# Headsprout® Early Reading for children with severe intellectual disabilities: a single blind randomised controlled trial

Corinna F. Grindle[1,2] (iD), Clodagh Murray[3], Richard P. Hastings[1], Tom Bailey[1], Helen Forster[4], Sabia Taj[4], Andreas Paris[1,2], Michael Lovell[2], Freddy Jackson Brown[2] and J. Carl Hughes[5]

[1]University of Warwick; [2]Centre for Behaviour Solutions Ltd; [3]Emirates College for Advanced Education; [4]Calthorpe Academy School; [5]Bangor University

Research evaluating phonics reading programs for children with severe intellectual disabilities (ID) is limited. The current study investigated whether using an online reading program (Headsprout® Early Reading; HER) as supplementary reading instruction for children with a severe ID leads to improvements in reading skills as compared to children not receiving this additional instruction. Fifty-five children from a special school were randomly allocated into the HER group or a waiting list control group. For six months, children in the intervention group received HER as supplementary instruction, whereas children in the control group received only reading as usual' teaching. Pre- and post-intervention tests on standardised reading measures were conducted. Analysis of data from outcome measures indicated that the HER group made improvements at post-intervention in comparison with the control group, with medium effect sizes on two domains from the main outcome measure. These results support the case for a larger research trial of HER for children with severe ID.

## Introduction

Children with intellectual disabilities (ID) can experience significant delays in reading development, and many do not go on to acquire functional reading skills (Ratz and Lenhard, 2013). Failing to develop reading proficiency can have a negative impact on other academic, vocational and life skills (Nally et al., 2018), impeding a person's access to inclusive settings. Researchers have examined three research foci to date: (1) implementing pre-existing reading curricula for neurotypical children but with modifications to meet the needs of children with ID, (2)

creating new comprehensive curricula and (3) creating targeted interventions for a specific skill difficulty.

For example, Mathes and Torgesen (2005) extended an existing curriculum, *Early Interventions in Reading*. This was a 240-lesson curriculum for neurotypical learners adapted by adding a new level targeting reading prerequisites for students with ID. The intervention contained several of the elements of effective instruction: phonemic awareness, fluency and comprehension (National Reading Panel, 2000), and incorporated explicit, systematic instruction with modelling and frequent opportunities to respond. Allor et al. (2010) evaluated this approach with 28 children with moderate ID, randomly assigned to intervention or control groups. The students in the intervention group received daily instruction for approximately 40 min per session, whereas students in the control group received 'teaching as usual'. Students in the intervention group outperformed the control group on outcome measures, with consistently large effect sizes.

The Early Literacy Skills Builder (ELSB; Browder et al., 2007) is an example of a new comprehensive reading curriculum designed for students with severe ID. It combines basic literacy skills and phonemic awareness activities to teach skills that lead into a phonics-based program. Browder et al. (2007) evaluated the impact of the ELSB with 23 students with severe ID. Twelve students in a control group received 'teaching as usual' instruction based on a sight word curriculum (Edmark: Austin and Boekman, 1990). Eleven students in the intervention group received the ELSB intervention. Gains on research team-designed measures of early literacy indicated that students who received the ELSB curriculum learned significantly more reading skills than those in the control group. In terms of focussed skills, other research groups have developed targeted interventions for skill instruction such as decoding

strategies (Tucker Cohen et al., 2008) and letter-sound correspondence and blending (Waugh et al., 2009).

Instructional practices incorporating explicit, systematic instruction, with modelling and frequent opportunities to respond, feature prominently in these three types of research on teaching children with ID to read, suggesting that explicit instruction may provide a strong foundation for evidence-based reading instruction for children with ID (Browder et al., 2006, 2009). However, individuals providing the instruction in research were typically highly trained and experts in the specific instructional procedures. This may limit the wide-spread adoption of these teaching practices where access to experts in reading instruction may be limited.

Computer-assisted instruction (CAI) may be one way that teachers and teaching assistants who are not extensively trained in evidence-based reading interventions may be able to support children's learning without other expert support. Pennington (2010) reviewed research using CAI to teach academic skills to students with autism and concluded that CAI was an effective intervention for teaching reading with this population. Pennington acknowledged that although educational software packages are increasingly being used by teachers in everyday school settings, there is a scarcity of research evaluating these programs. Pennington also lamented the lack of robust experimental designs being used in these research studies.

There are several reasons why CAI may facilitate the learning of reading for children with ID. Macaruso et al. (2006) suggested that CAIs are well-placed to provide teaching material matched to students' current levels in a highly motivating manner. Silver and Oakes (2001) have also suggested that CAI can deliver instruction in a uniform, consistent way that may promote greater efficiency of learning for students. Grindle et al. (2013) also argued that students with ID may find it difficult to discriminate relevant information in a learning situation (Baranek, 2002) and that concentrating on a computer screen during CAI, where minimal, pertinent, information is presented, may help to overcome these problems. Considering the potential benefits of CAI and the number of available software programs (Pennington, 2010), it seems imperative that these programs be evaluated for use with children with ID.

### Headsprout® Early Reading

Headsprout® Early Reading (HER) is an Internet-based reading program that teaches reading through phonics, focussing its instruction on phonemic awareness, grapheme-phoneme correspondence, and blending sounds to decode words phonetically, and incorporating elements of vocabulary, fluency and comprehension (Layng et al., 2004). HER is based on discovery learning principles, using a mixture of explicit instruction and engineered discovery (Layng et al., 2004). HER contains 80 20-minute

episodes with instruction presented in frames, where each frame is a different illustrated screen with a learning objective, requiring an active response on the part of the student. It takes the form of a game, with characters and appealing animations. HER begins at a basic level and gradually increases in difficulty. Errors are kept to a minimum as the instruction adapts based on the child's responses. The child experiences a high rate of correct responding and positive feedback. Children are also required to practice skills until they perform them at mastery level (Twyman et al., 2012).

