A randomized control trial of phonics training and sight word training in children with dyslexia

Research team: G McArthur, A Castles, S Kohnen, L Larsen, K Jones, T Anandakumar, E Barriball
(ARC Centre of Excellence in Cognition and its Disorders, Macquarie University)

Industry partners: S Davis (www.intrepica.com.au)
J Dyson (www.lexia.com.au)

Funding: NHMRC Project 488518
ARC Discovery Project 0879556

Disclaimer: This research has not yet been submitted for peer review. Please do not cite without written permission from G McArthur
1. Background

Around 5% of children find it unusually difficult to learn to read for no known reason: They have normal intelligence, no known neurological or psychological problems, and have had normal reading instruction. This condition is often called developmental dyslexia [1].

We have known for a long time that dyslexia is associated with long-term academic failure [2]. However, we are only just starting to understand how dyslexia affects emotional health. Children with poor reading are more likely to suffer from anxiety, depression, conduct disorder and hyperactivity than children with normal reading [3]. Given the large number of struggling readers, and the severe educational and emotional consequences of poor reading, it is important to treat poor readers early and effectively.

In the past, most treatment studies done with children with dyslexia have focused on “phonics”, which teaches children to use letter-sound rules to learn new words. For example, when a child first sees the word CAT, they have to (1) identify the letters, (2) translate each letter into a speech sound (“k” “a” “t”), and (3) blend the speech sounds into a spoken word (“kat”). Boxes 1, 2 and 6 in Figure 1 represent these three “components” of phonics reading, which collectively form the “phonics reading route” (also called the nonlexical reading route [4]).

![Figure 1](image.png)  
*Figure 1. The components of reading (boxes 1 to 6) adapted from the Dual-Route Model (DRM) of reading [4]. Most theoretical models of reading [5, 6, 7] incorporate some or all of these components. Thus, the outcomes of this project are applicable to most theoretical models of reading – not just the DRM.*

A systematic review of phonics training in children with reading difficulties has revealed that phonics only has small-to-moderate effects on reading difficulties [8]. A likely explanation for this limited effect is that many children with reading difficulties do not only have problems with phonics reading. They also have problems with “sight word reading”, which is the ability to recognise a whole written word by sight [9, 10]. For example, to read the word CAT by sight, a child needs to identify the letters (component 1 in Figure 1) and then scan their mental store of written word representations for CAT (component 3). If there is a match, this will trigger the meaning of the word (component 4) and the spoken representation of that word (component 5). This in turn will initiate the spoken output of the word “kat” (component 6). Together, these components form the “sight word reading route” (or the lexical reading route [4]).

Sight word reading is particularly important for reading English because one-third of written English words do not follow the letter-sound rules, and so have to be learned by sight (e.g., yacht; [4]). Given the importance of sight word reading in English, and given that most children with dyslexia have problems with sight word reading as well as phonics, it is surprising that no randomized control trial (RCT) of sight word and phonics training has ever been conducted with children with dyslexia. The aim of this study was the carry out the first RCT of sight word and phonics training in children with dyslexia to answer the following questions relating to the treatment of dyslexia:
1. Does 16 weeks of sight word training and phonics training have an effect on the sight word reading route or the phonics reading route of poor readers?
2. Does the order of sight word and phonics training matter?
3. Does 8 weeks of sight word training, phonics training, or mixed training have an effect on the sight word or phonics reading routes of poor readers?
4. Does the type of training (sight word, phonics, mixed) matter?

2. Methods

This study was a double-baseline, parallel-group, double-blind RCT. Below we report the methods according to the CONSORT 2010 guidelines [11].

2.1 Trial design

The order of testing sessions and training phases involved in this RCT is illustrated in Figure 3.

