



Evaluation of a School-Based Headsprout Intervention for Improving Literacy

Gerry McWilliams^{1,2}  · Julian C. Leslie¹ · Claire McDowell¹

Accepted: 5 August 2022
© The Author(s) 2022

Abstract

Headsprout Early Reading is a computer-based program designed on behavioral principles to enhance the basic skills that underpin the initial development of reading. In a within and between groups design, and using primary schools within Northern Ireland that had a currently high proportion of disadvantaged pupils, children who were behind their peers in progress with reading were randomly allocated to an intervention group ($n=79$), where the target was to work through 80 reading training episodes within a school year, or a teaching as usual group ($n=44$). Reading skills were assessed in all children before, at the midpoint, and after the intervention using a flashcard-based phonics identification test with three levels of difficulty, and before and after intervention using a standardized reading assessment, which generated a sentence reading age and a phonics reading age. Both groups showed increased scores on all measures over the 6 months of the study, but the intervention group showed markedly greater improvement. Importantly, the mean scores on sentence reading age and phonics reading age for the intervention group increased by over 17 months and 12.1 months, respectively, as opposed to 7.6 months and 7.8 months with the control group. These findings also validated the use of the flashcard-based phonics identification test with this population. This study indicates that widespread use of Headsprout Early Reading in mainstream education could be highly effective.

Keywords Headsprout early reading · Disadvantaged children · Sentence reading age · Phonics reading age · Mainstream education

Reading is one of the most critical academic skills that children acquire. Early reading skills predict later reading ability, language ability and general knowledge

✉ Gerry McWilliams
mcwilliams-g2@ulster.ac.uk

¹ School of Psychology, Ulster University, Coleraine, Northern Ireland, UK

² Massreagh, Kilmacrennan, Co. Donegal, Ireland

(Rigney et al., 2020). In contrast, illiteracy costs the UK economy £36 billion/year in lost earnings, benefit payments and increased medical costs (Hazell, 2018), there are also links between low literacy and depression, obesity, and life expectancy (Griggs & Walker, 2008). One societal group at greater risk of experiencing reading difficulties are children from disadvantaged backgrounds. Such children are more likely to live in cramped housing where they share rooms, are more likely not to have eaten before attending school, are less likely to have completed their homework and suffer from poorer attendance than their peers (Kellest & Dar, 2007). They are also more likely to engage in challenging behavior, aligned with poorer concentration and are more likely to be bullied for being poor (Thompson, 2020). Disadvantaged children are therefore more likely to suffer barriers to learning, and one result of this is a higher level of literacy difficulties. When people with poor literacy have children, there is likely to be a cycle of illiteracy whereby they are unable to help with homework, more likely to be unemployed, more likely to encounter social isolation, and suffer from lack of self-esteem and poor health (Hirsch, 2007).

In general, Northern Ireland (NI) has a lower level of educational attainment than the rest of the United Kingdom (UK). It also has a high number of people living in poverty: in 2014, around 17% of the population (370,000) lived in relative poverty of whom 110,000 were children (Joseph Rowntree Trust, 2018). Furthermore, there is a large gap in educational attainment between richer and poorer children, and NI had a consistently higher number of adults leaving education without qualifications compared to the rest of the UK (17% compared to 9% in 2014) (Barnard, 2018).

While there is unanimous agreement between educators and researchers that targeting literacy skills at an early age is crucial to closing the literacy gap between disadvantaged children and their non-disadvantaged peers, the most effective method of doing so is not agreed. To counter the impact of disadvantage, UK primary schools receive (since 2011) increased funding for every pupil who has been registered for free school meals at any time in the last 6 years. This money must be spent on activities designed to raise the attainment level of these disadvantaged pupils. Using these funds, many schools provide additional literacy activities in addition to regular, teacher-delivered classroom instruction, and invest in additional resources, including computer technology and equipment in efforts to support the pupils who need it most. However, increased school resources do not necessarily result in improved pupil performance (Hanushek., 1997), and this may be indicative of the lack of a systematic, evidence-based approach to the purchase and use of educational resources. Levačić and Vignoles (2002) argue that until better empirical evidence on the impact of using resources in different ways becomes available, it is difficult to provide guidance to head teachers on how best to allocate their resources.

Computer-assisted instruction (CAI) can be used to provide an additional means of support to teach beginning reading skills, with most programs designed to supplement teacher instruction (Blok et al., 2002). Headsprout Early Reading (HER), the intervention reading program used in this study, and Headsprout Comprehension are two CAI programs developed in a rigorous manner based on the scientific principles of behavior analysis (Layng et al., 2003, 2004, 2011). There is an emerging research base that shows HER to be useful for teaching students key foundational reading skills which match those identified as critical by the National Reading Panel (2000). HER

is designed to bring users to a proficient level of reading in 80-, 20-min episodes. Previous studies of the impact of HER on literacy performance (Huffstetter et al., 2010; Twyman et al., 2011) indicated that partial completion of HER still resulted in significant literacy gains, especially for completion of the first 23 episodes which concentrate on mastery of basic phonemic skills. HER has been evaluated in both school and home settings and with diverse populations of children, including typically developed children in home and school settings, those with intellectual disabilities, those with Autistic Spectrum Disorder and those from a care background (Layng et al., 2003, 2004; Clarfield & Stoner, 2005, Whitcomb et al., 2011; Grindle et al., 2013; Huffstetter et al., 2010, Tyler et al., 2015; Storey et al., 2017, 2020; Nally et al., 2021). However, a systematic review (Rigney et al., 2020) concluded that more rigorous research is required to support the widespread adoption of HER in mainstream schools, with publications to date showing only tentative support for effectiveness with this population. Rigney et al. suggested that future research should focus on school-based evaluations, employ more rigorous research designs, report effect sizes, and where possible use randomized assignment.