HER was initially developed as an individualised reading intervention for neurotypical learners who were struggling with learning to read, and it has performed favourably in several research studies with this population (Huffstetter et al., 2010; Twyman et al., 2011; Tyler et al., 2015a;). An emerging body of research implementing HER (with some adaptations) has focussed on meeting the needs of students with ID. For example, Grindle et al. (2013) investigated the feasibility of using HER with four children with autism and the adaptations needed to support their progression through the program. All four participants were able to complete the 80 HER episodes and demonstrated increases in standardised tests of reading that were maintained at 8-week follow-up.

Research has also evaluated using HER with students with mild or moderate ID. In one study, six students aged between 7 and 14 years completed all 80 HER episodes over 13–21 months. (Tyler et al., 2015b). One-to-one adult support to provide encouragement was used during episodes, as were frequency-building exercises to accompany the online sessions. All students improved their scores on standardised measures of phonemic awareness, nonsense word decoding and word recognition skills.

One recent study has also investigated feasibility questions related to conducting a full-scale randomised controlled trial (RCT) evaluation of HER (Roberts-Tyler et al., 2019). Employing a pilot randomised pre-test post-test group design with 26 students with ID, the researchers found that HER had a significant effect on reading skills when compared with 'reading as usual' with large effect sizes on the main outcome measure. A lack of information about reading education as usual for the three schools involved in the study limited conclusions about the effects of the intervention in comparison with the reading education that children in the control group received.

There are some research areas where greater clarity is needed. First, the evidence base focusses predominantly on studies with small participant numbers and lack of robust experimental designs. The absence of experimental control in existing studies is limiting. Second, existing studies on using HER with students with ID have typically included participants with mild or moderate ID.

Thus, it would be beneficial to evaluate HER in terms of its potential for teaching students with severe ID to read. Third, HER has typically been evaluated with students with ID supported by well trained, regularly supervised staff, and when the intervention is delivered with fidelity. Real-world conditions (e.g. with non-specialists delivering the intervention with minimal supervision from the research team) may pose a challenge to the effectiveness of HER.

Our aims were to evaluate HER with a reasonably large cohort of students with severe ID including associated conditions such as autism, adopting a randomised controlled trial design, and by training teachers and teaching assistants to implement HER in a special school setting with minimal ongoing supervision from the research team.

**Method**
Approval for this study was sought and obtained from the Psychology Research Ethics Committee at Bangor University.

*Design overview*
Fifty-five students who attended a school for children with a severe ID were randomly assigned to a HER group (supplemental instruction using HER) or a waiting list control group. Students in the control group received *only* the standard reading provision offered by the school (Reading as Usual). Assessments of the students' reading skills were undertaken at two data collection points: prior to intervention (baseline) and again seven months post-randomisation.

*Setting and participants*
The study took place in a large special school in the UK. At the time of the study, 385 students with severe ID were enrolled at the school. Of the 55 students, 21 were in elementary education classes (age 5−11), 30 in secondary education classes (age 11–16) and four were in post-16 education (aged 16–19). All students regardless of age used the same program in which cartoon characters appear in the online episodes. This was considered appropriate as other research had found no negative aspects to the child-friendly design of HER when it was implemented with older teenagers (Tyler et al., 2015b) and with adults (O'Sullivan et al., 2017) with ID.

Twelve students had a diagnosis of an autism spectrum disorder (ASD) alongside their severe ID. In the UK, the diagnosis of severe ID is identified by the Local Education Authority based on the primary needs of the student stated in their Education and Health Care Plan (EHCP). The EHCP is a legal document that describes a young person's special educational, health and social care needs. Generally speaking, for a student to be identified as having a 'severe' ID, they would need to have at least some of, but not all of, the following described in their EHCP:

(1) little or no speech, (2) finds it very difficult to learn new skills, (3) needs support with daily activities such as dressing, washing, eating and keeping safe, (4) has very limited social skills and (5) likely to need life-long support.

Descriptive information about students in the two study arms is provided in Table 1. Twenty-seven students (the HER group) were supported to access HER either in a separate room or in a quiet corner of their classroom. The remaining 28 students (the control group) received Reading as Usual instruction within their classrooms.

All students had received the typical reading instruction offered by the school prior to the study and had been identified by their class teachers as requiring additional support with learning how to read. No student selected for the study was able to decode phonemes (the sounds) that made up a word. That is, they either were not able to expressively 'sound out' the phonemes that made up a word to read a whole word, or they were not able to receptively identify the correct phonemes to compose a required target word. For example, when asked, 'make / cat/' when the phoneme cards '/c/a/t/' were arranged in a random array on the table in front of them, they could not move the phoneme cards to make the whole word.

Inclusion criteria were (a) ability to sit at a computer for a period of time (up to 10 min), (b) understanding and following one- or two-step instructions (e.g. 'clap hands, and turn around') and (c) ability to respond to feedback (praise or correction). As participants had severe ID, they were eligible for the study if they could a) imitate at least 10 different spoken sounds or words and b) make at least 10 self-initiated vocalisations across the day using single words (i.e. speaking in short phrases or sentences was not required). Students were selected for the study if they met the inclusion criteria and if the class teachers confirmed that the content of HER was consistent with their literacy goals as stated on their Individualised Education Plan (in

**Table 1: Participant characteristics**

| Variable | HER group (n = 27) | Reading as usual group (n = 28) |
|---|---|---|
| Gender | | |
| Male | 19 | 19 |
| Female | 8 | 9 |
| Main diagnosis | | |
| Severe ID | 21 | 22 |
| Severe ID and ASD | 6 | 6 |
| Age in months | | |
| Range | 66–207 | 67–200 |
| Months | 135 | 144 |
| SD | 41 | 35.5 |

the UK, as stated on their Education and Health Care Plan (their EHCP).

