidence-based instructional components, on the reading skills of children with ID (Browder, Gibbs, Ahlgrim-Delzell, Courtade, Mraz, & Flowers, 2009; Whalon, Otaiba & Delano, 2009).

Research indicates children with ID may have less access than typically developing peers to literacy activities at home (Fitzgerald, Roberts, Pierce & Schuele, 1995), and to reading instruction in school (Kliewer & Landis, 1999). Well-designed computer-assisted instruction can potentially provide many more practice and response opportunities than teacher-delivered instruction, as well as enabling more independent practice. This efficient use of instructional time may be especially

significant for children with ID who will likely require more input to develop reading skills (Browder & Spooner, 2006). Computer-assisted reading programmes are increasingly used to supplement reading instruction (Andrews, 2004). Although the quality of, and supporting evidence for, such programmes is somewhat variable, there is evidence they can have a positive effect on reading skills (e.g. Blok, Oostdam, Otter & Overmaat, 2002; Moran, Ferdig, Pearson, Wardrop & Blomeyer, 2008; Macaruso, Hook & McCabe, 2006; Torgesen & Zhu, 2003; NRP, 2000). Further, some research suggests computer-assisted reading programmes can help improve reading skills specifically for children with autism (e.g. Grindle, Hughes, Saville, Huxley, & Hastings, 2013; Soe et al., 2000) and children with ID (e.g. Coleman-Martin, Heller, Cihak & Irvine, 2005; Jones, Torgesen & Saxton, 1987; Torgesen, Waters, Cohen & Torgesen, 1988).

Headsprout® *Early Reading* is an Internet-based programme designed to teach the skills and strategies necessary for efficient, fluent reading. Comprising 80, 20-minute lessons (episodes), Headsprout® *Early Reading* (HER) is a computer-delivered systematic, synthetic phonics programme that includes instruction in phonemic awareness, print awareness, phonics, sounding out, segmenting and blending, and explicitly incorporates the five components of reading proposed by the NRP (Layng, Twyman, & Stikeleather, 2003). HER is an adaptive learning technology—every mouse-click forms data on individual learners' progress that is used to provide additional instruction or to ensure repeated practice of components not yet fluent. In this way, the instruction is individually adapted to each child's responses. In addition to online episodes, frequency-building exercises accompany HER. There are two tiers of this additional support—*Targeted Practice* and *Intensive Practice* (see Procedure below, or Layng, 2003, for more detail).

Although HER is designed for typically developing children, there is some evidence it can be beneficial for children with ADHD (Clarfield & Stoner, 2005) and autism (Grindle, Hughes, Saville, Huxley, & Hastings, 2013; Whitcomb, Bass, & Luiselli, 2011). Grindle et al., (2013) enrolled 4 children with a diagnosis of autism (aged between 5 and 7 years) in HER. With additional input (e.g. additional Discrete Trial Teaching for areas of difficulty, dividing episodes over 2-3 sittings, and delivery of additional reinforcers to increase motivation), all four children could access the programme. On completing HER, notable improvements were seen in early literacy skills and word recognition across participants. Through a similar series of case studies conducted as previous pilot work, we have demonstrated that children with ID can also access and benefit from HER with minimal or no adaptations (*anonymised et al.*, 2015). There is no published research to date relating to the use of HER with children with other disabilities. However, the current evidence-base indicates a broad range of children with ID and/or autism might benefit, which could be deemed promising for its use with other populations. In particular, given that Computer-assisted phonics instruction has been found to be effective for some students who use Alternative and Augmentative Communication (AAC; Coleman-Martin, Keller, Cihak & Irvine, 2005), HER may provide an opportunity for more students who use AAC to access phonics-based reading instruction more independently.

### **Educational research**

Randomised Control Trials (RCTs) have long been considered the ‘gold standard’ for informing evidence-based practice in medicine and healthcare (Milton, 2007). However, despite educational researchers also advocating RCTs in evaluation research (Oakley, 1998, 2000), their use in educational research has lagged behind healthcare in more recent decades (Torgersen & Torgersen, 2001; Oakley, 2006). As

such, educational policies are often introduced and implemented without sufficient evidence of their efficacy (e.g. National literacy and numeracy strategies; Torgersen & Torgersen, 2001). Although the detrimental effects of administering ineffective interventions in education may not be as pronounced as for life and death outcomes in medicine, it has been suggested that: “the exposure of children to educational harm when initiatives are not properly tested is a very real risk” (Hutchinson & Styles, 2010, p.7).

To design and conduct well-controlled RCTs, it is often necessary to conduct feasibility research to inform and trial aspects of design and methodology for a larger scale study. Models for testing complex interventions (Medical Research Council, 2008; Thornicroft, Lempp & Tansella, 2011) recommend feasibility research be conducted prior to conducting randomised studies to assess efficacy of interventions. The purpose of feasibility studies can be grouped into four general categories: Process (investigating feasibility of necessary steps of a main study, including determining recruitment, retention and adherence/compliance rates), resources (investigating potential time and budget requirements, including time taken to administer assessments and resources related to intervention implementation), management (investigating relevant management issues in participating settings), and scientific (investigating various aspects of the intervention, including estimation of treatment effect; Van Tiejlingen, Rennie, Hundley & Graham, 2001; Van Teijlingen & Hundley, 2001; Thabane et al., 2010). Through investigating these important parameters, the feasibility of conducting a full-scale evaluation can be better understood, and the chances of a subsequent full-scale evaluation being successful is greatly increased (Arain, Campbell, Cooper & Lancaster, 2010; Thabane et al., 2010).

Despite the obvious importance of feasibility and pilot studies, it is an aspect of the research process often neglected in research training (Thabane et al., 2010). Similarly, although feasibility work can potentially inform other researchers of important parameters in a given context, such work is seldom published (Van Teijlingen & Hundley, 2001). The increased dissemination of such studies could serve to reduce unnecessarily duplicating efforts of researchers in similar fields (Thabane et al., 2010).

### **Evaluating HER with children with ID**

Our initial pilot work implementing HER with children with ID (*anonymised et al., 2015*) has served to elucidate some important feasibility questions related to conducting a larger evaluation. Because HER is designed for typically developing children, our initial objectives were to investigate whether HER is accessible or can be made accessible for children with ID. We also explored the use of the additional tiers of support within HER. Table one outlines these initial feasibility objectives.

