Teaching children to read irregular words:

A comparison of three instructional methods

Danielle Colenbrander[1], Saskia Kohnen[1], Elisabeth Beyersmann[1], Serje Robidoux[1], Signy Wegener[1], Tara Arrow[1], Kate Nation[2] & Anne Castles[1]

[1]School of Psychological Sciences and Macquarie University Centre for Reading, Macquarie University, Australia

[2]Department of Experimental Psychology, University of Oxford, United Kingdom

Keywords: reading instruction, irregular word reading, exception word reading, mispronunciation correction, set for variability

Corresponding author: Danielle Colenbrander, School of Psychological Sciences and Macquarie University Centre for Reading, Macquarie University, Sydney, New South Wales, Australia, 2109; Email: danielle.colenbrander@mq.edu.au

This is the author accepted version of the manuscript, which has been accepted for publication in Scientific Studies of Reading. The published version of the manuscript is available here: http://dx.doi.org/10.1080/10888438.2022.2077653

Abstract

Purpose

Children learning to read in English must learn to read words with varying degrees of grapheme-phoneme correspondence regularity, but there is very little research comparing methods of instruction for words with less predictable or irregular spellings. Therefore, we compared three methods of instruction for beginning readers.

Method

Eighty-five Kindergarten children were randomly assigned to either Look and Say (LSay), Look and Spell (LSpell), mispronunciation correction (MPC), or wait-list control conditions. Children were taught 12 irregular words over three sessions. Amount of instructional time and number of exposures to the written and spoken forms of the words was controlled across the three experimental conditions. After training, children were assessed on reading aloud and orthographic choice measures.

Results

Children showed evidence of superior learning of trained words in the LSpell and MPC conditions, compared to LSay and control conditions. Differences between the LSpell and MPC conditions were not significant. There was no evidence of generalization to untrained items.

Conclusions

Findings indicate that active processing of a word's orthography is crucial for learning irregular words. These results have implications for initial reading instruction. Further research is required to determine whether differences between LSpell and MPC conditions emerge after longer periods of training.

In early reading instruction, effective use of limited instructional time is crucial. It is important that children reach a level of independence in their reading as quickly as possible, because independent reading is a powerful vehicle for further learning (Ehri, 2005, 2014; Nation, 2017; Share, 1995). Once children have some knowledge of grapheme-phoneme correspondences (GPCs), they can begin to decode simple words, and this process of decoding helps form lexical representations in memory. Repeated exposure to particular words results in increasingly detailed lexical representations of those words, which in turn results in faster recognition of the words in future instances (Ehri, 2005, 2014; Share, 1995). In this way, children build a bank of words that can be recognized fluently and efficiently. Children who are capable of fluent, automatic word recognition are likely to read more, and thus improve their own reading abilities to an even greater degree (Nation, 2017). Initial reading instruction should be designed to optimize children's ability to take advantage of this relationship.

However, words in the English language lie on a continuum of spelling-sound regularity (Colenbrander, Wang, Arrow & Castles, 2020). On one end of this continuum are words that can be accurately read using predictable, one-to-one correspondences between graphemes and phonemes (e.g., *vet*, *cat*, *catch*). On the other end are words containing numerous unpredictable grapheme-phoneme correspondences (e.g., *aisle*, *meringue*). Many of the most frequent words in the English language fall towards the irregular end of the continuum (e.g., *the*, *would, through*). Therefore, the ability to read "irregular" words is important from early on in reading instruction – children cannot read connected text (however simple) if they cannot read the most frequent words in their language. While there is a large body of research demonstrating the effectiveness of systematic, explicit instruction of GPCs (e.g., see National Reading Panel, 2000; Galuschka et al., 2014; McArthur et al., 2018), there is far less research addressing the instruction of irregular words. This is at least partially because there is disagreement in both the research and practical domains about how to define what is meant by “irregular words”. Some researchers and practitioners argue that very few words in English are irregular if children are taught sufficient information about the writing system (e.g., Joshi et al., 2008; Moats, 2005), but in-depth instruction into the nature of the writing system takes time, and there is debate as to how much is “sufficient” (e.g., see Bowers & Bowers, 2017; Moats, 2005; Solity & Vousden, 2009). The definition of regularity is also influenced by the knowledge of the reader (e.g., Joshi et al., 2008; Miles et al., 2018; Steacy et al., 2016). A greater number of words will appear “irregular” to a child who knows very few GPCs, than they will to a child with good knowledge of the most frequent GPCs in English. For the purposes of this study, we adopt the definition used by Colenbrander et al. (2020), whereby a word is considered regular if it can be correctly read using the most frequent pronunciation for each grapheme. All other words are considered irregular. This is a relatively strict functional definition, but it is likely to be appropriate for beginning readers, who generally have limited knowledge of GPCs.