*Materials*

A desktop or laptop computer with Internet access to HER was used. Dual headphone adapters ('headphone splitters') that allowed two headphone sets to be connected through to one audio source were also required. This allowed a supporting adult to listen in on the student's progress.

HER instructional components were (1) 80 online episodes (episodes) each of approximately 15–20 min in length, (2) Printable 'Sprout Stories' at the end of each episode, (3) 'Sprout Cards' (about 100 printable flash cards of sounds and words taught in the program) and (4) Progress Maps and Stickers (so students could mark off each completed episode). Licences for all students also allowed access to progress reports and further information on implementation (Headsprout Early Reading Teacher's Guide, 2010).

*Recruitment, randomisation and blinding procedure*

Fifty-seven students were identified by classroom teachers as meeting the inclusion criteria for the study. Parents of these students were sent a letter informing them about the aims of the study and about our intention to randomly allocate students into a HER intervention group or a waiting list control group receiving Reading as Usual. The study information sheet also explained to parents that students in the control group would receive the HER intervention during the following school year from teaching staff who would be trained by the research team during the course of the project. Fifty-five parents consented to their child's participation in the study, and following this, students were pre-tested on the standardised outcome measures. The pre-tests were carried out by the first, fifth and sixth authors.

A randomisation protocol was created for the needs of the study by the third author. The students attended 13 different classes within the school, and it was important for practical reasons that not too many students in any one class received the HER intervention at the same time. This was because the intervention required the students to be supported by the teaching assistants working in each class, and this would affect the staff-to-students ratio. Therefore, three balancing variables were used for randomisation: the class in which the students were enrolled, their gender and whether they had severe ID vs. severe ID with ASD. Students were allocated on a 1:1 basis to the two arms of the trial using the free-to-access Minim software that uses a dynamic allocation method to ensure balance between groups. Twenty-seven students were allocated to the HER intervention group and 28 to the control group. A consort style diagram summarising the design is presented in Figure 1.

At seven months post-randomisation (designed to be at the end of the planned intervention period), 52 students were again tested on the same standardised measures (3 students had left the school mid-way through the intervention period and were not available for the post-test). The post-tests were completed by master level trained students who had used the assessments in previous research studies. These assessors were not otherwise involved with the study, worked independently from the school and had no access to any intervention reports or data. Further, they were not informed of children's group status and so remained blind to allocation.

*Measures*

*Reading and early literacy skills.* The Dynamic Indicators of Basic Early Literacy Skills 6th Edition (DIBELS; Good and Kaminski, 2007) and the Word Recognition and Phonics Skills Test (WRAPS; Carver and Moseley, 1994) were used as outcome measures. These norm referenced instruments were chosen for their good psychometric properties and their use in published outcome studies on reading including with students with ID (e.g. Grindle et al., 2013; Tyler et al., 2015b). Measures were also selected on the basis that most could be accessed by students with limited vocal ability.

Students in this study had reading skills well below average for their age or no reading skills. Thus, all students were assessed using subtests from the DIBELS Kindergarten Benchmark Assessment. This assessment evaluates prerequisites for early reading that can be measured before a child learns to read, including a child's knowledge of letters and the awareness of speech sounds in words. The DIBELS Kindergarten assessment consists of five subtests: (1) the Initial Sound Fluency (ISF) measure that assesses the child's ability to identify and isolate the first sound of an orally presented word. This subtest does not require a vocal response as the assessor produces a sound and the child has to identify (either point to or say) which of four pictures begin with that sound; (2) the Letter Naming Fluency (LNF) measure that assesses the child's ability to label letter names (if the child produced the letter sound rather than the letter name this was scored as incorrect); (3) the Phonemic Segmentation Fluency (PSF) measure that assesses a child's ability to segment three and four phoneme words into their individual phonemes (e.g. being told the word 'mop' and being able to identify that the sounds in the word are 'm-o-p'); (4) the Nonsense Word Fluency (NWF) measure that assesses alphabetic principle skills (i.e. the ability to know that letters represent sounds in words and that letter sounds can be blended together to read/ decode words); and (5) the Word Use Fluency (WUF) measure that assesses expressive vocabulary skills (i.e. the ability to use words to convey a specific meaning for a particular label or word by putting a word into a short phrase or sentence). The