This study therefore aimed to extend the current evidence base of HER and fill a gap in knowledge by evaluating the impact on literacy skills of disadvantaged primary school children who were falling behind their peers in mainstream NI primary schools. Objectives were to investigate whether the addition of HER to existing school-based literacy instruction could bring about significant improvement in literacy skills, with particular focus on phonemic awareness, as well as word and sentence reading age. In a within and between groups design, children were randomly allocated to an intervention group, where the target was to work through 80 reading training episodes over 24 weeks of a school year, or a teaching as usual group. Reading skills were assessed in all children before and after intervention using a standardized reading assessment. This generated a sentence reading age and a phonics reading age. An additional assessment was carried out on pupils before, at the midpoint and after the intervention using a phonics identification test which contained three levels of difficulty. This was devised for this study in order to provide a measure that could show high and low rates, and could be applied several times during the study. It used sets of flashcards presented in one-minute timings. There were three levels of difficulty in the phonemes presented, and a correct and incorrect score was generated by each flashcard test. Ecological validity and sustainability of the intervention were planned to be enhanced by recruiting an HER coordinator at each participating school who was responsible for the children in the intervention group accessing and working through their HER episodes.

Existing research into the impact of HER has been carried out in various cohorts. This research is the first to look at the impact of HER on the literacy performance of primary school children in mainstream schools who are falling behind the expected reading level.

Methods

Participants and Design

Participants ($n=123$) were recruited for inclusion in the study if they met the following criteria:

- i. They were attending primary school with a higher-than-average uptake of free school meals at the beginning of the school year.
- ii. They were availing of free school meals at the beginning of the school year.
- iii. They had a reading age at least 1 year lower than their chronological age based on the results of the last annual literacy tests.

Each of eight participating schools was asked to select at least 15 potential participants who met the inclusion criteria.

Participants

As participants were under 18 years old, information and consent forms were distributed to their parents/guardians for completion. Schools were randomized to an intervention group or a waiting list control group at the pre-intervention stage by the process of simple randomization. This resulted in five schools (Schools 1–5) in the treatment group ($n=79$ age ranges from 5 to 9 years, 43 females and 36 males, average age 90 months) and three schools (School 6–8) in the control group ($n=44$, age range from 5 to 9 years, 21 females and 23 males, average age, 88 months.).

A repeated measure, between and within group design was used to evaluate the impact of engagement with the HER program on the literacy performance of disadvantaged primary school children. This was compared to the impact of teaching as usual. All participants were assessed before intervention, at midpoint of the HER intervention (approximately 12 weeks later) and post-intervention (approximately 26 weeks after the first assessment).

Measures

These were the Phonics Early Reading Assessment (PERA), a reading test providing a sentence reading age and a phonics reading age developed by McCarty and Ruttle (2012); and the Flashcard Identification Test (FIT), a bespoke phonics identification test which tested individual level of fluency for each pupil.

The PERA is a standardized phonics and early reading assessment which is based on the requirement to read a series of pre-determined words and non-words. PERA takes approximately 10 min to administer and tests phonic knowledge and fluency. It consists of two tests. The first test administered was the sentence reading age (SRA) assessment. Pupils were asked to read a short story from a pre-designed double-sided card which contained 50 words. No indication was given to the pupil

if answers given were correct or incorrect. The words were scored as pronounced correctly or incorrectly. The word on which the pupil made their fifth error indicated their sentence reading age in years and months. The second PERA assessment used was the phonics reading age (PRA) test. Pupils were asked to read 50 words spread over three sides of pre-designed A4 cards. A combination of real and nonsense words was used. There was no time limit set to read each card. The total number of correct answers provided by a pupil out of 50 was cross-referenced to a scoring table of established norms which provided a phonics reading age in years and months.

The FIT was designed to measure a participant's fluency in identifying the relationships between graphemes (written letters) and phonemes (spoken sounds) for the 44 phonemes that make up the English language. These 44 phonemes are represented by combinations of the 26 letters of the alphabet and are taught as part of the curriculum in NI primary schools by the end of the third primary school year. None of the phonemes in the FIT assessment should have been novel to any participants. Headsprout incorporates these 44 phonemes into the 80 episodes of its teaching system. Each of the three FIT tests contained 50 cards, and pupil's performance on seeing and saying the sound combinations were timed for one minute. Each 1-min timing provided a correct and incorrect rate of responding on the set tested. The level of difficulty of the phonemes used increased from FIT1 to FIT2 and again from FIT2 to FIT3. In each test, flashcards were held up in front of the pupil over a 60-s period, timed via a stopwatch app on a mobile phone. The pupils were asked to say the letters sound rather than the letter names on each card. The pack of assessment cards were held in the researcher's right hand. Each card was turned over and held up in front of the pupil by the researcher's left hand where both the pupil and a research assistant (RA) could see it. Each RA was blind as to which schools and pupils were part of the treatment and control groups. The RA sat beside the pupil opposite the researcher so they could score see the card and score the response. The researcher placed the card in a correct/incorrect pile based on the pupil's answer. Correct and incorrect answers were tallied at the end of each 1-min timing, and the researcher's scores (e.g., "21-4" with the first number the correct answers and the second the incorrect/no answer given score) were compared with the RA's scores to calculate inter-observer agreement. During the test, no verbal feedback was delivered. Copies of the FIT tests are available on application to any of the authors.