<table>
<thead>
<tr>
<th></th>
<th>Phonics+sight word (N = 36)</th>
<th>Sight word+phonics (N=36)</th>
<th>Mixed+mixed (N = 32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>Screening measures</td>
<td>Screening measures</td>
<td>Screening measures</td>
</tr>
<tr>
<td></td>
<td>Outcome measures</td>
<td>Outcome measures</td>
<td>Outcome measures</td>
</tr>
<tr>
<td>Test-retest 8 weeks</td>
<td>No training</td>
<td>No training</td>
<td>No training</td>
</tr>
<tr>
<td>Test 2</td>
<td>Outcome measures</td>
<td>Outcome measures</td>
<td>Outcome measures</td>
</tr>
<tr>
<td></td>
<td>2-3 hours</td>
<td>2-3 hours</td>
<td></td>
</tr>
<tr>
<td>Train 1 8 weeks</td>
<td>Phonics</td>
<td>Sight words</td>
<td>Mixed</td>
</tr>
<tr>
<td>Test 3</td>
<td>Outcome measures</td>
<td>Outcome measures</td>
<td>Outcome measures</td>
</tr>
<tr>
<td></td>
<td>2-3 hours</td>
<td>2-3 hours</td>
<td></td>
</tr>
<tr>
<td>Train 2 8 weeks</td>
<td>Sight words</td>
<td>Phonics</td>
<td>Mixed</td>
</tr>
<tr>
<td>Test 4</td>
<td>Outcome measures</td>
<td>Outcome measures</td>
<td>Outcome measures</td>
</tr>
<tr>
<td></td>
<td>2-3 hours</td>
<td>2-3 hours</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Testing and training phases for the three training groups. See Figure 3 for participant numbers. See Table 1 for screening and outcome measures.

At Test 1, children did the screening and outcome measures (see Table 1). After 8 weeks of no training, they returned to do the outcome measures only (Test 2). This no-training period indexed the test-retest effects for each outcome measure. Then Group 1 did 8 weeks of phonics training (followed by Test 3) and then 8 weeks of sight word training (followed by Test 4). Group 2 did the same except that the training was done in the reverse order. Group 3 did phonics and sight word training on alternate days for 8 weeks (followed by Test 3) and then the same for a further 8 weeks (followed by Test 4).

2.2 Participants

A flow diagram of the number of participants in each stage of the RCT is shown in Figure 3.
Figure 3. A flow diagram of the participant numbers at each stage of the RCT.

Children were recruited from schools, clinics, and via newspaper advertisements. They were included in the study if they (1) were aged from 7 to 12; (2) scored at least 1 SD below the age-mean on the Castles and Coltheart 2 (CC2) irregular-word reading test or nonword reading test [12]; (3) had no history of neurological or sensory impairment as indicated on the background questionnaire; and (4) used English as their primary language at school and at home (see screening measures in Table 1). It is noteworthy that while all children were tested for their non-verbal intelligence, children with non-verbal IQ scores below the average range were not excluded from the study since there is evidence that intelligence does not predict reading ability or response to treatment [13].

### 2.3 Randomisation

We allocated children into the three training groups with cluster randomisation using time to define the clusters [14]. Specifically, we allocated children recruited in months 1 to 6 to Group 1; months 7 to 12 to Group 2; and months 13 to 18 to Group 3. We used this type of cluster randomization for several interacting reasons. Specifically, we knew that we had to recruit children for the study from month 1 to month 18 because (1) all testing and training had to be completed in 2 years, and (2) the last child had to be recruited in month 18 since it would take them six months to do all the testing and training. However, we did not know how quickly we could recruit children who fulfilled our criteria, and we could not anticipate that the development of the sight word training would be unexpectedly delayed, so that only the phonics training was available in month 1 of the study. The solution to these unknown factors was to allocate children to the three groups by dividing the 18 month recruitment period into three periods of time. We then allocated children recruited in months 1 to 6 to the phonics+sight word training group, since the 8-week phonics training period would allow time for the sight word training to be finished. Children recruited in months 7 to 12 were allocated to the sight word+phonics group. Finally, children recruited in months 13 to 18 were allocated to the mixed+mixed group. This time-based cluster randomization appeared to work well since pre-training comparisons on screening and outcome measures showed that the groups were well matched prior to training (see Table 1).
2.4 Double-blinding

Unlike drug trials, it is difficult to guarantee double blinding in cognitive RCTs because the type of training cannot be completely concealed from a volunteer. However, neither parents nor children were told if they were in the experimental or control group, and it is unlikely that they had the expertise to ascertain the type of training (i.e., they were blind to group allocation). Regarding testers, it is unethical and impractical to hire and fire trained reading experts in a long-term continuous RCT such as this to ensure that different testers are used before and after training. Thus, we employed four casual testers (PhD students) to help the two principal testers. With careful planning, we ensured that no tester assessed the same child twice, and no tester was aware of the child’s group allocation (i.e., the tester was blind to group allocation).