**Figure 1: Consort style diagram.**
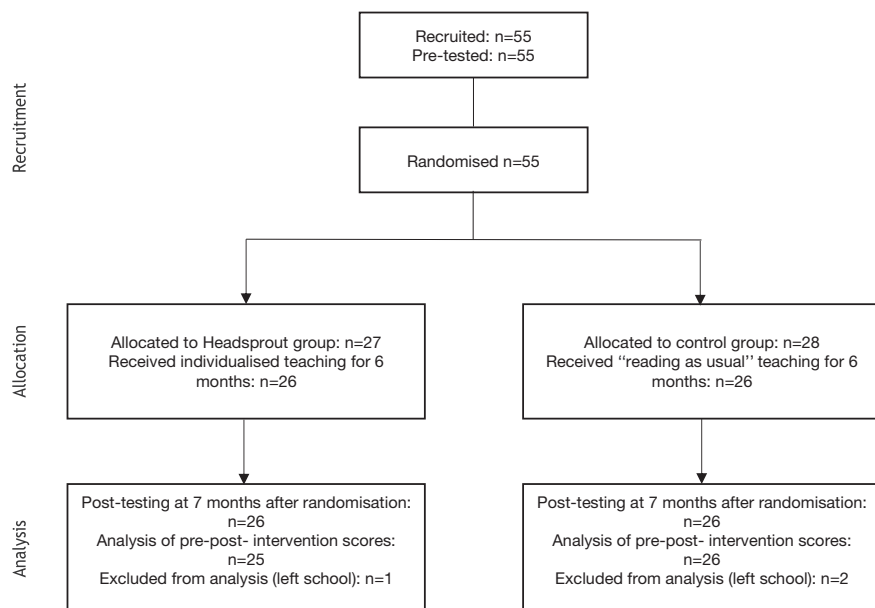


child's score for each subtest was how many correct responses they provided in one minute.

The LNF, the PSF and the NWF require some individual sound or single word verbal responses from children but do not require speaking in short phrases or sentences. These subtests were considered appropriate to use with the students as the response requirements matched the vocal ability eligibility criterion used. Although the WUF measure requires a student to put a named word into a short phrase or sentence (e.g. if told the word 'green', the student responds with 'grass is green' or 'green grass'), based on their knowledge of student vocal ability and understanding, class teachers confirmed that the students would be able to access this DIBELS sub test.

The Word Recognition and Phonic Skills assessment (WRAPS; Carver and Mosely, 1994) is a standardised measure that assesses students' developing word recognition skills and does not require a vocal response. The student is read a word, and the word is repeated again in a sentence (e.g. 'Orange', 'The Orange that we eat'). The student is then asked to choose the correct word in an array of five words. The student's total score is how many words, from 50, they can correctly identify. The WRAPS raw score was used. For students to achieve high scores on the WRAPS, they need to show that they can identify clusters and digraphs necessary for word recognition.

### Interventions

*'Reading as usual'.* Students in the control group had five 1-h session of literacy teaching per week as a part of the typical literacy instruction in the school (i.e. approximately 110 h of lesson time over the study period). Typically, each literacy lesson started with whole-class activities that focussed on teaching reading, writing and/or oral communication. The students were then divided into small groups (2–3 students) where they engaged in a variety of literacy activities (e.g. writing their name, sorting letter cards). Teachers embedded the teaching of reading into their literacy lessons in a variety of ways. Most included the teaching of reading into each daily literacy session, but a few teachers allocated specific times in the weekly timetable solely for teaching reading (e.g. one 1-h reading slot per week or 15 min every morning during timetabled 'Basic Skills' teaching). Teachers had autonomy in deciding how and when to teach reading to students in the school. The literacy co-ordinator in the school estimated that reading instruction occurred during at least half of the time designated for the teaching of literacy (i.e. approximately 55 h of lesson time), although no data were collected on this.

The school used a variety of different approaches to teach reading. For students in the elementary department, the 'Letters and Sounds' program (Department for Education, 2007) was used. The Reading for Pleasure curriculum (Atkinson, 2017; Department for Education, 2012) was used for children in both elementary and secondary departments. Finally, some students also followed the 'Symbols Making Sense' reading scheme. This scheme comprises more than 40 books, ranging from Introductory Book A (basic nouns in common use) to more complex sentences, social stories and non-fiction; all supported by simply drawn, colourful Widgit symbols that illustrate a single concept in a clear and concise way (e.g. Shepherd-Clarke, 2005).

*HER intervention group.* Students in the HER group received Reading as Usual with supplemental instruction using HER, or partial Reading as Usual with HER. Usually, the HER sessions were timetabled to occur

during the literacy lessons. On these occasions, students were taken out of the literacy lessons to receive HER, thus missing out on at least part of the usual lesson. When they had completed their HER session, however, they would rejoin the class and, if the literacy lesson was still ongoing, participate in the rest of the lesson.

During the HER sessions, the students were supported either by a teaching assistant who worked in the child's class, by their class teacher or by the 'learning mentor' in the school who had been allocated 1.5 days per week to help support the HER intervention and run sessions with the students. The literacy co-ordinator in the school was also allocated at least one half day per week on her time-table to help facilitate the intervention, although she rarely carried out 1:1 sessions with the students.

*Training.* Training the school staff to implement the HER intervention involved a number of steps. Initially, three 2-hr training sessions on the program were conducted by the first author with two staff members in the school (the literacy co-ordinator and the learning mentor) who had been allocated the role of *Headsprout coaches* in the school. It was intended that the Headsprout coaches be trained up to a high degree of expertise so that they could train other staff in the school on HER and offer ongoing support (using a 'train the trainer' model), thus helping to build long term sustainability of HER implementation in the school after the cessation of the study. Headsprout coaches received some ongoing mentoring and supervision throughout the study period from the first author. Meetings were held at least once every four weeks to discuss implementation issues, including strategies that might be helpful for individual children who were struggling to progress through the online episodes.

After the Headsprout coaches had received training, the first author and the coaches conducted a training session with all the teachers and teaching assistants in the school who were going to deliver the intervention (hereafter: instructors). This took place in a weekly 'twilight' after school training slot of 1-hr duration. During this training, (1) key features of the program were briefly described, (2) instructors were shown how to navigate the HER webpage, and (3) instructors were told how to effectively support students with the program (e.g. that children should do a minimum of three sessions a week, that students should 'speak out loud' when required to do so by the program). Instructors were given written resources summarising the training.