Although the research on irregular word reading in initial reading instruction is sparse, two broad approaches show promise. The first approach is usually known as “sight word instruction”. Such instruction involves repeated exposure to pairings of the spoken and written forms of whole words, which are memorized. Sight word instruction has been controversial, with some researchers and practitioners arguing that it may be counterproductive, resulting in a guessing strategy or detracting attention away from the GPCs in the word (e.g., Miles et al., 2018; Potter, 2012). This is likely to be true if sight word instruction is used as the sole or primary method of initial reading instruction, but when a limited set of frequent irregular words is taught in conjunction with systematic instruction in frequent GPCs, there is no evidence to support the view that sight word instruction is detrimental. In fact, there is evidence to suggest that sight word instruction can be effective even with struggling readers (e.g., Broom & Doctor, 1995; Kohnen, Nickels, Brunsdon, et al., 2008; McArthur, Castles, et al., 2015; McArthur, Kohnen, et al., 2015; Shapiro & Solity, 2016). There is also a small amount of evidence that sight word instruction can result in generalization to words with similar spellings to trained words, words with high numbers of orthographic neighbours, and words that are frequent in written and spoken language (e.g., Brunsdon et al., 2005; Kohnen, Nickels, Brunsdon, et al., 2008; Kohnen, Nickels, Coltheart, et al., 2008; Terzopoulos et al., 2018).

Many studies of sight word instruction use methods that can be broadly termed "Look, Say, Cover, Write, Check", in which children are asked to look at the orthographic form of an irregular word, say it aloud, cover the word and then spell it from memory. The act of spelling the word may be a crucial aspect of such instruction, because it may help form a stable representation of a word's orthographic form in memory. Correctly spelling a word requires explicitly retrieving the graphemes that represent the phonemes in the word, in the correct order (Perfetti, 1997). Written spelling can be very laborious for young children and those with limited writing abilities, but the additional processing demands of spelling (as opposed to reading) may actually be an advantage when it comes to forming long-term representations in memory (Perfetti, 1997; Conrad, Kennedy, Saoud, Scallion & Hanusiak, 2018). However, no studies to date have explored which elements of sight word instruction are crucial for learning, and which are not. Furthermore, while sight word instructional methods have been compared to business-as-usual control conditions, they have rarely been compared to alternative training methods (e.g., see Colenbrander et al., 2020).

In recent years, research on a second broad approach to irregular word reading instruction, and reading instruction in general, has emerged. This approach is known as "set-for-variability" or "mispronunciation correction" (henceforth, we will use the term "mispronunciation correction"). Readers may spontaneously self-correct their own partial decoding attempts using spoken vocabulary knowledge or information from context. Mispronunciation correction involves binding of graphemes to phonemes, and the linking of a partially decoded form (sometimes called a “spelling pronunciation”) to the correct phonological form of the word in memory (e.g., Elbro, de Jong, Houter & Nielsen, 2012; Kearns et al., 2016; Tunmer & Chapman, 2012). There is evidence that mispronunciation correction skill can vary at both the child and word level. In other words, mispronunciation correction may be both a metalinguistic skill, and a way of building orthographic representations at the item level (Steacy et al., 2019). Thus, teaching children to correct their mispronunciations from the beginning of reading instruction may help them to develop a flexible strategy that they can apply to any unfamiliar word (both regular and irregular), accelerating the process of self-teaching. It may also be an effective way to teach the orthography of specific words.