2.5 Measures

2.5.1 Screening measures

The children were given three screening tests. We tested nonword reading and irregular-word reading accuracy with the CC2 reading test [12]. The test comprises 40 nonwords that can be read accurately using the GPC rules (e.g., grenty), 40 irregular words (e.g., yacht), and 40 regular words (e.g., marsh). Items were presented in increasing difficulty on separate cards in a fixed random order. Children were asked to read the item on each card until they made five errors in a row for any one type of item (e.g., irregular words), at which point the presentation of that type of item was stopped. Presentation of the other types of items continued until the child made five errors for each of the other types of items or they reached the end of the test. A child was given 10 seconds to read each word before being prompted to try the next word. Scores were expressed as z scores that had a mean of 0 and SD of 1.

We measured non-verbal IQ with the Kaufman Brief Intelligence Test 2 (KBIT-2) Matrices subtest [15]. In each trial, children saw an incomplete picture matrix and had to select the missing portion from six possible options. Scores were standardised with a mean of 100 and an SD of 15.

We used a questionnaire to determine if children had any known problems with their hearing, vision, neurology, or psychology, which could account for their reading problem. We also used this questionnaire to establish if the children used English as their primary language at both school and home.

2.5.2 Outcome measures

2.5.2.1 Phonics reading route

We tested the effect of the training programs on three measures of the phonics reading route. All three tests used nonwords as stimuli since nonwords can only be read correctly by the phonics reading route (i.e., regular words can be correctly read by the phonics route and the sight word reading route).

We tested nonword reading accuracy using 20 untrained nonwords printed on flashcards. A child was asked to read each nonword aloud. All items were monosyllabic, comprised 3 or 4 letters (e.g., urk, ump, mouk, clon), and translated to two, three or four phonemes. Half the items contained digraphs (e.g., th, ai, oo), and half single-letter correspondences (e.g., t, p, e). Scores were the total correct out of 20.

We indexed nonword reading fluency using the Test of Word Reading Efficiency (TOWRE) nonword subtest [16]. This comprised 63 increasingly difficult nonwords that can be read correctly using the GPC rules. The child was asked to read as many nonwords as they could in 45 seconds. Scores were standard scores with a mean of 100 and SD of 15.

We tested the application of the phonics reading route to the task of nonword spelling using the Diagnostic Spelling Test for Nonwords (DiSTn; ref), which tests a person’s knowledge of English sound-letter-rules. The
DiSTn is composed of 74 items (i.e., nonwords) in which children spell a nonword that is spoken by the tester. Scores were the total correct out of 74.

2.5.2.2 Sight word reading route

We measured the accuracy of the sight word route with three measures. Each used irregular words as stimuli because irregular words can only be read correctly by the sight word reading route (i.e., regular words can be correctly read by the phonics route as well as the sight word reading route).

We measured irregular word accuracy by presenting 60 irregular words on flash cards. Thirty of the words were in the sight word training program and 30 were not. Children were asked to read each word on each flash card. Scores were the total correct trained irregular words (out of 30) and the total correct untrained irregular words (out of 30).

We measured the ability to apply the sight word reading route to the task of irregular-word spelling using the Diagnostic Spelling Test for Irregular Words (DiSTi). In each of 74 items, the tester named an irregular word and the put that word in a sentence. The child was asked to spell each word. Scores were the total correct out of 74.

2.5.2.3 Both reading routes

We tested the output of both reading routes using two general measures of reading. The first used the TOWRE sight word subtest [16] to measure word reading fluency. This comprises 104 irregular (sight word route) and regular words (sight word and irregular word route) that increased in difficulty. A child was asked to read as many words as possible in 45 seconds. Scores were standard scores with a mean of 100 and SD of 15.