Regarding the accessibility of HER, we determined HER can be used with children with ID. Some children required additional input, although predominantly only if they did not understand the concept of negation (e.g. “Which of these is not a fish?”), which is crucial to progression beyond HER Episode 4. Furthermore, reading skills did appear to improve following completion of HER (as measured by subtests of the Dynamic Indicator of Basic Early Literacy Skills – Good & Kaminski, 2007 – and the Word Recognition and Phonic Skills Assessment – Carver & Moseley, 1994).

Regarding the implementation of HER with children with ID, we found that the episode data (collected online by HER) of most children with mild-moderate ID who were enrolled demonstrated similar performance (in terms of percentage scores and time taken per episode) to TD children (*anonymised et al, 2015*). This suggests

they did not necessarily require the level of additional support provided through completing the *Intensive Practice* activities. Furthermore, based on observations of rate of progress through HER, we concluded that conducting the *Intensive Practice* tier of support as standard provision greatly increases the amount of 1:1 input required. In a number of cases, this was prohibitive to episode progress, and in fact appeared to reduce overall intensity of the programme for these children.

The main objective of a full-scale RCT to evaluate the use of HER with children with ID in the UK would be to determine the efficacy of the programme to improve the reading skills of children with ID when compared with children with ID receiving either ‘treatment as usual’ or another specified reading programme. However, further feasibility work is required to effectively design and conduct such an evaluation. In the present research, we conducted a pilot RCT to investigate the feasibility of a RCT design evaluating an online reading programme with children with ID attending special needs schools. The feasibility objectives span the four categories previously outlined. Table 2 was devised for the purpose of this study, and is based on the general guidelines for conducting feasibility research provided by Thabane et al., (2010). The table outlines these objectives and the specific questions under investigation, along with how these objectives were assessed within the current study.

## **Method**

### **Design and Methodology**

This study employed a pre-test post-test randomised group design, in which participants were randomly allocated to the intervention (HER) group, or a waiting-list-control group. Those in the HER group received this intervention in place of other

formal reading instruction they might otherwise have received, however still participated in other literacy activities, including ‘group reading’ in class.

As would be expected with a RCT design, this study utilised a predominantly quantitative approach. Although arguably some of the feasibility objectives and questions of interest could lend themselves to a more qualitative approach (e.g., some aspects of resources and management), due to the nature of HER, many of these questions could be answered objectively via the data gathered directly through use of the programme.

### **Participants**

Participants were 26 students (aged 5-19 years), who were all identified in their school records as having mild-moderate ID. All participants were recruited from three special schools in North Wales. Participants were selected by school staff in accordance with eligibility criteria established by the research team. These criteria were based on previous feasibility work, and were designed to ensure participants had an appropriately low reading level to potentially benefit from HER, and the prerequisite skills to access the programme without requiring significant additional input. As such, participants had to be able to complete the *Mousing Around* episode independently (see intervention procedure), and have a word reading age below 7 years as measured by the Schonell Reading Test (1971). This particular test was used due to low cost availability of the assessment to distribute to the schools, thus providing a convenient approximate measure of reading level of participants.

### **Intervention**

HER comprises 80 online episodes, averaging around 20 minutes, during which the programme directly delivers instruction to each learner. The episodes include explicit instruction in synthetic phonics, incorporating fluency-based activities to

ensure concepts are mastered in each lesson (see Procedure or Layng, Twyman, & Stikeleather, 2003, for more detail). Episodes were delivered on computers that were available within the schools.

As previously outlined, HER also includes two tiers of additional support—*Targeted Practice* and *Intensive Practice*. In previous pilot work using HER with children with ID, we used the *Intensive Practice* programme (comprising over 100 additional activities). However, in the current study, the *Targeted Practice* programme (comprising only 25 activities) was used to remove a previously identified potential barrier to programme completion.

HER also includes 80 stories comprising material covered in the programme. These were printed for participants to read after specified episodes. Licenses for all participants allowed access to progress reports and further information on implementation protocol (*Headsprout Teacher's Guide*, 2010). Teachers also printed a progress map from the Headsprout website for each child as a visual representation of progress.

Three checklists were used in this study. These included: an initial screening checklist for prerequisite skills during *mousing around*, an implementation checklist, and a school feedback checklist (see *Training and implementation fidelity*).

### **Outcome measures**

Although evaluation of outcomes was not the focus of this feasibility study, reading assessments were conducted pre and post intervention for all participants (baseline and 6-months post-baseline), to investigate characteristics of outcome measures and provide some information about potential effects of HER in this population. The following assessments were investigated:

The *Diagnostic Reading Analysis* (Crumpler & McCarthy, 2007) and the *Oral Reading Fluency* (ORF) subtest of the *Dynamic Indicator of Basic Early Literacy* (Good & Kaminski, 2007) were used to assess progress in oral reading. The DRA comprises of passages of increasing difficulty, and provides an accuracy score, standardised score, and reading age for each participant. The DRA provides standardised scores for children up to the age of 16years 5months. The ORF subtest consisted of three passages at Year 2 equivalent level. The child reads as many words from each passage as they can in one minute, and the median score is taken.

The *Word Recognition and Phonic Skills assessment* (Carver & Moseley, 1994) was used to assess progress in word recognition skills. In this assessment, the child is read a word and asked to choose the correct word from a choice of four or five. The assessment places children within a word recognition stage, from one (almost no word recognition knowledge) to ten (moving towards mastery of clusters and digraphs necessary for word recognition).

*Assessment reliability.* Parallel forms are used in both the DRA and the WRaPS; we administered form A at pre-test, and form B at post-test. For the DRA, the reliability of these parallel forms is reported to be 0.93 for reading accuracy, however no psychometric data on validity are available (Phillips, Kelly, Symes, Bowen & Scott, 2013). The WRaPS assessment has been found to have high internal consistency, with a Cronbach's alpha value of .97 for both forms (Keong, 2013). For the DIBELS ORF, test-retest reliability has been found to range from .92 to .97 for children of primary school age (Good, Simmons & Kame'enui, 2001). There are no test-retest reliability data available on use of this subtest with older children.