Procedure

Pupils in the treatment group had access to a Laptop, PC, iPad/Tablet with headphones and a wired/Wi-Fi internet connection to access the HER program. The school HER coordinators were present, while pupils used the program. Treatment schools committed to arranging for four 30-min sessions per pupil per school week at the outset of the research. HER sessions were carried out at the five schools in the treatment group in computer suites, classrooms and corridors based on the available resources each day in each school. Each school HER coordinator was responsible for timetabling HER on a weekly basis then and informing relevant staff of the days and times sessions were scheduled for pupils.

Assessments and interventions were conducted as following in all eight schools:

1. Assessment period 1: Baseline performance assessment using PERA and FIT measures. The HER intervention then began in the five treatment group schools immediately following this assessment. HER sessions ran over a 6-month period from January to June. Pupils were expected to complete one lesson or episode per session. HER coordinators interacted with pupils only to help with technical difficulties as all instructions are contained within the HER program. While pupils in the treatment group were receiving the HER intervention, pupils in the control group received teaching as usual (TAU). The TAU content and methods used varied from school to school but generally involved phonics training using programs such as Jolly Phonics (Lloyd & Wernham, 2005), teacher word recognition tasks and sentence reading using the student's current classroom reading book.
2. Assessment period 2: This occurred at the midpoint of the intervention, around 12 weeks after Assessment period 1, and only the FIT measures were used.
3. Assessment period 3: Post-intervention assessment occurred once HER intervention ended, using the PERA and FIT measures. This occurred about 26 weeks after Assessment period 1.

Pupils were given clear verbal instructions of what they were being asked to do for each section of the assessment. Short breaks were offered to any pupils who required them during the assessment.

Inter Observer Agreement (IOA) A research assistant observed and recorded all student responses during all assessments. IOA was calculated by dividing the number of observer agreements by the number of judgments. IOA was above 98% for all baseline, midpoint and post-intervention assessments. All data collected from the FIT assessments were subject to IOA. No formal review of procedural integrity was carried out as computer-based instruction was used. This monitored student progress and ensured lessons were completed in the correct order.

Program Completion

All participants completed at least the first 23 episodes of HER which focus on mastery of basic phonemic skills. The average number of episodes completed within the treatment group was 49 at an average completion time of 31 min per episode. Pupils completed an average of approximately 2 episode per week over the 24 weeks of the study.

Results

Of the 123 pupils in the study, 27 achieved baseline PERA sentence/phonics reading age scores at Assessment 1 that were within 12 months of their chronological age. The results for these pupils were not included in the PERA results analyses; for

those analyses, $n = 96$. The FIT assessment results were analyzed for all participants, $n = 123$.

Figure 1 (upper panel) shows the group mean SRA scores taken at baseline and post-intervention for the pupils in the treatment group ($n = 54$) and the control group ($n = 37$). There was a marked increase in mean SRA of 17 months from 61.5 months to 78.5 months for the treatment group. There was also an increase in mean SRA for the control group of 7.6 months from 59.5 months to 67.1 months. A summary of all ANOVAs carried out is shown in Table 1. In each of these, there was one within-subjects factor, time of testing, and one (group) or two (group, and FIT level) between-subjects factors. As shown in the table, very many effects reported were significant with a large majority also having a large effect size. For SRA, there was main effect of group, and of time of testing, with interaction between these two factors. Independent sample t tests showed no significant difference in SRA between the treatment group and the control group at baseline, $t(93) = 2.193$, NS, but a significant post-intervention difference in SRA between treatment and control groups,

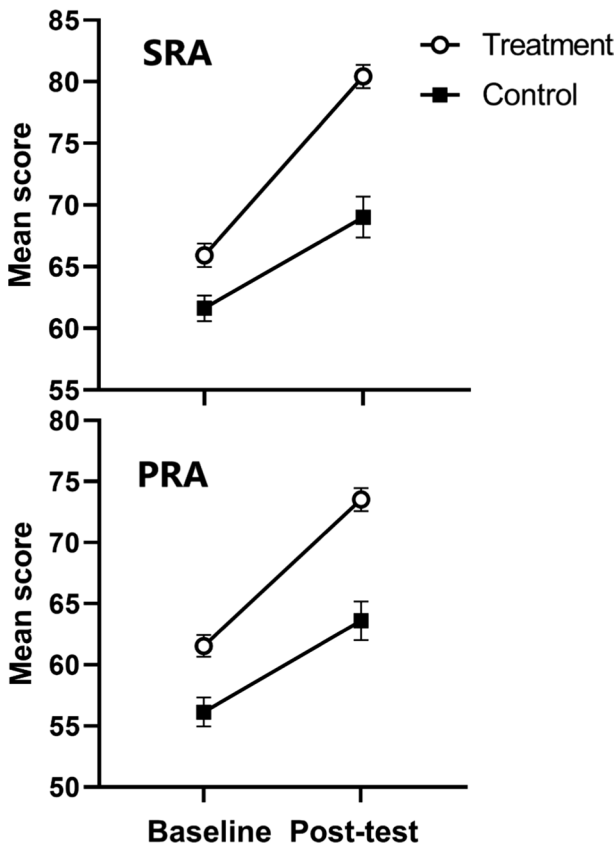


Fig. 1 Group means (\pm SE) in months of sentence reading ages (upper panel) and phonic reading ages (lower panel) at baseline and at post-intervention test