The second test measured reading comprehension in the context of everyday reading tasks. The Test of Everyday Reading Comprehension (TERC; [17]) presented children with 11 examples of everyday reading stimuli, such as a text message, the label on a medicine bottle, a party invitation, or a canteen menu. They were asked two literal (i.e., not inferential) questions based on the information written in each example. Scores were the total answers correct out of 22.

2.6 Training programs

2.6.1 Phonics training program

Children were asked to do the phonics program at home for 30 minutes per day, 5 days per week, for 8 weeks. The phonics training was a modified version of the Lexia® Strategies for Older Students. The modifications - provided by the Managing Director of Lexia® Learning Australia, John Dyson - involved removing all exercises that included irregular words. This included exercises that presented sentences and paragraphs of text. Thus, the program focused on training grapheme-phoneme correspondences either alone, within subsections of words (i.e., syllables), and within regular words.

2.6.2 Sight word training

The sight word training was also done at home for 30 minutes per day, 5 days per week, for 8 weeks. Each training session trained one of 30 lists of 24 irregular words that increased in difficulty between and within each list. Irregular words were used to train the sight word reading route because regular words train the phonics reading route as well as the sight word reading route. In the first sight word training session, a parent tested their child’s ability to read the easiest list of 24 irregular words by presenting each word on a flash card. All the words that the child read correctly were placed in a “correct pile”. All the words that they read incorrectly were placed in an “incorrect pile”. Next, the child spent 20 minutes playing a computerised memory game called Dingo Bingo by MacroWorks®, which had been adjusted by
Shane Davis (the CEO) to present the same list of irregular words in the session as the flash cards. The parent noted any words that the child failed to read in Dingo Bingo, and added them to the incorrect pile of flash cards (Part 2 of the session). Finally, the parent re-presented the child with the flashcards in the incorrect pile. If a child got less than 2 incorrect (i.e., 0 or 1 mistake), then they moved onto the next list the next day (List 2). Otherwise they redid the same list in the next training session.

2.6.3 Mixed training

The mixed training was the same as the phonics and sight word training except that each type of training was done on alternate days. So, on the first day the child did phonics training, the second day sight word training, the third day phonics training, and so on.

3. Results

3.1 Equality of groups prior to training

Before analysing the effects of training on the outcome measures, it was important to establish that the training groups did not differ on the screening or outcome measures prior to training. This ensured that any improvements in outcome measures after training were not due to regression to the mean effects (i.e., according to statistical probability, any person or group with an extreme score at one point in time is likely to have a more moderate score at a second point in time even if there is no actual change in their ability). We used a between groups ANOVA to compare the screening and outcome measure scores of the three training groups prior to training (i.e., at Test 1). The relevant data is shown in Table 1.

Table 1. Means and standard deviations (M/SD) for the screening and outcomes measure for the three training groups at Test 1 prior to training.