Indeed, there is some evidence to indicate that a mispronunciation correction approach may be effective. In one study (Dyson et al., 2017), five- to seven-year-old children were taught to correct their own mispronunciations, as well as discuss word meanings, write words and match them visually, and complete rhyming tasks. A control group received business-as-usual instruction that did not involve targeted exposure to the trained words. Relative to the control group, the children in the mispronunciation correction group improved their ability to read trained irregular words and untrained irregular words of similar difficulty, though there was no effect of training on a standardized test of irregular word reading. In another, larger study (Savage et al., 2018), two multi-component training programs for poor readers in Grade 1 were compared. One training program (called Direct Mapping and Set for Variability, or DMSfV) involved mispronunciation correction instruction alongside phonics instruction, with close mapping between taught GPCs and the content of books used for shared reading activities. The other program involved phonics and phonological awareness instruction as well as sight word instruction, and taught GPCs were not closely mapped to the content of shared reading books. Children in the DMSfV condition showed greater reading improvements than children in the comparison condition, though irregular words were not assessed separately.

However, these were multi-component instructional programs, and it is not clear which aspects of these programs were the “active ingredients” leading to the demonstrated reading gains. To date only one study (Zipke, 2016) has explored the effectiveness of teaching mispronunciation correction in isolation. In this study, children in Grades 1 and 2 were assigned to either an experimental group receiving mispronunciation correction training, or a control group who simply practiced reading aloud. All children received five 20-minute lessons. Students in the mispronunciation correction group read aloud irregular words embedded in texts. If they made an error, they were given prompts which encouraged them to say the word another way, think of another word that sounded similar, or try a different sound for a particular letter. Children were post-tested on their ability to read a list of irregular words that they had not seen during training.

After training, children in the mispronunciation correction group made more attempts to correct their own errors than controls, but those attempts were not more likely to result in a correct response. In other words, children had learnt a strategy, but it did not improve their reading. This may be because children in the study had been taught to read via whole-word methods, and did not have the decoding abilities to arrive at the correct response. It is also relevant that the researcher did not measure children’s ability to read the words they saw during instruction. Thus, this study measured the effects of teaching mispronunciation correction as a strategy, but did not measure the effects of mispronunciation correction on learning of particular words.

In summary, there is research to indicate that sight word instruction and mispronunciation correction methods may be effective for teaching irregular words as part of programs of initial reading instruction. However, it is not clear which aspects of sight word instruction and mispronunciation correction are essential for learning to occur, whether one of these methods is more effective than the other, or what consequences these different methods will have for generalization to untrained words.

Several studies demonstrate that the act of spelling can result in better orthographic learning than reading alone (e.g., Conrad et al., 2018; Ouelette, 2010; Shahar Yames and Share, 2008). This suggests that sight-word instruction involving spelling will be more effective than sight word instruction that does not. On the other hand, while methods of sight word instruction have never been directly compared to mispronunciation correction instruction, there are reasons to believe that mispronunciation correction may be more effective. Firstly, as outlined above, the process of mispronunciation correction involves decoding a word and using a partially decoded (regularized) form to retrieve the word's correct phonological form, binding together the graphemes and phonemes in the word, as well as binding the regularized form or "spelling pronunciation" to the existing phonological form. This spelling pronunciation may act as a mnemonic to help a child remember the graphemes in the irregular word (Hilte & Reitsma, 2006), and the whole process may result in a more detailed and stable orthographic representation. Secondly, if children can learn to apply mispronunciation correction procedures to unfamiliar words, this type of instruction may be more likely to result in generalization than sight word instruction.