Table 1 Results of ANOVAs

Variable analyzed	Factor	Degrees of freedom	F value	Significance level	Effect size
SRA	Group	198	28.8	***	0.24
	Time	189	280.0	***	0.76
	Group × time	189	40.3	***	0.31
PRA	Group	1110	25.8	***	0.19
	Time	1110	267.9	***	0.70
	Group × time	1110	12.8	***	0.14
Correct FIT measures	Group	1112	18.4	***	0.14
	Time	2224	685.5	***	0.86
	Level	2224	260.9	***	0.70
	Group × time	2224	19.4	***	0.14
	Group × level	2224	18.2	***	0.47
	Time × level	4448	31.3	***	0.22
	Group × time × level	4448	20.7	***	0.16
FIT1-correct	Group	1112	25.5	***	0.19
	Time	2224	226.9	***	0.67
	Group × time	2224	41.5	***	0.27
FIT2-correct	Group	1112	17.4	***	0.13 ^m
	Time	2224	200.2	***	0.64
	Group × time	2224	6.35	**	0.05 ^m
FIT3-correct	Group	1112	7.12	**	0.02 ^s
	Time	2224	101.9	***	0.48
	Group × time	2224	1.86	NS	
Incorrect FIT measures	Group	1112	18.8	***	0.14
	Time	2224	45.6	***	0.29
	Level	2224	679.8	***	0.86
	Group × time	2224	10.7	***	0.09 ^m
	Group × level	2224	0.31	NS	
	Time × level	4448	1.46	NS	
	Group × time × level	4448	3.20	NS	
FIT1-incorrect	Group	1112	21.1	***	0.20
	Time	2224	238.1	***	0.81
	Group × time	2224	15.5	***	0.24
FIT2-incorrect	Group	1112	17.3	***	0.13 ^m
	Time	2224	41.1	***	0.38
	Group × time	2224	4.57	NS	
FIT3-incorrect	Group	1112	7.12	**	0.06 ^m
	Time	2224	17.7	***	0.14
	Group × time	2224	6.86	***	0.06 ^m

Effect sizes are calculated as partial eta-squared. All effect sizes shown are large except those marked^m or ^swhich are moderate or small, respectively. Significance levels: ** $p < 0.01$; *** $p < 0.001$

$t(89)=5.98$, $p<0.001$, $d=1.31$, a large effect size. Because both groups showed rises in mean SRA scores between the two times of testing, differences scores (post-intervention SRA–baseline SRA) were calculated and compared between groups with an independent samples t test. The mean increase for pupils in the intervention group (17.0, $SD=6.75$) was significantly greater than for the control group (7.65, $SD=7.12$), $t(88)=6.28$, $p<0.001$, $d=1.36$, a large effect size.

Figure 1 (lower panel) shows the group mean PRA scores taken at baseline and post-intervention for the treatment group ($n=71$) and the children who had teaching as normal in the control group ($n=41$). As seen in the figure, there was a marked increase in mean PRA of 12.1 months from 61.3 months to 73.4 months for the pupils in the treatment group. There was also an increase in mean PRA for pupils in the control group of 7.8 months from 55.8 months to 63.6 months ($n=39$). As shown in Table 1, there was a main effect of group, of time of testing, with an interaction between these two factors. Independent sample t tests showed a significant difference in PRA between treatment and control groups at baseline, $t(119)=3.54$, $p<0.01$, $d=0.55$, a moderate effect size, and a significant difference in PRA between the treatment and control groups post-intervention, $t(119)=5.049$, $p<0.001$, $d=1.11$, a large effect size. Because both groups showed rises in mean PRA scores between the two times of testing, differences scores (post-intervention PRA–baseline PRA) were calculated and compared between groups with an independent samples t test. The mean increase for pupils in the intervention group (12.2, $SD=5.43$) was significantly greater than for the control group (7.80, $SD=7.36$), $t(110)=3.13$, $p<0.01$, $d=0.82$, a large effect size.