<table>
<thead>
<tr>
<th></th>
<th>Phonics+ sight word (ph-sw)</th>
<th>Sight word+ phonics (sw-ph)</th>
<th>Mixed+mixed (m-m)</th>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>9.42/1.71</td>
<td>9.19/1.64</td>
<td>9.21/1.67</td>
<td>ns</td>
</tr>
<tr>
<td>Non-verbal IQ (ss)</td>
<td>97.5/14.16</td>
<td>95.56/17.12</td>
<td>101.12/14.25</td>
<td>ns</td>
</tr>
<tr>
<td>Irregular words (z)</td>
<td>-1.31/0.67</td>
<td>-1.38/0.64</td>
<td>-1.26/0.58</td>
<td>ns</td>
</tr>
<tr>
<td>Nonwords (z)</td>
<td>-1.50/0.57</td>
<td>-1.27/0.54</td>
<td>-1.32/0.53</td>
<td>ns</td>
</tr>
<tr>
<td>Regular words (z)</td>
<td>-1.41/0.57</td>
<td>-1.29/0.59</td>
<td>-1.30/0.59</td>
<td>ns</td>
</tr>
<tr>
<td>Nonword accuracy (raw)</td>
<td>7.06/6.35</td>
<td>9.08/6.49</td>
<td>8.31/7.29</td>
<td>ns</td>
</tr>
<tr>
<td>Nonword fluency (ss)</td>
<td>11.67/9.24</td>
<td>13.08/6.95</td>
<td>14.66/11.19</td>
<td>ns</td>
</tr>
<tr>
<td>Trained irregular accuracy (raw)</td>
<td>11.17/7.77</td>
<td>10.97/8.09</td>
<td>11.16/8.07</td>
<td>ns</td>
</tr>
<tr>
<td>Untrained irregular accuracy (raw)</td>
<td>8.58/6.39</td>
<td>8.5/6.37</td>
<td>9.03/6.28</td>
<td>ns</td>
</tr>
<tr>
<td>Irregular word spelling (raw)</td>
<td>14.57/10.52</td>
<td>13.28/11.51</td>
<td>15.41/15.04</td>
<td>ns</td>
</tr>
<tr>
<td>Word reading fluency (ss)</td>
<td>42.47/16.88</td>
<td>40.83/16.84</td>
<td>41.78/16.74</td>
<td>ns</td>
</tr>
<tr>
<td>Reading comprehension (raw)</td>
<td>14.25/8.37</td>
<td>13.75/8.2</td>
<td>14.59/7.89</td>
<td>ns</td>
</tr>
</tbody>
</table>
There was no significant difference between the three groups prior to training on the screening measures or the outcomes measures.

### 3.2 Outcome measures

Figure 4 shows the gains in raw scores on each outcome measure after 8 weeks and 16 weeks of training for each training group. Specifically, each line in each graph – with its corresponding circle (the mean) and arms (95% confidence intervals = variability within the group) - represents the shift in a child’s raw score on a test from one test time (e.g., Test 1 before training) to another (e.g., Test 4 after 16 weeks of training). A positive difference score means that a child’s score was better at the second test time than the first. The first three lines in each graph belong to Group 1 who did phonics and then sight word training. The second three lines relate to Group 2 who did sight word training then phonics training. The last three lines relate to Group 3 who did two phases of mixed training.

The first line for each group, labelled T1T2, shows the shift in raw scores from Test 1 to Test 2. These two tests flanked an 8-week double-baseline period of no training, and so any positive shift reflects an improvement due to test-retest effects. Any T1T2 line that does not fall across the 0 line (i.e., the point of no difference between raw scores at each test time) is statistically significant (p <= .05). These are marked with an *.

The second line for each group, labelled T1T3, reflects the difference in raw scores on the test between Test 1 (before training) and Test 3 (after the first 8 weeks of training). If this line does not cross the 0 line (i.e., no difference in raw scores between Test 1 and 3) then this means there was a statistically significant increase in raw scores between tests (marked with an *). For this significant increase to be a “true” training effect, it had to be significantly larger than the test-retest effect (i.e., T1T2; also marked with an *). Any T1T3 line that represents a significant and true training effect is marked **.

The third line for each group, labelled T1T4, is the difference in raw scores between Test 1 (before training) and Test 4 (after 16 weeks of training). If this line does not fall over the zero line (marked with an *), and if the raw difference score is significantly larger than the test-retest effect (T1T2; marked with an *) then there is a true effect of 16 weeks of training (i.e., marked with **). For T1T2, T1T4, and T1T3, we have provided the within-subject equivalent of Cohen’s d effect sizes for each effect underneath the * marks ([www.cognitiveflexibility.org; [18]).

As well as testing if each type of training had a true effect, we wanted to determine if there was any difference between the size of the effects for the different types of training done over 8 weeks (i.e., phonics versus sight words versus mixed) and done over 16 weeks (i.e., phonics+sight words versus sight words+phonics versus mixed+mixed). For each outcome, we used a between-group ANCOVA (controlling for each group’s test-retest effects on the outcome measure) to compare (1) T1T3 effects for Group 1 (phonics) and Group 2 (sight words) and Group 3 (mixed); and (2) T1T4 effects for Group 1 (phonics+sight words) and Group 2 (sight words+phonics) and Group 3 (mixed+mixed). We explored any significant group difference using post-hoc t-tests to determine which groups differed from which. These groups are marked with a horizontal line in Figure 4.