Therefore, in this study, we aimed to pare back sight word and mispronunciation correction instruction to their essential elements, and compare their effectiveness for irregular word reading instruction in beginning readers. To do so, we developed three training programs that we administered to children in their first year of formal schooling. The three training conditions were compared to a wait-list control condition. The first training condition, Look and Say, involved teaching children the pairings between written and spoken forms of whole irregular words. The second condition, Look and Spell, included all the elements of Look and Say, with the addition of written spelling. The third condition, Mispronunciation Correction, involved simplified mispronunciation correction training. Because we wished to pare back these methods of instruction to their most basic forms, we taught words in isolation rather than in sentence context. Children in all conditions were shown a picture corresponding to each irregular word, and children in the mispronunciation correction condition were taught to correct their mispronunciations using the pictures as cues. We aimed to answer the following questions:

1. Which of the three methods of instruction is most effective for improving children's ability to read and recognize correct spellings of the trained words?
2. Do any of the methods of instruction result in generalization to untrained words?

Effectiveness was measured using reading aloud and orthographic choice tasks. For both tasks, we predicted that all three methods of instruction would be more effective than the wait-list (business-as-usual) control. We predicted that mispronunciation correction would be more effective than Look and Spell and Look and Say, because it involves both explicitly processing the links between graphemes and phonemes and actively connecting partially decoded responses to existing vocabulary knowledge. In turn, we predicted that Look and Spell would be more effective than Look and Say because Look and Spell involves active processing of the details of a word's orthography. We predicted that generalization would be greatest for the mispronunciation correction group, because they were taught a strategy that could, in theory, be readily applied to unfamiliar words. We also predicted that there would be a greater degree of generalization on the reading measure than the orthographic choice measure, because choosing between multiple similarly spelled alternatives requires more precise orthographic knowledge than recognizing the written form of the word. Hypotheses were pre-registered on the Open Science Framework (https://osf.io/xku47/).

Materials and methods

Participants

Participants were 94 Kindergarten children from two government primary schools, one in metropolitan Sydney and one in a suburban area of coastal New South Wales, Australia. The first school had an Index of Community Socio-educational Advantage (ICSEA) score of 1104, indicating relative socio-educational advantage compared to the average value of 1000. The second school had an ICSEA score of 940, indicating relative socio-educational disadvantage. Systematic, explicit phonics instruction occurred at both schools, though the scope and sequence of instruction was different. Participating children were in their third term of their first year of formal schooling at the time of the study.

This study received ethical approval from the Macquare University Human Sciences Ethics Committee and the New South Wales State Education Research Applications Process (SERAP). Schools distributed information and consent forms to all Kindergarten children, but only those who returned signed parental consent forms and gave verbal assent took part. Two children were withdrawn from the study due to disruptive behaviour. A further four children were excluded because they were absent for two or more of the training sessions. There was missing data for one child who was absent at post-testing, and for two children who were absent for vocabulary pre-testing. Thus, the final sample was 85 children. The average age of participants was 6 years of age (range: 5 years 1 month – 6 years 8 months). Demographic characteristics of the sample are shown in Table 1. Letter Sound Test results

(Larsen et al., 2015; see below) indicated that our sample represented a very broad range of decoding abilities (mean percentile rank = 51, *SD* = 31, min = 1, max = 95).

Given the size of our sample, it was possible that simple randomization at the individual level may have led to baseline imbalances between groups. Therefore, class teachers were asked to assign participating children into mixed-ability groups of up to five children. Using an online randomization tool (https://www.graphpad.com/quickcalcs/randomize2/), each group was then randomly assigned to one of four conditions: Look and Say (LSay), Look and Spell (LSpell), Mispronunciation Correction (MPC) or control. There were five groups per condition.