The FIT measures were conducted at baseline, midpoint and post-intervention. Mean scores for treatment and control groups of correct responses at these three time points for FIT1, FIT2 and FIT 3 are shown in Fig. 2. As is evident from the figure, all three measures increased with time, but these increases were more marked in the treatment group. These effects were confirmed by statistical analysis. As shown in Table 1, there was a main effect of group, of time of testing, and of FIT level, with two- and three-way interactions between these factors. By a least significant differences test, scores on FIT1 were significantly higher than those on FIT2 which were higher than those on FIT3. All other differences were explored by carrying out separate two-way ANOVAs for each FIT level. For the FIT1 measure, Table 1 shows there was a main effect of group, of time of testing, with an interaction between these two factors. Independent sample t tests showed no significant difference in FIT1 scores between the treatment group and the control group at baseline $t(119)=0.62$, NS, but a significant difference at the midpoint test, $t(116)=6.11$, $p<0.001$, $d=1.58$, a large effect size and at post-test, $t(115)=6.04$, $p<0.001$, $d=1.62$, a large effect size. For the FIT2 measure, there was a main effect of group, of time of testing, with an interaction between these two factors. Independent sample t tests showed no significant difference in FIT2 scores between the treatment group and the control group at baseline $t(119)=2.59$, NS, but a significant difference at the midpoint test, $t(116)=4.18$, $p<0.001$, $d=1.19$, a large effect size and at post-test, $t(115)=4.32$, $p<0.001$, $d=1.27$, a large effect size. For the FIT3 measure there was a main effect of group, and of time of testing, with no interaction between these two factors. Independent sample t tests showed no significant difference in FIT2 scores

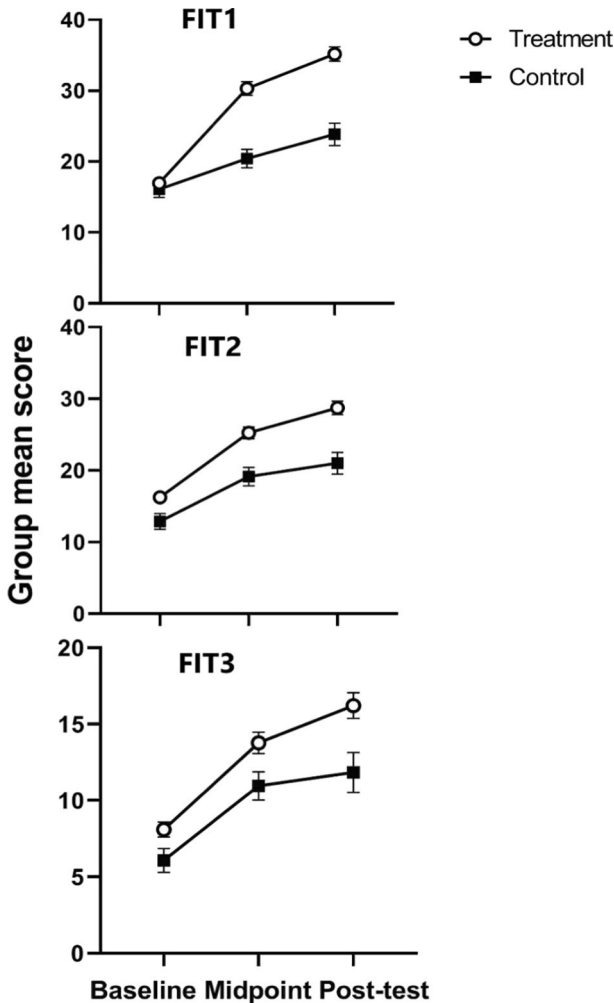


Fig. 2 Group mean (\pm SE) correct scores on Flashcard Identification Tests, FIT1 (upper panel), FIT2 (middle panel), and FIT3 (lower panel) at baseline mid-point and at post-intervention test

between the treatment group and the control group at baseline $t(119)=2.24$, NS, or at the midpoint test, $t(116)=2.44$, NS, but a significant difference at post-test, $t(115)=2.82$, $p < 0.01$ $d = 0.95$, a large effect size.

Mean scores for treatment and control groups of incorrect responses at the three time points for FIT1, FIT2 and FIT 3 are shown in Fig. 3. As is evident from the figure, all three measures decreased with time, but these decreases were more marked in the treatment group. These effects were confirmed by statistical analysis. As shown in Table 1, there was a main effect of group, of time of testing, and of FIT level, and an interaction between time and group, but no other interactions. By a least significant differences test, scores on FIT1 were lower than those on