## Procedure

**Recruitment and screening.** Three special needs schools in North Wales were asked to participate in the study. One of these schools had used HER prior to this, through previous involvement with the research team. Prior to participant recruitment and assessment, each school screened potential participants to ensure they met eligibility criteria previously outlined. This included the Schonell reading test (1971), and the *Mousing Around* episode – a short introductory online episode that familiarizes the child with the instructional language of HER and provides practice of appropriate responding prior to introducing the reading episodes). To ensure each school was making similar decisions on performance on this episode, a checklist was devised to guide teachers on important prerequisite skills. This included items such as ‘*Responds appropriately to speak out loud activities without continuous prompting*’, ‘*Can click at appropriate speed in fluency activities*’.

Ethical approval was granted by the researchers’ institution. Once eligible participants had been identified by the participating schools, informed parental consent was obtained in writing for all participants prior to data collection and beginning the intervention.

**Randomisation and blinding.** Once consent was obtained, participants were randomly allocated to either the HER group or the control group prior to baseline assessment. Participants were randomised within each school using simple randomisation strategy involving random number lists in Microsoft Excel, ensuring each school had half of their participants in each group. For example, if there were six participants in one school, the numbers one to six would be sorted to occur in a random order. The first three numbers would be allocated to the HER group, and the second three numbers to the control group. The numbers would then be randomised

again next to a list of participant identifiers to determine which participants had been allocated to which group. Pre-test assessments were conducted after randomisation, but were blind to intervention condition. Thirty-six percent of post-test assessments were conducted blind to intervention condition.

**HER online episodes.** Episodes were conducted according to implementation guidelines provided by Headsprout®. Participants engaged in episodes at a computer ready to access their individual profile. A staff member remained with the child during episodes, however, the only interaction with the child was to offer encouragement to stay on task. This was to eliminate interference with the sophisticated correction procedure built into HER, ensuring responses made provided accurate feedback of the child's current ability and progress. When each child finished an episode, online data were checked to ensure the required accuracy was attained, set at 90% in each episode. If this was attained, they chose a sticker to place on their progress map that indicated which lesson they had completed.

**Sprout Stories®.** In accordance with implementation guidelines, children were also required to read stories after specified episodes. If the child struggled, staff were advised to remind them to sound out the word, and implemented the Model-Lead-Test error correction procedure as described below in the *Targeted Practice* exercises.

**HER Targeted Practice flashcards and Oral Reading Fluency.** The *Targeted Practice* tier of HER was conducted after the episodes specified in the programme protocol. This comprises around 25 frequency-building exercises consisting of individual sounds and words and 10 oral reading fluency exercises designed to ensure children were fluent on materials taught in specific episodes before they progress to the next episode. In this study we used a modified flashcards procedure rather than the sheets provided with the *Targeted Practice* materials. A teacher or classroom assistant

was advised to work with the child at the table and conduct practice sessions using a Model-Lead-Test format (Engelmann & Carnine, 1982). This involved demonstrating the procedure by responding to four cards (model), then repeating this along with the child (lead), and then the child responding alone (test). This ensured participants understood the procedure, and also served as a warm-up activity prior to timing. Participants were then timed for one-minute. Because HER was not designed specifically for children with ID, we reduced the number of correct responses required for reaching criterion to that recommended for children aged between five and six years, which varied between 25 and 50 correct responses per minute. Correct and incorrect responses were recorded on each child's data sheet. To demonstrate mastery, participants had to obtain the target for the specific activity over three timings before moving onto the next episode. Staff were advised to employ a correction procedure after each timing, again using the Model-Lead-Test format outlined previously. This was repeated until the participant responded correctly to all errors made during the timing.

Oral reading fluency measures also form part of the *Targeted Practice* programme. Participants were required to read a short passage, and the number of words read correctly per minute was recorded. As with the flashcards, oral reading fluency targets had to be met in three timings before progressing, and the same error correction procedure was employed.

**Benchmark Reading Assessments.** Twelve of the 80 stories are considered *Benchmark Reading Assessments*, to be conducted after specified episodes. For the Benchmark readers, data on reading accuracy were taken (i.e. number of words read correctly), and a rating of reading proficiency of *Independent* (read with few errors), *Satisfactory* (read with some errors and slight hesitation), or *Needs Practice* (read

with frequent errors). Staff were instructed to record these data either electronically through the HER site, or on printed sheets available to download. These data were then used, alongside the programme data, to guide decisions on whether additional frequency-building activities were required.

We repeated assessments with all participants at the end of the school year, regardless of which episode had been reached. (See results for information on the flow of participants).

**Training and implementation fidelity.** One training session was conducted in each school so that a teaching assistant or teacher could implement the programme with each participant. This training session lasted 1-2 hours, and provided an overview of the programme using Powerpoint and example excerpts from episodes of HER. The Powerpoint training covered information important to implementation fidelity, including the importance of: monitoring episode data, completing 3 episodes per week, completing benchmarks and stories, use of the *Targeted Practice* materials, and not prompting within online episodes. Correct procedures were modelled for these aspects by the trainer.

Researchers were present for either the initial session or a session early in the programme, after which we monitored online episode data to ensure fidelity of implementation. An implementation checklist to guide the sessions was adapted from Huffstetter et al. (2010) for use during training and thereafter. This included items such as: ‘Have you responded to any requests for help by redirecting the child back to the programme?’ and, ‘Have you checked each child has achieved 90% accuracy immediately after episode completion?’.

In response to school feedback and to encourage reporting of difficulties implementing HER, an additional checklist was introduced early on in the research

period. This required one staff member at each school to ask other staff implementing HER key fidelity questions (e.g. ‘Have you completed three episodes this week?’, ‘Have you completed the *Targeted Practice* materials?’) and ask whether they were experiencing any difficulties. This information served to check for implementation fidelity and as a request for assistance if required. This checklist was to be emailed to the lead researcher each week.

### **Implementation fidelity criteria**

*Episodes.* According to HER protocol, and the guidelines given to staff, at least three episodes should be completed each week. This can include episode repetitions, which should occur if episode scores are below 90%. It is also essential children do not receive external prompts during episodes, and that they speak out loud when necessary.

*Benchmark stories.* These should be read by the child, scored and rated (as previously outlined) by the staff member following relevant episodes. If an ‘N’ (*Needs practice*) is recorded, the story must be repeated until it is rated an ‘S’ (*Satisfactory*), and the child should not progress onto the next episode until this is achieved.

*Targeted Practice.* These activities should be completed following relevant episodes. Correct and incorrect responses per minute should be recorded, and the target met or surpassed on three timings prior to moving onto the next episode.