There were also opportunities for ongoing help and support throughout the study. Headsprout coaches scheduled frequent observations of HER teaching sessions to ensure that procedures were implemented in a consistent fashion. They provided feedback regarding how the instructors were supporting students and provided further advice

where necessary (e.g. how to deliver appropriate prompts to students). If children did not make expected progress, the Headsprout coaches discussed the issues with the first author in their regular meetings and requested that the first author carry out an observation to provide additional feedback. This was rarely considered necessary.

*Teaching procedure.* The teaching procedure adopted was similar to that described in Grindle et al. (2013) and Tyler et al. (2015b) Students had to demonstrate proficiency with basic computer skills such as clicking and dragging with the mouse as well as following simple instructions delivered by the computer, before beginning episode 1 of HER where these skills would be utilised. Progress data were collected automatically by HER, and a mastery criterion of 90% correct was required on each online episode before moving on to the next episode. After completing episodes, students were required to read the Sprout Story booklets without hesitation (as per the HER implementation guidelines) before students could progress with the online tuition. It was also emphasised during training that instructors could use the Sprout stories, the Sprout cards, and the progress maps and stickers to help support students' progress through the online episodes.

Some students struggled to stay on task for the 20-min duration of one online episode. For these students, episodes were divided into two sessions (one in the morning and one in the afternoon). Certain times of the day were also associated with higher rates of off-task behaviour for all students (e.g. the start or end of the school day, immediately before lunch), so episodes were not completed at these times. During each online episode, an instructor familiar to the child sat next to them. It was important that the data collected by the computer were based on the child's unprompted performance of reading ability, so any prompts delivered by the instructor predominantly consisted of reminders to attend to the computer screen or to speak out loud when required to do so by the program.

Reinforcement for correct responding was provided within the program in the form of 'gold stars' which could be traded in for cartoons or games on the HER website. Many students enjoyed these activities and did not require additional incentives. A few students required additional reinforcers to maintain their motivation for completing the episodes. Instructors were advised to use a token system, where tokens were given to the child as they engaged with the program and completed tasks. These could then be exchanged at the end of the episode for preferred items and activities.

*Checking administration of outcome measures*
All assessors exercised every caution to obtain reliable and valid data. All assessments were administered in a distraction free environment in the school (often using

school office space, rather than a classroom). For 31% of testing sessions (25% for WRAPS and 37% for DIBELS), a second assessor scored the child's performance. At baseline, first and second assessors were members of the research team; at post-test, they were working independent of school and also blind to group status. Agreement was calculated using a percentage agreement index method (i.e. total agreements divided by the number of agreements plus disagreements × 100%). Inter-scorer agreement averaged 99% (range 92–100%).

*Analysis procedure*
Data analysis compared the HER Intervention and Reading as Usual groups post-intervention, adjusting for baseline scores on the respective DIBELS and WRAPS outcome measures as well as the prognostic factors accounted for in the randomisation (gender, ASD or not, and classroom). Classrooms were included in all models as fixed effects (dummy-coded), due to a small number ($N = 13$) of classes (McNeish and Stapleton, 2016), a number of near-zero variances at the class-level when random intercepts were initially specified for classes, and an absence of predictors at the classroom level.

For each regression model, follow-up scores for the each of the respective outcomes were the independent variables. These outcomes were adjusted for baseline scores, alongside the three balancing variables from our randomisation procedure, that is gender, diagnosis (with or without ASD) and classroom. These regression models are analytically equivalent to analysis of covariance (ANCOVA) models when evaluating RCT outcomes. In this study, data were counts (with the exception of Initial Sound Fluency) and thus did not follow a normal (Gaussian) distribution. Instead, cases where the mean and variance of the distribution was roughly equal followed a Poisson distribution, whereas when variances were larger than the mean (overdispersion) a negative binomial model was applied (the Poisson and negative binomial models are equivalent in the absence of any overdispersion). A further correction can be made to this model to account for excess zeros (known as a zero-inflated model). Accordingly, the analysis approach offered greater flexibility than an ANCOVA model, to account for the different distributional characteristics of tested outcomes.

The type of regression models used in this study varied, based on the distribution of the outcome variables. Three of the outcome variables (namely Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency) were comprised of integer values ≥0, with overdispersion and a number of zero values. Accordingly, the appropriate model to assess these three outcomes was a zero-inflated negative binomial regression. The outcome variable Word Use Fluency also had integer values ≥0 and overdispersion but had fewer zero values and was therefore assessed with a negative binomial model (which does not account for zero-inflation). WRAPS raw scores (no overdispersion or zero integers) and Initial Sound Fluency (continuous, non-integer values, ranging from 0 to 22.5) outcomes were assessed with Poisson and linear regression models respectively.

Beta coefficients for all models are presented in their exponentiated form (i.e. as incident rate ratios – IRRs), with the exception of the linear regression analysis for Initial Sound Fluency. These IRRs quantify the percentage change in the counts for a one-unit change in the predictor, holding the other variables constant. Effect sizes for all models were estimated as standardised mean differences (SMD; cf. Coxe, 2018). These effect sizes are equivalent to Cohen's (1988), yet accounting for heteroscedasticity and overdispersion where appropriate, quantifying the change (in standard deviation units) in the outcome variable for every one-unit change in the predictor variable.

Across all outcomes, the maximum number of missing observations was three for the WRAPS raw score at six month follow-up (5.45% of observations for that variable), whilst five other variables had two missing observations at six month follow-up representing 3.64% of observations for each of the variables. Analyses were thus conducted based on the Intention to Treat (ITT) principle and complete case analysis.