We interpret the outcomes of these analyses for each outcomes measure in separate sections below.
Figure 4. Increases in raw scores between different test times for outcome measures for each group.
3.2.1 Nonword reading accuracy

Sixteen weeks of phonics and sight word training had a significant and moderate-to-large true training effect on nonword reading accuracy, regardless of order of training. That is, there was no significant difference between the outcomes of the group who did phonics then sight word training, the group who did sight word then phonics training, and the group who did mixed training.

Eight weeks of phonics, sight word, or mixed training also had moderate-to-large true training effects on nonword reading accuracy. However, there was no difference between the outcomes for the groups who did phonics training, sight word training, or mixed training.

3.2.2 Nonword reading fluency

Sixteen weeks of phonics and sight word training had a significant and moderate-to-large true training effect on nonword reading fluency in the phonics+sight word group and mixed+mixed group. The sight word+phonics group showed very similar effects, but these were not significantly larger than their test-retest effect. There was no significant difference between the groups who did the two types of training in different orders.

Similarly, 8 weeks of phonics, sight word, or mixed training had moderate-to-large true training effects on nonword reading fluency in the phonics+sight word group and mixed+mixed group; and the effect in the sight word+phonics group was not significantly larger than the test-retest effect. Again, there was no significant difference between the groups who did different types of training.

3.2.3 Nonword spelling

Sixteen weeks of phonics and sight word training did not have a true training effect on nonword spelling in any group. Similarly, 8 weeks of phonics, sight word, or mixed training did not have a true training effect on nonword spelling.

3.2.4 Trained irregular word accuracy

Sixteen weeks of phonics and sight word training had a very large and significant true training effect on trained irregular word accuracy. There was no significant difference between the groups who did the training in different orders.

Eight weeks of phonics or sight word or mixed trained also had very large and significant true training effects on trained irregular word accuracy. In this case, there was an effect of training type because the groups that received irregular word training (i.e., the sight word+phonics group and mixed+mixed group) made significantly greater gains than the group that did phonics training (i.e., phonics+sight word group).

3.2.5 Untrained irregular word accuracy

Sixteen weeks of phonics and sight word training had a significant and very large true training effect on untrained irregular word accuracy. There was a significant difference between groups who did the training in different orders: The groups who did sight word training in the last 8 weeks of their training (phonics+sight word group and mixed+mixed group) made significantly greater gains than the group who did phonics in the last 8 weeks (sight word+phonics group) who made no gains during their phonics training period.

Eight weeks of phonics, sight word training, or mixed training also had very large and significant true training effects on trained irregular word accuracy. There was no statistically significant effect of training type on untrained irregular word accuracy.
3.3.3 Irregular-word spelling

Sixteen weeks of phonics and sight word training had a large and significant true training effect on irregular word spelling in all groups. There was no statistically significant difference overall between the three groups that did the training in different orders.

Eight weeks of phonics training, sight word training, or mixed training had a moderate to large, significant, true training effect on irregular word spelling. There was no significant difference between training type.

3.4.1 Word reading fluency

Sixteen weeks of phonics and sight word training had a large and significant true training effect on word reading fluency. There was no significant difference between the groups who did different orders of training.

Eight weeks of phonics training, sight word training, or mixed training also had a large and significant effect on irregular word spelling. There was no difference between the groups that did different types of training.

3.4.2 Reading comprehension

Sixteen weeks of phonics and sight word training had a significant and large-to-very large true training effect on reading comprehension regardless of the order of training. Eight weeks of phonics training, sight word training, or mixed training also had large to very large significant effects on reading comprehension, regardless of the type of training.

4. Discussion

The aim of this study was the carry out the first RCT to measure the effects of both sight word and phonics training in children with dyslexia. The outcomes were used to address four questions relating to the treatment of dyslexia. These are discussed in turn below.

4.1 Does 16 weeks of sight word and phonics training have a significant effect on the sight word reading route or the phonics reading routes of poor readers?