## **Results**

### **Recruitment and retention rates**

**Timeline for initial recruitment.** Recruiting schools, selecting and screening participants, obtaining consent, staff training and pre-test assessments took up to 4-

months. All schools began HER intervention in January, leaving approximately six months to the end of the school year for delivery of the intervention.

**Retention.** Figure 1 outlines the flow of participants from screening to post-test data collection using a Consort-style presentation. Consent for participation in the study (including random allocation to intervention or Control group) was obtained for all 26 eligible children. However, following randomisation, two participants from the HER group and one from the control group were excluded due to non-compliance with baseline assessments. A further participant from the control group was not available for assessment 6 months post-baseline. This demonstrates a retention rate of 85%. All 11 children enrolled in HER remained in the study.

#### **Equipment availability and resource preparation**

All schools had computers with Internet access available prior to the study. However, each school reported difficulties with teaching staff suggesting that existing facilities were inadequate for efficient delivery of HER. A number of staff members reported slow computer start up times and slow Internet access as barriers to completing more of the programme. No staff reported preparation of the additional materials as a barrier to using the additional materials.

#### **Rates of progress**

During the 6-month intervention period, none of the participants completed the programme. As reported in Table 3, episodes completed ranged from 21 to 73, with 8 participants reaching the second half of the programme (episode 40+).

#### **Implementation fidelity, adherence and compliance**

**Episodes.** The episodes were completed with the required accuracy (90%) and repeated when this was not attained. Similarly, according to the self-report fidelity checklists, participants were not prompted during the episodes and complied with the

'speak-out-loud' component of the episodes. However, as can be seen in Table 3, only 5 of the 11 participants completed 3 or more episodes per week.

**Benchmark stories.** Table 4 outlines the fidelity and compliance figures for the *benchmark stories*. This aspect of the programme was only used with fidelity with 38% of participants, and overall compliance (correct or attempted use) was 62%.

**Targeted Practice.** Table 5 outlines the fidelity and compliance figures for the *Targeted Practice* materials across each school. Overall fidelity of this aspect was 27% of all participants, and overall compliance was 72%. The following were noted as violations of procedure resulting in coding as 'used, but not with fidelity': Missing activities (sporadic data or occasional missing data); deliberate skipping of difficult activities (e.g. nonsense words or oral reading); conducting activities but not reaching frequency aim; conducting activities, reaching frequency aim, but not three times.

### **Staffing**

No requirements were specified by the research team for staffing prior to beginning HER. Staffing levels and organisation, therefore, varied across settings. Teaching assistants predominantly implemented HER, with the exception of two children (from schools two and three) for whom the class teacher was also involved. All schools chose to implement HER on a 1:1 basis. Initially, all schools allocated a staff member who was working in each child's class to be responsible for HER. Schools two and three continued with this model, with the number of participants (five and two, respectively) matching the number of staff members involved. Due to difficulties timetabling episodes within the classroom and other responsibilities of the assistant, school one changed their model of staffing approximately one month into the research period. One assistant was then responsible for implementing HER with

three of the four participants, taking them out of their respective classes for episodes and additional activities. HER was still conducted 1:1 for these children.

There was no notable difference in progress of participants (Table 3), fidelity of implementation (Tables 4 and 5) and outcomes (Table 6) across settings and modes of delivery.

### **Training**

Although initial training and support was reported by staff to be adequate, the low fidelity and compliance rates with some aspects of HER suggests that training had not resulted in sufficient competence or adherence to the protocols to ensure high fidelity and quality implementation.

### **Defining 'Education as usual'**

Both the form and the frequency of reading instruction varied between each school. One school reported use of a specific programme once per week, with generic class reading each day; one reported use of a different programme with varying frequency, whereas another reported more general work on alphabetic knowledge and sight words that varied from daily to weekly depending on the child.

### **Baseline data**

Tables 6 and 7 outline the baseline age and DRA Reading age and Accuracy scores of participants in both groups. Pre-test DRA reading ages were similar between groups, ranging from <5years to 6yrs 4months in the HER group, and from <5yrs to 6yrs 2months in the Control group. However, there was a marginally significant difference in DRA accuracy scores at pre-test, with the HER group scoring higher prior to enrolling in HER ( $F(1,19) = 1.30, p = .062$ ). A large effect size was found for this difference ( $d = 0.85$ ). Group means for all reading measures can be seen in Table 9.

### **Exploratory statistical analysis of reading assessment outcomes**

Although there are limitations of the outcome data in the current study (namely the small sample size), there is a good case for analysing initial evidence for some kind of effect, therefore exploratory statistical analysis was conducted. Table 8 shows the mean scores, results of ANCOVA and t-test analysis, and effect sizes for all measures for the HER and Control group. To control for potential pre-test differences between the groups despite initial random assignment, a one-way analysis of covariance model was used. Because no Control group participants attained a standardised score on the DRA at pre-test, an independent samples t-test was conducted to compare post-test scores only for this outcome measure. Effect sizes based on Cohen's *d* were calculated using the mean change scores for the HER and Control group and the pooled pre-test Standard Deviation (Rosnow & Rosenthal, 1996). For the DRA reading age and WRaPS word recognition age, effect sizes based on Cohen's *d* were calculated using the means at post-test for each condition and the pooled post-test Standard Deviation.

The results indicate that for the DRA Reading Age, WRaPS Raw Score and word Recognition Stage, there were significant differences in the HER group at 6 months post-baseline assessment compared to the control group, and marginally significant differences in Standardised score, Accuracy Score, and DIBELS Oral Reading Fluency. Large effect sizes were found for DRA Accuracy Score and Reading Age. A medium effect size was found for DIBELS ORF and small effect sizes for WRaPS Raw Score and Stage.

## **Discussion**

Overall, this study demonstrates the importance of feasibility research to inform aspects of a larger investigation that would require significantly more resources. In addition, exploratory statistical analysis suggested evidence of potential intervention effects to examine in future research. Thus, with attention to design and methodology issues, a future large scale RCT of HER in special schools appears to be feasible.

### **Recruitment, retention rates and assessment**

Recruitment was feasible at a rate of 6-7 children per month, and schools and parents were willing to consent to randomisation. Eligibility criteria did not appear to be unduly exclusive; each school was able to identify potential participants. The retention rate of 85% of participants in the research evaluation suggests a relatively low rate of loss to follow-up.