## Results

*Adherence*
Our initial intent was for every child to have a minimum of three intervention sessions per week (as per HER recommendations). However, adherence to this proposed level of intervention intensity occurred for only five students (19% of sample). Seventeen students completed on average one episode a week (65% of sample), and four students completed on average two episodes a week (16%). Events such as school outings, staffing shortages or a child in the classroom requiring extra attention frequently resulted in teaching assistants and teachers not being able to find the necessary time to implement the HER intervention.

*Distribution of students into study arms*
As reported in Table 1, the randomisation protocol resulted in approximately equal groups with respect to gender and ASD. Ages (in months) for the Reading as Usual ($M = 144.25$ months) and HER group ($M = 135.41$ months) did not significantly differ ($t(53) = .856$, $p = .802$).

Overall, children in the study were from 13 classrooms, class sizes varied from nine students (1 class) to 2 students (3 classes), with a mean class size of 4.23 students (SD = 2.28). At least one child from each of the HER and Reading as Usual groups were in each class. A number of classes (7 classes) had a single student from the

HER group, although the number of children from the HER group in each class was as large as 5 (1 class). The number of children from the Reading as Usual group in each of the respective classes ranged from 1 (3 classes) to 5 (1 class).

*Between-group comparisons*

Outcome data at baseline and 6 months follow-up are presented in Table 2. Phoneme Segmentation Fluency scores at follow-up were 1.82 times higher in the HER group (Incidence Rate Ratio; IRR = 1.82, $p$ = .012) in comparison with the Reading as Usual group. Nonsense Word Fluency scores at follow-up were 2.27 times higher in the HER group (IRR = 2.27, $p$ = .006) when compared to the Reading as Usual group. Scores at follow-up did not significantly differ on any of the four other outcomes between the Reading as Usual and HER groups (all $p$'s < .05).

As can be seen in Figure 2, when comparing the HER and Reading as Usual groups, the largest effect size was observed for Phoneme Segmentation (Standardised Mean Difference; SMD = .72, 95% CI = .18–1.65), with Nonsense Word Fluency (SMD = .72, 95% CI = .14–1.92) also showing a moderate effect size. Effect sizes for the other four outcomes were small (SMD range = −.07–.29).

**Discussion**

Children with severe ID who received HER made gains compared to the Reading as Usual group on each of the domains measured by standardised reading assessments. This reflects the positive findings from the pilot RCT conducted by Roberts-Tyler et al., (2019). While statistically significant differences were observed on two of the six domains measured, the skills measured in these domains are pivotal for fluent reading. The first domain was Nonsense Word Fluency where participants were asked to read (decode) as many nonsense words as they could in one minute (e.g. pov, sig). This assessment focusses on use of the alphabetic principle (that letters represent sounds) and the ability to blend sounds together, indicating phonemic awareness (Kaminski and Good, 1996). The second domain was Phoneme Segmentation Fluency and, here, an even larger effect size was observed. This subtest involved segmenting three or four phoneme words into their individual phonemes (e.g. being told the word mop and being able to identify the sounds in the word are m-o-p). Given that HER focusses on teaching phonemic awareness as one of the key five reading components, it is particularly encouraging that this effect was observed. These results are similar to Tyler et al. (2015b) where the most notable outcomes were for Nonsense Word Fluency and Phoneme Segmentation Fluency on the DIBELS.

**Table 2: Between-group comparisons for primary and secondary outcomes**

|  | Reading as Usual | HER | Adjusted incidence rate ratio* | *p*-value |
|---|---|---|---|---|
| WRAPS Raw Score[1] |  |  |  |  |
| Baseline | 11.32 (6.93) | 14.37 (8.10) | – |  |
| After Intervention (6 months) | 12.04 (8.80) | 17.15 (11.23) | 1.14 (.974–1.33) | .104 |
| Letter Naming Fluency[2] |  |  |  |  |
| Baseline | 2.29 (4.26) | 6.41 (10.16) | – |  |
| After Intervention (6 months) | 3.04 (4.89) | 7.88 (9.15) | 1.51 (−.81 to 2.82) | .192 |
| Phoneme Segmentation Fluency[2] |  |  |  |  |
| Baseline | 3.18 (7.26) | 3.41 (9.13) | – |  |
| After Intervention (6 months) | 3.04 (5.81) | 6.88 (11.00) | 1.82 (1.21–2.75) | .012 |
| Nonsense Word Fluency[2] |  |  |  |  |
| Baseline | 13.25 (23.19) | 12.19 (12.10) | – |  |
| After Intervention (6 months) | 9.81 (14.31) | 14.04 (17.17) | 2.27 (1.26–4.08) | .006 |
| Word Use Fluency[3] |  |  |  |  |
| Baseline | 1.89 (4.72) | 2.96 (7.21) | – |  |
| After Intervention (6 months) | .65 (1.50) | 3.04 (8.53) | 1.45 (.23–9.04) | .687 |
| Initial Sound Fluency[4] |  |  | Adjusted mean difference* |  |
| Baseline | 4.65 (4.97) | 6.41 (6.50) | – |  |
| After Intervention (6 months) | 8.66 (14.04) | 10.50 (9.80) | −.82 (−6.40–4.76) | .768 |