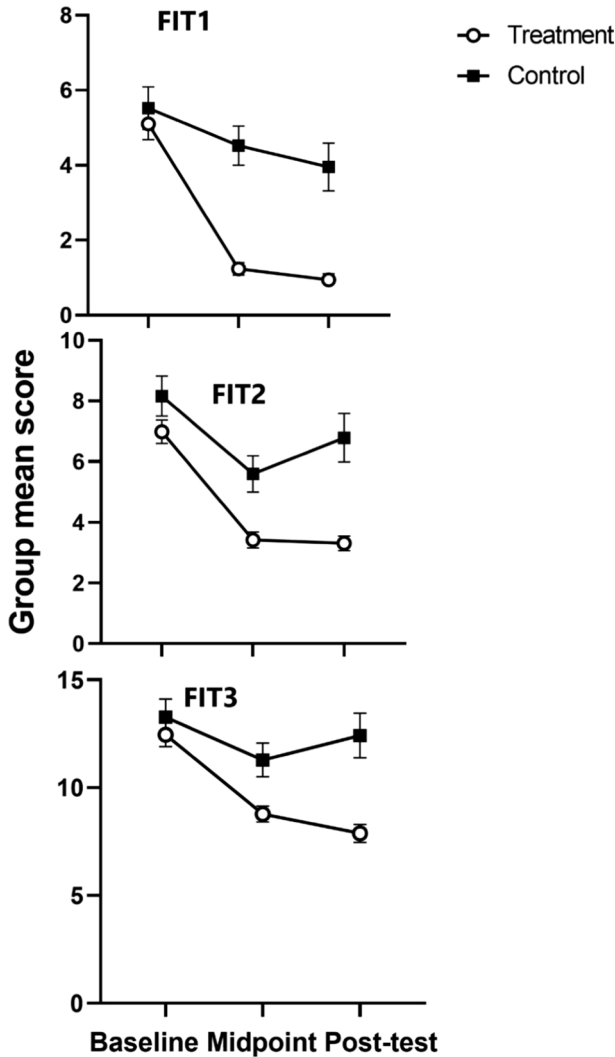


Fig. 3 Group mean (\pm SE) incorrect scores on Flashcard Identification Tests, FIT1 (upper panel), FIT2 (middle panel), and FIT3 (lower panel) at baseline mid-point and at post-intervention test

FIT2 which were lower than those on FIT3 ($p < 0.001$ in each case). All other differences were explored by carrying out separate two-way ANOVAs for each FIT level. For the FIT1 measure, there was a main effect of group, of time of testing, and an interaction between these two factors. Independent sample *t* tests showed no significant difference in FIT1 scores between the treatment group and the control group at baseline $t(119) = 0.59$, NS, but a significant difference at the midpoint test, $t(116) = 6.04$, $p < 0.001$, $d = 1.01$, a large effect size, and at post-test, $t(115) = 4.60$, $p < 0.001$, $d = 0.69$, a moderate effect size. For the FIT2 measure there was a main effect of group, and of time of testing, with no interaction between these two factors.

Independent sample *t* tests showed no significant difference at baseline $t(119)=1.54$, NS, but a significant difference at the midpoint test, $t(116)=3.33$, $p<0.01$, $d=0.35$, a small effect size, and at post-test, $t(115)=4.17$, $p<0.001$, $d=0.58$, a moderate effect size. For the FIT3 measure there was a main effect of group, a main effect of time of testing, and an interaction between these two factors. Independent sample *t* tests showed no significant difference in FIT3 scores between the treatment group and the control group at baseline $t(119)=0.82$, NS, or at the midpoint test, $t(116)=2.90$, NS, but a significant difference at post-test, $t(115)=4.07$, $p<0.001$, $d=0.51$, a moderate effect size.

Discussion

This study evaluated the impact of engagement with the HER literacy program on the core reading skills of disadvantaged primary school children in NI and who were initially at least a year behind their peers in reading attainment. The researchers sought to make the study more sustainable by involving teachers from the schools as HER coordinators and using the equipment and facilities already in the schools where possible. The results indicate that the children in the treatment group who received the Headsprout intervention made significantly greater gains in the various measures used to test literacy performance than pupils in the control group receiving teaching as usual.

Importantly, average sentence reading age improved by 17.6 months for treatment group children in comparison to 7.1 months for control group children, while average phonics reading age improved by 12.1 months for treatment group children in comparison with 7.8 months for control group children (see Fig. 1). On average children in the HER treatment groups gained an additional 11.6 months on sentence reading age and an additional 6.1 months on phonics reading age (relative to standardized age group norms), while those in the control group gained on average only 1.1 and 1.8 months, respectively, in the 6-month time period. Similar findings were obtained by Storey et al. (2019) evaluating the potential of HER as a supplementary support for at risk pupils in NI. However, the present study increased the sample size of study participants to 123 pupils across eight schools, with randomization to treatment or waiting list control condition (the control schools were offered access to HER training after the study reported here was completed). In addition, this study reports large effect sizes for both the main effect of treatment and the interaction between treatment and time, on both sentence reading age and phonics reading age. These findings add further evidence for a body of literature showing positive impacts on children's reading using HER (Huffstetter et al., 2010; Pindiprolu & Forbush, 2009; Storey et al., 2019; Twyman et al., 2011; Tyler et al., 2015; Watkins et al., 2016) and expand on current literature by employing a larger sample size, randomization and calculating effect sizes and well as standardized scores. The accelerated learning seen for the HER intervention group is vital in the context of pupils who are struggling with literacy and falling behind the performance of their peers (Rose, 2009). These findings are very encouraging, especially in relation to efficiency and efficacy, considering the 6-month timeframe of the study.

The study also incorporated repeated measures of fluency using 1-min timed test using sets of flashcards, with phonemes at three levels of difficulty. The FIT measures showed good validity in that lower correct scores and higher incorrect scores were obtained at baseline as the level of difficulty increased from FIT1 to FIT3. Over the six months of the study, correct scores tended to increase (see Fig. 2) and incorrect scores tended to decrease for all participants (Fig. 3). However, children in the HER treatment group showed greater improvements in phoneme fluency across FIT1, FIT2, and FIT3 assessments with higher rates of correct scores and lower rates of incorrect scores. This pattern supports the results of the standardized reading test measures for both groups, and provides some validation of the measures developed for this study to test accuracy and speed of the key elements of reading fluency (Hudson et al., 2005; Rasinski, 2012; Shanahan, 2012). In addition, similar to findings reported by Nally et al. (2021), the significant increases in rate of word reading and non-word reading by participants in the treatment group suggest that HER is effective in promoting fluency, important for the generalization of decoding strategies (Kent & Street, 2013).