The current study also demonstrated that assessors can be blind to condition. Whilst only 36% of post-test assessments were conducted blind to condition, this was only due to resources to fund assessors and training. No assessments were unblinded; we planned for the remaining 64% of assessments to be conducted by the lead researcher (who was not blind to condition), due to limited resources. If funding allowed for further independent assessors, it appears that all assessments could be conducted blind to condition.

### **Equipment availability**

Although all schools had seemingly adequate computing facilities prior to beginning HER, staff indicated this was a factor that affected implementation and programme intensity.

### **Rates of progress**

No participants completed HER in the 6-month research period, and some of those reaching the second half of the programme had not received the *Targeted Practice* elements with the recommended intensity. This suggests that a longer intervention period would be required to give participants the opportunity to complete the programme prior to post-test assessment.

### **Adherence, implementation fidelity and training**

Although all participants enrolled in HER received the intervention, adherence in relation to the ‘dose’ or frequency of the intervention was low, with only 45% of participants completing the recommended 3 episodes per week, with an average completion rate of 2 episodes per week across all schools. This indicates that further support is required to enable schools to schedule and embed the programme into literacy provision in special schools, in particular looking to address the barriers identified in this study.

Implementation fidelity with HER is perhaps not as complex as with other educational interventions, due to the instruction within lessons being delivered directly from the programme. However, use of the *Benchmark Assessments* to ensure children do not progress prior to mastering the concepts taught in each episode, and the addition of the *Targeted Practice* tier of support – both involving specific delivery protocol – do require fidelity checks to monitor implementation. The poor levels of implementation fidelity for these aspects of HER indicate staff training and monitoring of these aspects (which occurred via self-report checklists) was not adequate in the current study, despite staff reporting satisfaction with the training provided. Staff reported that the implementation session checklist (adapted from Huffstetter, et al., 2010) served the functions of reminding staff of important aspects

of delivery and providing an easy way to allow other staff to supervise a child on the episodes if required. However, the self-report checklist incorporated two months into the research period was only partially successful; it did alert the researchers to some difficulties, but some staff members continually reported no difficulties when in fact there were violations of procedure and sometimes complete omissions of aspects of the programme. This indicates that both checklists might be best accompanied by periodic observations of and feedback on implementation to ensure fidelity in all aspects of delivery. Further, an extended or additional training session to allow for direct practice and feedback on implementation of *Benchmark Stories* and *Targeted Practice* materials would also likely enhance delivery of these elements. Despite these challenges, we did find that teaching assistants were able to support the delivery of the intervention in schools.

### **Considerations and recommendations for a future, full-scale RCT**

In a future definitive RCT of HER, consideration would need to be given to the control or comparison. Initially, a comparison with reading education as usual would probably be important. However, we have only very limited information about education as usual from the three schools involved in the current study. A priority for future feasibility research would be to gather more detailed information to inform a later RCT.

The current study used individual randomisation within each school; each school had some participants in the HER group and some in the control group. This procedure was successful in this context. However, the feasibility of individual randomisation in a larger, full-scale evaluation is important to consider because of the potential for contamination (teachers using HER with children allocated to the control group).

An alternative approach that is increasingly common in RCT studies is to use a cluster randomised design, in which random allocation to groups is at the level of the group or cluster (Moore, Graham & Diamond, 2003; Donner & Klar, 2000). In the context of a HER evaluation in special schools using this design, randomisation would be at the level of the school. Schools would be recruited on the basis that they would be allocated to deliver HER or to a control group in which they continue to deliver their usual provision.

Three assessments were investigated in the current study: the DRA (comprising a reading accuracy score, a standardised reading score, and a reading age); the *Dynamic Indicator of Basic Early Literacy Skills* Oral Reading fluency subtest; and the *Word Recognition and Phonic Skills assessment*. The first two feasibility objectives related to these assessments were to investigate the appropriateness of the assessments for measuring reading skills with children with ID, and to ascertain whether the measures detected changes in reading skills. With the exception of the two participants excluded from the HER group due to non-compliance during baseline assessments, there were no difficulties in administering each of the assessments according to the respective published protocol. However, there were some notable observations concerning the WRaPS assessment with some participants. Within the assessment, there are up to four distractor items alongside the target word. All assessors were trained to administer this at a slow pace and to remind participants throughout the assessment to read each word prior to selecting their answer, however, a number of participants continued to select their responses rapidly and seemingly without surveying the possible answers. Although there were significant differences in WRaPS scores between the groups at post-test, this could account for the variability in performance on this measure, with some participants – in

both groups – giving fewer correct responses at post-baseline assessment. This brings into question the validity of this measure in the current study, and indicates it might not be a suitable measure in a full-scale RCT with this population.

The DRA did detect changes in reading skills. However, there was a floor effect for standardised scores in the control group at both baseline and post-baseline assessment. Although participants in the control group did attain an accuracy score, these scores were too low to attain a standardised score. This has implications for the use of standardised scores as the primary output of interest from this assessment, as without any mean or standard deviation values it is not possible to run ANCOVA or to calculate an effect size for this output. Because we are targeting children at beginning reading level – as HER is designed for – this problem is likely to occur with standardised scores. Similarly, although not the case in the current study, failure to attain a reading age in such a cohort is also a distinct possibility. It might therefore be necessary to use reading accuracy scores from the DRA as the primary output of interest from this assessment.

The DIBELS ORF also detected changes in reading skills. However, because correct words per minute can be calculated from the DRA output, this assessment would not be required in addition to the DRA. Further, despite changes in ORF scores being detected, the scores attained on this measure were generally low. Whilst this could be due to the weak implementation fidelity demonstrated in the study (in that we might arguably expect greater improvements under more faithful execution of the intervention, and upon completion of more episodes), it could indicate the importance of including other subtests of the DIBELS as outcome measures. HER as a whole is designed to improve reading skills at a composite level (as measured by composite measures such as oral reading skills); it does so through building skills in phonemic

awareness, alphabetic knowledge, and fluency in applying such skills to decoding text effectively. As such, inclusion of DIBELS subtests designed to measure these component skills (such as phoneme segmentation fluency and nonsense word fluency) should be considered for a future RCT.

**Declaration of interest statement:**

There are no conflicts of interest to report for any authors of this paper.