Data are mean (SD) or mean (95% CI). [1]Analysis based on a Poisson regression model. [2]Analysis based on a zero-inflated negative binomial model. [3]Analysis based on a negative binomial model. [4]Analysis based on a linear regression model. *Models adjusted for baseline scores, gender, diagnosis (ASD, no ASD), and classroom.
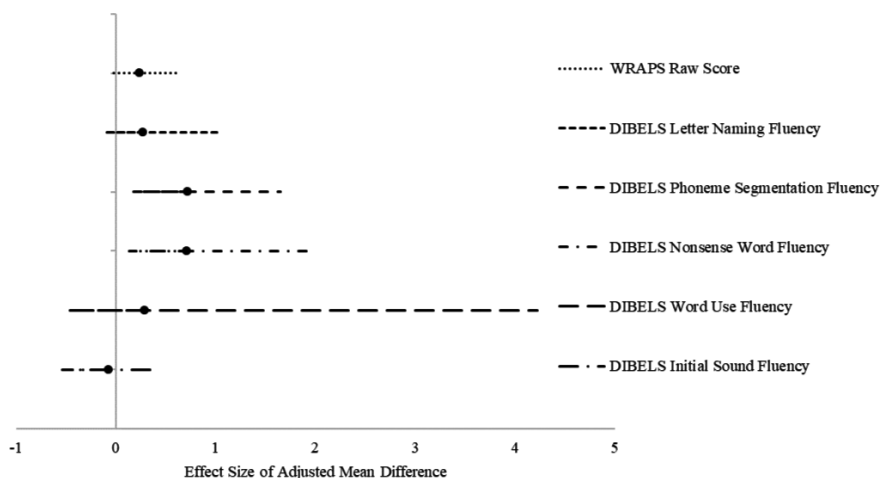
**Figure 2: Effect size estimates and 95% confidence intervals post-intervention. The forest plot shows effect size estimates for all main outcomes of the study. The effect sizes are based on standardised mean differences between the HER and Reading as Usual group, adjusted for baseline scores, gender, diagnosis (with or without ASD) and classroom.**

Although non-significant differences were found for four of the six outcomes (WRAPS Raw Score, Letter Naming Fluency, Word Use Fluency, Initial Sound Fluency), smaller effect sizes may have gone undetected within this study due to a lack of statistical power. Post hoc power calculations (based on a type 1 error rate of 5%, type 2 error rate of 20%) suggested that this study was adequately powered to detect an effect size of 0.80 or larger. Thus, a larger HER RCT is required in future.

The participants in this study had severe ID or ASD with severe ID. As noted by Browder and Spooner (2011) reading interventions for this group had not previously incorporated the components of reading instruction identified by the National Reading Panel (2000). The current study addressed this gap in the literature by evaluating an intervention that does incorporate the recommended components, moving away from a focus on sight word instruction. This study also makes an important contribution to the extant research into using HER with diverse populations, contributing to the evidence base of studies that have included a RCT of HER with children with ID. We have also demonstrated the feasibility of conducting RCTs in special education settings (including obtaining parental consent, maintaining blind post-testing etc.).

Another important aspect of this study is the fact that it is one of the first studies to demonstrate the potential effectiveness of HER with children with ID using the standard intervention (episodes and stories alone, with little specialist support). It is unclear whether incorporating additional support elements (e.g. as per Grindle et al., 2013; Tyler et al., 2015b) would have resulted in greater improvements for the HER group. Although we successfully completed initial training and delivered a coaching support model for instructors, the recommended number of HER sessions was not completed for most of the participants in the HER group. Greater gains may be possible under conditions where participants complete the recommended three sessions per week.

There were a number of other limitations in the current study. First, measures of cognitive functioning were not implemented so the degree of impairment of participants and how this relates to outcomes on following HER is unknown. However, the groups were balanced in terms of the numbers of participants who also had ASD. No fidelity data were recorded during the initial training that was delivered either to the Headsprout coaches or to the instructors. Nor were there fidelity data collected during the implementation of the intervention. Thus, we cannot be certain that participants did not receive additional prompts or that the read aloud sections were accurate. The automated nature of HER means that the opportunities for staff to make mistakes are limited; nonetheless, future research should place more of an emphasis on measuring fidelity across all stages of the intervention.

Further research is needed to test the effectiveness (and ideally cost-effectiveness) of using HER with children with ID, especially in a larger RCT where the design and methods take into account the learning from now several smaller-scale implementation studies including the current RCT. In particular, the current RCT was carried out in a single school. Future research is needed that adopts a cluster randomised design in which larger numbers of children are involved and schools (or potentially classrooms across schools) are the unit of randomisation.

*Address for correspondence*
Corinna F. Grindle,
Centre for Educational Development,
Appraisal and Research (CEDAR),
New Education Building,
Westwood Campus,
University of Warwick,
Coventry CV4 7AL,
UK
Email: c.grindle@warwick.ac.uk.

# References

Allor, J. H., Mathes, P. G., Roberts, J. K., Jones, F. G. & Champlin, T. M. (2010) 'Teaching students with moderate intellectual disabilities to read: an experimental examination of a comprehensive reading intervention.' *Education and Training in Autism and Developmental Disabilities*, 45, pp. 3–22.

Atkinson, P. (2017) Promoting Reading for Pleasure. [online] UKLA. <https://ukla.org/resources/details/promoting-reading-for-pleasure> (Accessed 25 January 2019).

Austin, P. & Boekman, K. (1990) *Edmark Functional Word Series*. Redmond, WA: PCI Educational.

Baranek, G. (2002) 'Efficacy of sensory and motor interventions for children with autism.' *Journal of Autism and Developmental Disorders*, 32, pp. 397–422. https://doi.org/10.1023/A1020541906063

Browder, D. M., Gibbs, S., Ahlgrim-Delzell, L., Courtade, G. R. & Lee, A. (2007) *Early Literacy Skills Builder*. Verona, WI: Attainment.

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, pp. 269–82. https://doi.org/10.1177/0741932508315054

Browder, D. M. & Spooner, F. (2011) *Teaching Students with Moderate and Severe Disabilities*. New York: The Guilford Press.