Based on the literacy benefit demonstrated in this study, the potential return on investment of use of HER is significant. Whereas existing research into the evidence base of reading programs and interventions used in the UK found a wide variety of gains, effects sizes and claims of efficacy (Brooks, 2016), the results from this study support the findings of previous studies which resulted in improved literacy performance from children in mainstream schools using Headsprout in both the US, UK and in NI (Twyman et al., 2011; Tyler et al., 2015; Watkins et al., 2016; Storey et al., 2017; Storey et al., 2021). Results of this research and others, along with performance figures from the Department of Education for NI (Department of Education, 2015, 2019), demonstrate that teaching as usual incorporating an eclectic approach is not always effective for this population. The implications of ineffective strategies for children struggling with literacy in terms of education and future wellbeing are profound. The attainment gap appears early and continues to increase throughout school (Department of Education, 2015, 2019) and, as lessons become more difficult, it is likely the attainment gap will increase (Goodman & Gregg, 2010). Subsequent studies (Nally et al., 2021; Thompson et al., 2021; Pindiprolu & Forbush, 2009) have shown the use of evidence-based practice (EBP), such as HER, offers a possible solution to the attainment gap, indicating that well-targeted resources can have important effects for modest outlays. Although research suggest the impact of EBP is limited to between 10 and 30% of the variance in pupil outcomes (Teddle & Reynolds, 2000; Muijs, 2009), an increase of up to 30% in performance would make a significant impact in closing the literacy attainment gap.

The small amount of qualitative information gathered from participating teachers showed that schools can utilize evidence-based CAI approaches easily and effectively, when initial training and ongoing support are offered. While there were some minor issues reported, these were mainly concerning the availability of school resources and adequate Wi-Fi signal, rather than issues with HER software. In fact, respondents reported that they would continue to use Headsprout in the future having seen the positive impact on pupil progress. These findings, although from a small sample of teachers, are important to informing future research and development of

the training and support required by schools and services to better equip them in decision making and use of funding available. Further well-designed research is warranted into the impact of Headsprout Early Reading on literacy skills (Rigney et al., 2020). Equally important is a focus on what educators require in order to identify and select EBP methods, and effectively and efficiently utilize them to close the attainment gap.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10864-022-09489-y>.

Declarations

Conflict of interest There are no potential financial or non-financial conflicts of interest within this research. The authors have no affiliation with any school involved in this research or Headsprout Early Reading, the intervention used with pupils.

Human Participants and/or Animals Consent forms were sought for all participants: (1) School principals signed a consent form for their school to be involved. (2) Parents of all participants signed a consent form allowing their child to participate in the research. (3) Participants who were randomly allocated to the control group were given access to the intervention in the school year following the research free of charge.

Informed Consent Child were given an easy read outline of the research, told what was involved and given the chance to decline to be involved at the outset, or at any point of the research.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barnard, H. (2018). *Poverty in Northern Ireland 2018*. Retrieved August 25, 2019, from <https://www.jrf.org.uk/report/poverty-northern-ireland-2018>. <https://doi.org/10.1016/j.enpol.2017.12.018>
- Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research*, 72(1), 101–130. <https://doi.org/10.3102/00346543072001101>
- Brooks, G. (2016). *What works for children with literacy difficulties: The effectiveness of intervention schemes* (5th edn). Frensham: Dyslexia—SpLD Trust.
- 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. <https://doi.org/10.1080/02796015.2005.12086286>
- Department of Education. (2015). *Statistical Bulletin 5. Qualifications and destinations of Northern Ireland school leavers 2013/14*. Retrieved March 17, 2020, from <https://www.education-ni.gov.uk/sites/default/files/publications/de/qualifications-and-destinations-1314.pdf>
- Department of Education. (2019). *Statistical Bulletin 4. Qualifications and destinations of Northern Ireland school leavers 2017/18*. Retrieved from March 17, 2020, from <https://www.education-ni.gov>.