## References

- Andrews, R. (2004). *The impact of ICT on literacy education*. RoutledgeFalmer, London.
- Arain, M., Campbell, M. J., Cooper, C. L., & Lancaster, G. A. (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Medical research methodology*, *10*, 67, 1-7
- Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted Instruction in support of beginning reading instruction: a review. *Review of educational research*, *72*, 1, 101-130
- Browder, D. M., & Xin, Y. P. (1998). The meta-analysis and review of sight word research and its implications for teaching functional reading to individuals with moderate and severe disabilities. *The Journal of Special Education*, *32*, 130-53
- Browder, D. M., Wakeman, S. Y., Spooner, F., Ahlgrim-Delzell, L., & Algozzine, B. (2006). Research on reading instruction for individuals with significant cognitive disabilities. *Exceptional Children*, *72*, 392–408
- Browder, D. M., Gibbs, S., Ahlgrim-Delzell, L., Courtade, G. R., Mraz, M., & Flowers, C. (2009). Literacy for students with severe developmental disabilities: What should we teach and what should we hope to achieve? *Remedial and Special Education*, *30*, 269-282
- Carver, C., & Moseley, D. (1994). *A group or individual diagnostic test of Word Recognition and Phonic Skills (WRaPS)*. London: Hodder & Stoughton.
- Campbell, N. C., Murray, E., Darbyshire, J., Emery, J., Farmer, A., Griffiths, F., Guthrie, B., Lester, H., Wilson, P., & Kinmonth, A. L. (2007). Designing and evaluating complex interventions to improve health care. *British Medical Journal*, *334*, 7591, 455-459

- Clarfield, J. & Stoner, G. (2005). The Effects of Computerized Reading Instruction on the Academic Performance of Students Identified with ADHD. *School Psychology Review*, 34 (2), 246-254
- Coleman-Martin, M. B., Heller, K. W., Cihak, D. F., & Irvine, K. L. (2005). Using computer-assisted instruction and the nonverbal approach to teach word identification. *Focus on autism and developmental disabilities*, 20, 80-90
- Crumpler, M., & McCarthy, C. (2007). *Diagnostic Reading Analysis: Manual* (2nd ed.). Totten, Hampshire: Hobbs the Printers Ltd.
- Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications*. New York: Irvington.
- Fitzgerald, J., Roberts, J., Pierce, P., & Schuele, C. M. (1995). Evaluation of home literacy environment: An illustration with preschool children with Down syndrome. *Reading and writing quarterly*, 11, 311-334.
- Good, R. H., & Kaminski, R. A., (Eds.) (2007). *Dynamic indicators of Basic Early Literacy Skills* (6<sup>th</sup> edition).
- Graf, S., & Lindsley, O. (2002). *Standard Celeration Charting 2002*. Youngstown, OH: Graf Implements.
- Grindle, C. F., Hughes, J. C., Saville, M., Huxley, K., & Hastings, R. P. (2013). Teaching children with autism to read using Headsprout Early Reading. *Behavioral Interventions*, doi: 10.1002/bin.1364
- Gueron, J. M. (2002). *The politics of random assignment: Implementing studies and affecting policy*. In F. Mosteller & R. F. Loruch (Ed.s) (2002). Evidence matters: Randomized trials in education research. Brookings institution.
- Huffstetter, M., King, J. R., Onwuegbuzie, A. J., Schneider, J. J., & Powell-Smith, K. A. (2010). Effects of a computer-based reading program on the early reading

- and oral language skills of at-risk preschool children. *Journal of education for students placed at risk*, 15, 279-298
- Hutchinson, D., & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers*. National foundation for educational research.
- Jones, K. M., Torgesen, J. K., & Saxton, M. A. (1987). Using computer guided practice to increase decoding fluency in learning disabled children: A study using the Hint and Hunt I program. *Journal of learning disabilities*, 20, 122-128
- Joseph, L.M., & 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, 88-94
- 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, 3-15
- Kliewer, C. and Landis, D. (1999). Individualizing literacy instruction for young children with moderate to severe disabilities. *Exceptional children*, 66, 85-100.
- Layng, T. V. J., Twyman, J. S., & Stikeleather, G. (2003). Headsprout Early Reading: Reliably teaching children to read. *Behavioral Technology Today*, 3, 7-20
- Macaruso, P., Hook, P. E., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading*, 29, 2, 162-172  
<https://doi.org/10.1111/j.1467-9817.2006.00282.x>
- Marks, J. (2000). *What are special educational needs? An analysis of a new growth industry*. Centre for Policy Studies.
- Moran, J., Ferdig, R. E., Pearson, P. D., Wardrop, J., & Blomeyer, R. L. (2008). Technology and reading performance in the middle-school grades: A meta-

- analysis with recommendations for policy and practice. *Journal of Literacy Research*, 40.1: 6-58
- Moore, L., Graham, A., & Diamond, I. (2003). On the feasibility of conducting randomised trials in education: case study of a sex education intervention. *British educational research journal*, 29, 5, 673-689
- National Reading Panel (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC: National Institute for child health and development.
- Oakley, A. (1998). Public policy experimentation: Lessons from America. *Policy studies*, 19, 2, 93-114
- Oakley, A. (2000). *Experiments in knowing: Gender and method in the social sciences*. Cambridge: Polity Press.
- Oakley, A. (2006). Resistances to 'new' technologies of evaluation: education research in the UK as a case study. *Evidence and policy*, 2, 1, 63-87
- Soe, K., Koki, S., & Chang, J. M. (2000). Effect of computer-assisted instruction (CAI) on reading achievement: A meta-analysis. Pacific Resources for Education and Learning, Honolulu, HI. Retrieved from: <http://www.prel.org/products/products/Effect-CAI.pdf>
- Torgesen, C., Waters, M. D., Cohen, A. L., & Torgesen, J. L. (1988). Improving sight-word recognition skills in LD children: and evaluation of three computer program variations. *Learning disability quarterly*, 11, 125-132
- Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rlos, L., Robson, R., Thabane, M., Glangregorio, L. & Goldsmith, C. H. (2010). A tutorial on pilot studies: the what, why and how. *BMC Medical research methodology*, 10, 1, 1-10