Browder, D. M., Wakeman, S. Y., Spooner, F., Ahlgrim-Delzell, L. & Algozzinexya, B. (2006) 'Research on reading instruction for individuals with significant cognitive disabilities.' *Exceptional Children*, 72 (4), pp. 392–408.

Carver, C. & Moseley, D. (1994) *Group or individual diagnostic test of Word Recognition and Phonic Skills (WRAPS): Manual*. London: Hodder Arnold H&S.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences* (2nd edn). Hillsdale, NJ: Erlbaum.

Coxe, S. (2018) Effect size measures for nonlinear count regression models. Poster presented at the ASA Conference of Statistical Practice, Portland, Oregon.

Department for Education (2007) Letters and Sounds: Principles and Practice of High Quality Phonics, Primary National Strategy. [online] gov.uk. <https://www.gov.uk/government/publications/letters-and-sounds-principles-and-practice-of-high-quality-phonics-phase-one-teaching-program> (Accessed 25 January 2019).

Department for Education (2012) Research evidence on reading for pleasure. [online] London, pp. 3–7. <https://www.gov.uk/government/publications/research-evidence-on-reading-for-pleasure> (Accessed 25 January 2019).

Good, R. H. & Kaminski, R. A. (2007) *Dynamic Indicators of Basic Early Literacy Skills* (6th edn). Eugene: Institute for Development of Educational Achievement.

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*, 28, pp. 203–24. https://doi.org/10.1002/bin.1364

Headsprout (2010) Headsprout Early Reading Teacher's Guide. <https://issuu.com/headsprout/docs/her_teacher_guide/43> (Accessed 25 January 2019)

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, pp. 279–98. https://doi.org/10.1080/10824669.2010.532415

Kaminski, R. A. & Good, R. H. (1996) 'Toward a technology for assessing basic early literacy skills.' *School Psychology Review*, 25 (2), pp. 215–27.

Layng, J., Twyman, J. & Stikeleather, G. (2004) 'Selected for success: How Headsprout reading basics teaches beginning reading.' In D. J. Moran & R. W. Malott (eds), *Evidence-based Educational Methods*, pp. 171–97. San Diego: Elsevier Academic Press.

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), pp. 162–72.

Mathes, P. G. & Torgesen, J. K. (2005) *Early Interventions in Reading Level 1*. Columbus, OH: SRA/McGraw-Hill.

McNeish, D. & Stapleton, L. M. (2016) 'Modeling clustered data with very few clusters.' *Multivariate behavioral research*, 51 (4), pp. 495–518. https://doi.org/10.1080/00273171.2016.1167008

Nally, A., Healy, O., Holloway, J. & Lydon, H. (2018) 'An analysis of reading abilities in children with autism spectrum disorders.' *Research in Autism Spectrum Disorders*, 47, pp. 14–25. https://doi.org/10.1016/j.rasd.2017.12.002

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.

O'Sullivan, D. V., Grindle, C. F. & Hughes, J. C. (2017) 'Teaching early reading skills to adult offenders with intellectual disability using computer delivered instruction.' *Journal of Intellectual Disabilities and Offending Behavior*, 8, pp. 122–31.

Pennington, R. C. (2010) 'Computer-assisted instruction for teaching academic skills to students with autism spectrum disorders: a review of literature.' *Focus on Autism and other Developmental Disabilities*, 25, pp. 239–48.

Ratz, C. & Lenhard, W. (2013) 'Reading skills among students with intellectual disabilities.' *Research in developmental disabilities*, 34 (5), pp. 1740–8. https://doi.org/10.1016/j.ridd.2013.01.021

Roberts-Tyler, E. J., Hughes, J. C. & 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, pp. 14–26.

Shepherd-Clarke (2005) Symbols Making Sense, Introductory book A. Just Special Marketing.

Silver, M. & Oakes, P. (2001) 'Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others.' *Autism*, 5, pp. 299–316.

Tucker Cohen, E., Wolff Heller, K., Alberto, P. & Fredrick, L. D. (2008) 'Using a three-Step decoding strategy with constant time delay to teach word reading to students with mild and moderate mental retardation.' *Focus on Autism and Other Developmental Disabilities*, 23 (2), pp. 67–78. https://doi.org/10.1177/1088357608314899

Twyman, J. S., Layng, J. & Layng, Z. R. (2012) Mimiosprout: The power of adaptive online learning to teach reading. <http://130.17.111.21/bitstream/handle/011235813/1162/BurnsE%20Summer%202017.pdf?sequence=1>

Twyman, J. S., Layng, T. V. J. & Layng, Z. R. (2011) 'The likelihood of instructionally beneficial, trivial, or negative results for kindergarten and first grade learners who complete at least half of Headsprout Early Reading.' *Behavioral Technology Today*, 6, pp. 1–19.

Tyler, E. J., Hughes, J. C., Beverley, M. & Hastings, R. P. (2015a) 'Improving early reading skills for beginning readers using an online program as supplementary instruction.' *European Journal of Psychology of Education*, 30, pp. 281–94. https://doi.org/10.1007/s10212-014-0240-7

Tyler, E. J., Hughes, J. C., Wilson, M. M., Beverley, M., Hastings, R. P. & Williams, B. M. (2015b) 'Teaching early reading skills to children with intellectual and developmental disabilities using computer-delivered instruction: a pilot study.' *Journal ofInternational Special Needs Education*, 18 (1), pp. 1–11. https://doi.org/10.9782/2159-4341-18.1.1

Waugh, R. E., Fredrick, L. D. & Alberto, P. A. (2009) 'Using simultaneous prompting to teach sounds and blending skills to students with moderate intellectual disabilities.' *Research in Developmental Disabilities*, 30, pp. 1435–47. https://doi.org/10.1016/j.ridd.2009.07.004