[Insert Table 1 about here]

Procedure and Design

All methods and analysis procedures for the study were pre-registered on the Open Science Framework (https://osf.io/xku47). Data collection took place in Term 3 of the 2019 school year, which started in February and ended in December. Data collection was conducted in School 1 in late August, and in School 2 in early September. At each school, data collection took place over nine school days. The study timeline is shown in Figure 1. Children in the three training groups received 10 minutes of training in groups of five, over three consecutive days. Children in the wait-list control group received one training session from each of the three training conditions, after the final post-test had been completed.

[Insert Figure 1 about here]

Background assessment

Letter Sound Test (LeST, (Larsen et al., 2015). On this task, children are shown a series of 51 graphemes (e.g., m, a, sh, igh). They are asked to give the sound for each grapheme. The test has Australian norms. Reported test-retest reliability (intra-class correlation) for the test is 0.88. Cronbach’s alpha for the study sample was 0.92.

Word reading. Children were also asked to read aloud fourteen frequent words, some of which were regular and some of which contained an irregular grapheme-phoneme correspondence (*me, no, and, had, from, back, over, time, great, around, before, inside, through, children*). None of these words appeared in the training lessons. Cronbach's alpha was 0.90. Test administration for this task and the LeST was audio recorded, and online scoring was checked against these audio recordings. Ten percent of all data was triple-scored by three research assistants. Inter-rater agreement was very high (98.63%).

Vocabulary. The vocabulary task was an experimenter-designed, group administered multiple-choice task. Using a smart board, children were shown a written word and four pictures. The word was then read aloud and children were asked to tick the box on their worksheet that best matched the word. There were three practice items and 20 test items. In order to prevent copying, children were seated away from each other where possible and were given pieces of paper to cover up their work as they went.

Test items were selected by consulting a list of 297 words with age-of-acquisition norms derived from picture naming (Morrison, Chappell & Ellis, 1997), a set of age-of-acquisition norms derived from adult ratings (Kuperman et al., 2012), and a list of vocabulary words selected to be appropriate for American Kindergarten children (Flocabulary, 2020). From this larger pool, we selected words with an age-of-acquisition between 4 and 8.5. From this list, any duplicates and function words were removed. Items were piloted on five Kindergarten students not involved in the training aspect of the study. Using this data, twenty words were selected for inclusion in the final assessment (*heavy, giggle, gentle, jungle, peek, liquid, rhino, dangle, tornado, binoculars, nibble, escalator, cube, nostril, exhausted, cheetah, scrub, antlers, saxophone,* and *stationery*). Target and distractor pictures were then selected from Unsplash (https://unsplash.com) and Google Images. The task was relatively easy for children but there was no ceiling effect (min = 8, max = 19, mean = 15.0, *SD* = 2.64). Cronbach's alpha was 0.95.

Outcome measures

At post-test, children were assessed on their ability to read aloud and recognize the spellings of twelve trained and eight untrained words. We chose words that were likely to be familiar to children in their spoken but not their written forms, and we did not pre-test children on these words because we did not wish to expose them to the written forms prior to training. We used slightly fewer untrained words in order to reduce the length and difficulty of assessment sessions. In order to select stimuli, a long list of irregular words was compiled by examining stimuli from the Dyson et al. (2017) and Zipke (2016) studies and by selecting irregular words from lists of frequent words such as the Oxford Wordlist (https://www.oup.com.au/primary/oxford-wordlist). From this long list, words were retained if they were imageable and frequent in spoken language, but less frequent in written language. Words were rejected if they were homophones or if their decoded forms sounded similar to another non-target word. The resulting shortlist was piloted on five Kindergarten students not involved in the training aspect of the study. Items were rejected if more than two of the pilot participants could read them correctly. From the remaining words, 12 trained and eight untrained items were selected that could be matched on letter length, spoken frequency (CBBC frequencies from Subtlex UK; van Heuven et al., 2014), number of irregular spelling-sound correspondences, number of syllables, and number of orthographic neighbours calculated using both CLEARPOND (Marian et al., 2012) and the Children's Printed Word Database (Masterson et al., 2010) – see Table 2.