- Torgerson, C., & Zhu, D. (2003). *A systematic review and meta-analysis of the effectiveness of ICT on literacy learning in English, 5-16*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Thornicroft, G., Lempp, H., & Tansella, M. (2011). The place of implementation science in the translational medicine continuum. *Psychological medicine, 41*, 10, 2015-2021
- Torgerson, C., & Torgerson, D. (2001). The need for randomised controlled trials in educational research. *British journal of educational studies, 25*, 129–143.
- U.S. Department of Education, Institute of Education Sciences, National Assessment of Educational Progress (NAEP) Data, 2007, Retrieved January 7<sup>th</sup> 2019, from <http://nces.ed.gov/nationsreportcard/>
- Van Teijlingen, E. R, & Hundley, V. (2001). The Importance of Pilot Studies. *Social Research Update*
- Van Teijlingen, E. R, Rennie, A.M, Hundley, V., & Graham, W. (2001). The importance of conducting and reporting pilot studies: the example of the Scottish Births Survey. *Journal of advances in nursing, 34*, 289-295
- Wehmeyer, M. L. (2006). Beyond access: Ensuring progress in the general education curriculum for students with severe disabilities. *Research and practice for persons with severe disabilities, 31*, 322-326
- Whalon, K. J., Otaiba, S. A., & Delano, M. E. (2009). Evidence-based reading instruction for individuals with autism spectrum disorders. Focus on autism and other developmental disabilities, 24, 3-16
- Whitcomb, S. A., Bass, J. D., & Luiselli, J. K. (2011). Effects of a computer-based early reading program (Headsprout) on word list and text reading skills in a student with autism. *Journal of physical and developmental disabilities, 23*,

491-499

Wise, B. W., & Olson, R. K. (1995). Computer-based phonological awareness and reading instruction. *Annals of Dyslexia*, 45, 99-122

Table 1. *Initial feasibility objectives for investigating the use of HER with children with ID.*

---

<b>Initial feasibility objectives</b>
<b>Accessibility:</b> <ul style="list-style-type: none"><li>• Can HER be accessed by children with ID?</li><li>• Is additional input required to enable access?</li><li>• Do children with ID appear to benefit from HER?</li></ul>
<b>Implementation feasibility (episodes + <i>Intensive Practice</i> activities):</b> <ul style="list-style-type: none"><li>• How does episode performance of children with ID compare to TD children?</li><li>• How feasible is the use of the <i>Intensive Practice</i> tier of support as standard provision for children with ID using HER?</li></ul>

---

Table 2. *Feasibility objectives, specific questions of interest and method of measurement of objectives.*

<b>Feasibility objectives and specific questions of interest</b>	<b>Measurement of objective</b>
<b>Recruitment and retention rates (process):</b>	
<ul style="list-style-type: none"> <li>• Are schools and parents willing to consent to randomisation?</li> <li>• What is the rate of retention?</li> <li>• What eligibility criteria might be appropriate? Is the chosen eligibility criteria appropriate (i.e. not unduly exclusive or too inclusive)?</li> <li>• How long is the recruitment, consent and screening process?</li> </ul>	Collating recruitment and retention data
<b>Assessing (process):</b>	
<ul style="list-style-type: none"> <li>• Can assessors be blind to condition?</li> <li>• How long is the assessment process?</li> </ul>	Collating assessment data
<b>Equipment availability and resource preparation</b>	
<b>(resources &amp; management):</b> <ul style="list-style-type: none"> <li>• Do schools have adequate computing facilities to implement the programme?</li> <li>• How feasible is preparation of additional resources?</li> </ul>	Informal staff interview
<b>Rate of progress through the programme (process):</b>	
<ul style="list-style-type: none"> <li>• What is the average rate of progress through the episodes?</li> </ul>	Aggregated episode data from HER
<b>Implementation fidelity/compliance rates (process):</b>	
<ul style="list-style-type: none"> <li>• Are implementation guidelines adhered to for the episodes (e.g. 3 per week), the <i>Targeted Practice</i> activities (completed prior to episode progress), and the benchmark reading assessments?</li> <li>• To what extent does this vary across settings?</li> </ul>	HER data and informal staff interview

---

<ul style="list-style-type: none"> <li>• What fidelity measures might be used, and are these sufficient?</li> </ul>	
---	--

---

<p><b>Staffing and training requirements (resources &amp; management):</b></p> <ul style="list-style-type: none"> <li>• What initial training and subsequent support might be provided, and is this sufficient?</li> <li>• What level of staffing for HER implementation might influence progress, fidelity and outcomes?</li> </ul>	<p>Informal staff interview</p>
--	---------------------------------

---

<p><b>Appropriate outcome measures and potential effects of HER in this population (scientific):</b></p> <ul style="list-style-type: none"> <li>• Are the chosen outcome measures appropriate for this population?</li> <li>• Do the chosen outcome measures detect change?</li> <li>• Are there any statistically significant effects of the intervention?</li> </ul>	<p>Analysis of pre and post-test reading assessment data</p>
--	--

---

<p><b>Defining ‘Education as usual’:</b></p> <ul style="list-style-type: none"> <li>• What is ‘treatment as usual’ for reading instruction with children with ID in Special needs schools?</li> <li>• To what degree is this consistent across settings?</li> </ul>	<p>Informal staff interview</p>
---	---------------------------------

---

Table 3. *Individual episode progress, including episode reached, number of repetitions, and average number of episodes completed each week.*

<b>School</b>	<b>Participant</b>	<b>Episode reached</b>	<b>Number of repetitions</b>	<b>Average completed per week</b>
<b>1</b>	1	37	3	<3
	2	46	0	<3
	3	61	3	>3
	4	73	0	>3
<b>2</b>	5	21	0	<3
	6	40	0	<3
	7	59	0	>3
	8	61	0	>3
	9	73	0	>3
<b>3</b>	10	37	1	<3
	11	43	0	<3

Table 4. *Fidelity and compliance figures for use of benchmark reading assessments in each school.*

<b>School</b>	<b>Used with fidelity</b>	<b>Used but not with fidelity</b>	<b>Not used</b>	<b>% of children used with fidelity</b>	<b>% compliance (correct or attempted use)</b>
1 ( <i>n</i> =4)	3	0	1	75	75
2 ( <i>n</i> =5)	2	1	2	40	60
3 ( <i>n</i> =2)	0	1	1	0	50
All	5	2	4	38	62

Table 5. *Fidelity and compliance figures for use of Targeted Practice materials in each school.*