The answer is yes. Sixteen weeks of sight word and phonics training had significant and true training effects on all of the outcome measures except for (1) nonword spelling for all groups, and (2) nonword reading fluency for the sight word+phonics group. The largest training effects were on the sight word reading route, with very large effects on trained and untrained irregular words, and large effects for irregular word spelling. Large effects were also seen on the measures that taxed both routes (sight word reading fluency and reading comprehension). The training effects on the phonics route, which were significant and true, were moderate to large.

4.2 Does the order of sight word and phonics training matter?

The answer is “it depends on the outcome measure”. The order of phonics and sight word training did affect gains in untrained irregular words. The groups who did their sight word training in the last 8 weeks made the largest gains (phonics+sight word group and mixed+mixed group). In contrast, the gains made by the sight word+phonics group occurred during their sight word training; and they made no gains during their phonics training. While there were trends for order effects for other outcome measures (e.g., trained irregular words, irregular word spelling), these trends were not statistically significant.
4.3 Does 8 weeks of sight word training, phonics training, or mixed training have a significant effect on the sight word or phonics reading routes of poor readers?

The answer is yes. Like 16 weeks of training, the 8-week training programs had significant and true training effects on all outcome measures except for nonword spelling (all groups) and nonword reading fluency (sight word+phonics group). Further, the effects of 8 weeks of training were largest for the sight word reading route (large to very large effects), followed by both routes (large effects), and the phonics reading route (moderate to large effects). Eight weeks of training differed from 16 weeks of training in having a smaller effect on irregular word spelling (moderate-to-large rather than large effect).

4.4 Does the type of training (sight word, phonics, mixed) matter?

Again, the answer is “it depends on the outcome measure”. Children who did 8 weeks of sight word training made significantly greater gains in trained irregular word accuracy than children who did 8 weeks of phonics training. Interestingly, the former included children in the mixed+mixed group who made similar gains on the trained irregular word test to the sight word+phonics group even though they did half as much sight word training. It is also interesting to note that the phonics+sight word training group “caught up” with the other two groups as soon as they did their 8 weeks of sight word training. It is also noteworthy that there were trends for training-type effects for other outcome measures (e.g., nonword reading accuracy, untrained irregular word accuracy), but these trends did not reach statistical significance.

4.6 Limitations

A possible limitation of this study is the size of the treatment groups. It took our team 2 years to identify 141 children with dyslexia who fulfilled the research criteria. Thirty-seven children dropped out of the study. This is perhaps not surprising given the large, yet necessary, amount of training and testing. The remaining 104 children were divided into three groups of 36, 36, and 32 children. While these sample sizes had power enough to allow the moderate, large, and very large within-group effects to reach statistical significance, they may not have been large enough to allow potential training order and type effects to reach statistical significance. For example, it can be seen in Figure 4 that the groups whose first training phase included phonics (phonics+sight word group and mixed+mixed group) made more rapid gains in nonword reading accuracy than the group who did sight word training (sight word+phonics group). However, the latter group made (impressive) gains once they later did phonics training. This trend is not only clear in the data, but makes sense theoretically. However, it was not statistically significant. Thus, we must make our conclusion that type and order of training only affects two outcomes measures (trained and untrained irregular word accuracy) with the caveat that larger-scale studies, with greater statistical power, may reveal other training order and type effects.

4.7 Summary

In sum, the outcomes of this project suggest that 16 weeks of phonics and sight word training has large or very large effects on the general functioning of the phonics and sight word reading routes of children with dyslexia. Further, 8 weeks of phonics, sight word, or mixed training had moderate to very large effects in the children with dyslexia. The effects on the sight word route, and both routes combined, were large to very large. The effects on the phonics reading route, being moderate to large, were slightly smaller. This suggests that the sight word reading route may be easier to train than the phonics reading route. These findings are important because they represent an advance in treatment of dyslexia. Specifically, the mean effect found in the current study was large (Cohen’s d = 1.06), which is a big improvement on the small-to-moderate effects found by previous phonics training studies with children with dyslexia [8]. This superior effect size supports the idea that children with dyslexia need treatment for sight word reading and phonics reading, and not just phonics training alone.
5. References