- uk/sites/default/files/publications/education/qualifications-and-destinations-of-northern-ireland-school-leavers-20171.pdf
- Goodman, A., & Gregg, P. (2010). *Poorer children's educational attainment: How important are attitudes and behaviour?* Joseph Rowntree Foundation. Retrieved from <https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/poorer-children-education-full.pdf>
- Griggs, M. J., & Walker, R. (2008). *The costs of child poverty for individuals and society: A literature review*. Retrieved February 14, 2019, from <https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/2301-child-poverty-costs.pdf>
- Grindle, C. F., Carl Hughes, J., Saville, M., Huxley, K., & Hastings, R. P. (2013). Teaching early reading skills to children with autism using MimioSprout Early Reading. *Behavioral Interventions*, 28(3), 203–224. <https://doi.org/10.1002/bin.1364>
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, 19, 141–164. <https://doi.org/10.3102/01623737019002141>
- Hazel, W. (2018). *Illiteracy costs UK £36 billion a year*. Times Educational Supplement. Retrieved from December 1, 2019, from <https://www.tes.com/news/illiteracy-costs-uk-ps36-billion-year>
- Hirsch, D. (2007). *Experiences of poverty and educational disadvantage*. Joseph Rowntree Foundation. Retrieved from September 14, 2019, from <https://www.jstor.org/stable/1164207>
- Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher*, 58(8), 702–714. <https://doi.org/10.1598/RT.58.8.1>
- Huffstetter, M., King, J. R., Onwuegbuzie, A. J., Schneider, J. J., & Powell-Smith, K. A. (2010). Effects of a computer-based early reading program on the early reading and oral language skills of at-risk preschool children. *Journal of Education for Students Placed at Risk*, 15(4), 279–298. <https://doi.org/10.1080/10824669.2010.532415>
- Kellett, M., & Dar, A. (2007). *Children researching links between poverty and literacy*. Joseph Rowntree. Retrieved March 26, 2020, from <https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/2094-literacy-poverty-children.pdf>
- Kent, J. R., & Street, E. M. (2013). *Response to intervention and precision teaching. Creating synergy in the classroom*. The Guildford Press.
- Layng, T. J., Twyman, J. S., & Stikeleather, G. (2003). Headsprout early reading: Reliably teaching children to read. *Behavioral Technology Today*, 3, 7–20.
- Layng, T. J., Twyman, J. S., & Stikeleather, G. (2004). Selected for success: How Headsprout reading basics teaches beginning reading. *Evidence-based educational methods* (pp. 171–197). Academic Press. <https://doi.org/10.1016/B978-012506041-7/50012-7>
- Layng, T. V., Sota, M., & Leon, M. (2011). Thinking through text comprehension: Foundation and guiding relations. *The Behavior Analyst Today*, 12(1), 3. <https://doi.org/10.1037/h0100706>
- Levačić, R., & Vignoles, A. (2002). Researching the links between school resources and student outcomes in the UK: A review of issues and evidence. *Education Economics*, 10(3), 313–331. <https://doi.org/10.1080/09645290210127534>
- Muijs, D. (2009). Effectiveness and disadvantage in education. In C. Raffo, A. Dyson, H. Gunter, D. Hall, L. Jones, & A. Kalamouka (Eds.), *Education and poverty in affluent countries*. Routledge.
- Nally, A., Holloway, J., Lydon, H., et al. (2021). A randomized controlled trial of Headsprout on the reading outcomes in children with autism using parents as facilitators. *Behavior Analysis in Practice*, 14, 1–14. <https://doi.org/10.1007/s40617-021-00597-1>
- National Institutes of Health. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Retrieved February 8, 2018, from <https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf>
- Pindiprolu, S. S., & Forbush, D. (2009). Computer-based reading programs: A preliminary investigation of two parent implemented programs with students at-risk for reading failure. *Journal of the International Association of Special Education*, 10(1), 66.
- Rasinski, T. V. (2012). Why reading fluency should be hot! *The Reading Teacher*, 65(8), 516–522. <https://doi.org/10.1002/TRTR.01077>
- Rigney, A. M., Hixson, M. D., & Drevon, D. D. (2020). Headsprout: A systematic review of the evidence. *Journal of Behavioral Education*, 29(1), 153–167. <https://doi.org/10.1007/s10864-019-09345-6>
- Rose, J. (2009). *Identifying and teaching children and young people with dyslexia and literacy difficulties: An independent report*. Retrieved June 15, 2019, from <http://www.thedyslexia-spldtrust.org.uk/media/downloads/inline/the-rose-report.1294933674.pdf>

- Shanahan, T. (2012). Developing fluency in the context of effective literacy instruction. In T. Rasinski, C. Blachowicz, & K. Lems (Eds.), *Fluency instruction: Research-based best practices* (pp. 17–34). Guilford Press.
- Storey, C., McDowell, C., & Leslie, J. C. (2017). Evaluating the efficacy of the Headsprout® reading program with children who have spent time in care. *Behavioral Interventions*, 32(3), 285–293. <https://doi.org/10.1002/bin.1476>
- Storey, C., McDowell, C., & Leslie, J. C. (2020). Headsprout early reading for specific literacy difficulty: A comparison study. *Journal of Behavioral Education*, 29(3), 619–633. <https://doi.org/10.1007/s10864-019-09336-7>
- Teddlie, D. R., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London: Falmer Press.
- Thompson, I. (2020). Poverty and education in England: A school system in crisis. In *Poverty in education across the UK: A comparative analysis of policy and place*. Bristol University Press.
- Twyman, J. S., Layng, T. J., & Layng, Z. R. (2011). The likelihood of instructionally beneficial, trivial, or negative results for kindergarten and first grade learners who complete at least half of Headsprout Early Reading. *Behavioral Technology Today*, 6, 1–19.
- Tyler, E. J., Hughes, J. C., Beverley, M., & Hastings, R. P. (2015). Improving early reading skills for beginning readers using an online programme as supplementary instruction. *European Journal of Psychology of Education*, 30(3), 281–294. <https://doi.org/10.1007/s10212-014-0240-7>
- Watkins, R. C., Hulson-Jones, A., Tyler, E., Beverley, M., Carl Hughes, J., & Hastings, R. P. (2016). Evaluation of an online reading programme to improve pupils' reading skills in primary schools: Outcomes from two implementation studies. Cylchgrawn Addysg Cymru/Wales. *Journal of Education*, 18(2), 81–104. <https://doi.org/10.16922/wje.18.2.7>
- 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 Developmental and Physical Disabilities*, 23(6), 491–499. <https://doi.org/10.1007/s10882-011-9240-6>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.