<b>School</b>	<b>Used with fidelity</b>	<b>Used but not with fidelity</b>	<b>Not used</b>	<b>% of children used with fidelity</b>	<b>% compliance (correct or attempted use)</b>
1 ( <i>n</i> =4)	1	2	1	25	75
2 ( <i>n</i> =5)	2	2	1	40	80
3 ( <i>n</i> =2)	0	1	1	0	50
All	3	5	3	27	68

Table 6. *HER Group Ages at Pre-test, Reading Ages, and Reading Accuracy DRA scores at Pre-test and Post-test*

<b>School</b>	<b>Participant</b>	<b>Age (years,months)</b>	<b>Pre-test Reading age (years,months)</b>	<b>Pre-test Reading accuracy</b>	<b>Post-test Reading age (years,months)</b>	<b>Post-test Reading accuracy</b>
<b>1</b>	1	17,1	<5	36	<5	32
	2	12,1	<5	40	6,0	80
	3	14,4	<5	30	<5	38
	4	14,8	<5	39	<5	42
<b>2</b>	5	11,6	<5	20	<5	20
	6	7,6	6,4	90	6,6	95
	7	18,0	<5	41	6,2	89
	8	5,5	<5	28	8,3	130
	9	17,6	<5	42	<5	43
<b>3</b>	10	11,10	<5	52	<5	40
	11	11,2	<5	41	8,6	136

Table 7. *Control Group Ages at Pre-test, Reading Ages, and Reading Accuracy DRA scores at Pre-test and Post-test*

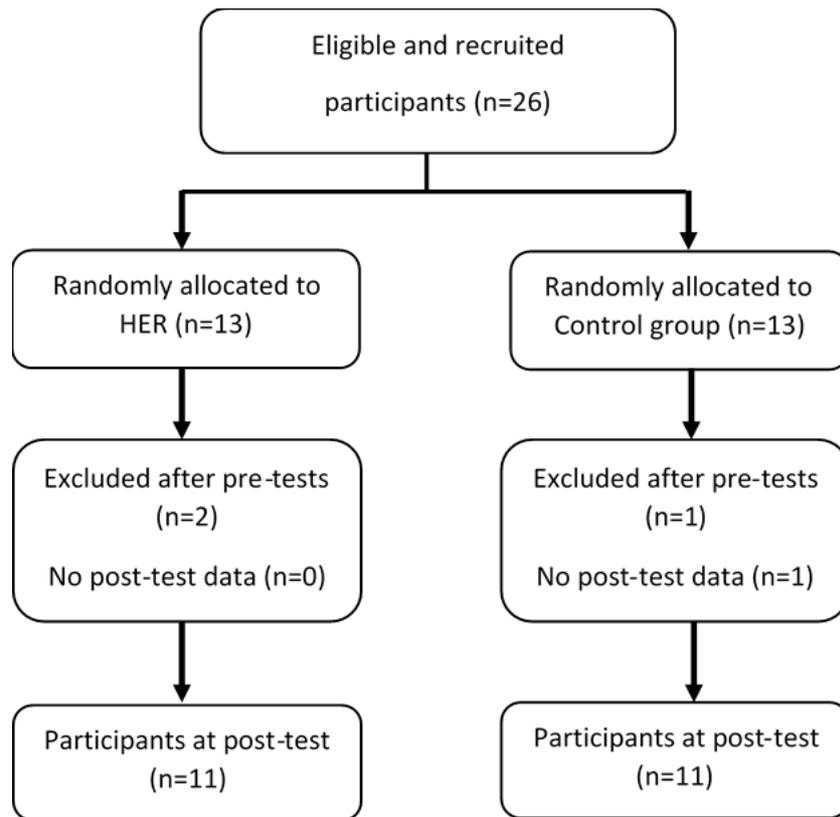
<b>School</b>	<b>Participant</b>	<b>Age (years,months)</b>	<b>Pre-test Reading age (years,months)</b>	<b>Pre-test Reading accuracy</b>	<b>Post-test Reading age (years,months)</b>	<b>Post-test Reading accuracy</b>
<b>1</b>	12	13,1	<5	4	<5	13
	13	16,3	6,2	86	7,0	109
	14	13,1	<5	1	<5	11
<b>2</b>	15	9,4	<5	28	<5	27
	16	14,2	<5	12	<5	15
	17	13,0	<5	0	<5	0
	18	12,7	<5	47	<5	11
	19	18,7	<5	19	<5	12
<b>3</b>	20	12,3	<5	0	<5	0
	21	10,5	<5	29	<5	42
	22	12,0	<5	25	<5	27

Table 8. Results of ANCOVA analysis including all participants on measures of reading outcome.

		HER		Control		ANCOVA		
		Pre-test	Post-test	Pre-test	Post-test	<i>F</i> (2, 19)	<i>p</i>	<i>d</i>
		<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )			
Diagnostic Reading Analysis	Standardised score <sup>1</sup>	7.73 (25.63)	17.55 (39.57)	0 (0)	0 (0)	13.76	.086	n/a <sup>2</sup>
	Accuracy Score	41.73 (18.12)	67.73 (40.46)	22.82 (25.79)	24.27 (30.66)	3.48	.078	0.99
	Reading age (months)	6.91 (22.91)	38.63 (45.31)	6.73 (22.31)	7.64 (25.33)	5.06	.036	1.39
WRaPS	Word recognition Raw score	34.18 (6.15)	38.55 (5.45)	18.36 (9.37)	20.45 (8.99)	4.84	.04	0.22
	Word recognition stage	6.82 (1.54)	8.00 (1.18)	3.18 (2.14)	3.64 (2.02)	8.13	.01	0.31
	Word recognition Age (months) <sup>1</sup>	82 (5.88)	86.18 (5.04)	56.73 (28.97)	63.27 (22.53)	2.09	.164	0.09
DIBELS	Oral Reading Fluency (correct words per minute)	22.45 (9.85)	35.18 (16.90)	10.64 (19.87)	14.09 (19.97)	3.48	.078	0.51

<sup>1</sup> An Independent samples T-Test was conducted on post-test scores for this output due to no participants in the control group attaining scores at pre-test.

<sup>2</sup> Cohen's *d* could not be calculated for this output due to no participants in the control group attaining scores at pre or post-test.



*Figure 1.* The flow of participants from initial screening to post