[Insert Table 2 about here]

Reading aloud. In the reading aloud task, children were asked to read aloud the trained and untrained words, interspersed, from a printed list. Children were told that they might know some of the words, but not others. No corrective feedback was given. Test administration for this task was audio recorded, and online scoring was checked against these audio recordings. All data for this task was initially entered and checked by a single research assistant, who was present at two control training sessions at one of the schools, but was otherwise blind to group membership. Data was then checked by a second research assistant. Inter-rater agreement was high (90.22%). Cronbach's alpha was 0.93.

Orthographic choice. The orthographic choice task was group-administered. In this task, children were given a printed worksheet and a pencil. For each trained and untrained word, the correct written form was printed along with two foils. One of the foils was always a regularized version of the word, while the other was one letter different from the correct spelling (e.g. *yung*, *yaung*, *young*). The order in which the correct form and foils appeared was randomized, as was the order in which trained and untrained items appeared. However, all children completed a single version of the test and saw the items in the same order. Children heard each word spoken aloud and were asked to circle which of three options was the correct way to spell it. Children completed two practice items with feedback before completing the remaining test items without feedback. Cronbach's alpha for all items (trained and untrained) was 0.70.

Instruction

Training sessions were scripted – examples are shown in Figure 2 and full scripts are available at https://osf.io/wnrt5/. Typically, MPC instruction involves reading words in sentence context (e.g., see Dyson et al., 2017), while sight word instruction (e.g., look-say-cover-write-check) involves teaching single words in isolation. In this study, we wished to compare the instructional methods in their most basic forms in order to determine their "active ingredients". Therefore, we decided to teach MPC at the single word level, using pictures to provide semantic context. The use of pictures rather than sentence context allowed us to equate instructional time and degree of contextual support across conditions. A further reason to teach MPC in single words is that our participants were only in their third term of formal schooling, so the majority were only able to read very simple connected text, limiting the usefulness of written sentence context.

[Insert Figure 2 about here]

All four groups learnt the same twelve words (see Table 2), but the wait-list control group received instruction after post-testing was completed. Exposure to the written form of the word was equal across all experimental conditions; the written form was displayed on the whiteboard throughout each training session. Exposure to the correct spoken form was also controlled; each word was said aloud four times by the trainer. Thus, in total across all lessons, children were exposed to the written form of the word three times, and heard the correct spoken form of the word 12 times.

Training was delivered by the first, fifth and sixth authors, and two additional research assistants. Due to the availability of personnel, it was not possible to evenly distribute trainers across groups - the first author delivered 47% of the sessions in the LSay condition, 93% of sessions in the LSpell condition and 73% of sessions in the MPC condition. However, every effort was made to ensure that instruction was consistent across groups. Firstly, instruction was tightly scripted in each condition. Secondly, quality of instruction was carefully controlled. All trainers received two hours of instruction in how to deliver the lessons from the first author. After this, they observed the first author deliver at least one training session. The first author then observed each trainer's initial training session and provided feedback before they conducted training sessions independently.

Note that in the LSpell and MPC conditions, all but one group had at least two out of the three lessons delivered by the first author. In the LSay condition, three of the groups had at least two lessons delivered by the first author, while two of the groups did not have any lessons delivered by the first author. We therefore examined the raw data for the groups in the LSay condition. This revealed no indication of trainer effects (see Appendix B).

Analyses

Data was analysed in two multilevel logistic regression models, one predicting post-test reading accuracy and the other predicting post-test spelling (orthographic choice) accuracy. Continuous variables were grand-mean centered and standardised before analysis. Analyses were conducted in the R software environment using *lme4* (Bates et al., 2015), *lmerTest* (Kuznetsova et al., 2017) and *emmeans* (Lenth, 2021) packages. Data and analysis scripts are available on the Open Science framework: https://osf.io/cj7xr/.

Models contained fixed factors for group and item type (trained or untrained). They also contained fixed factors for pre-test reading score (composite LeST and word reading scores) and pre-test vocabulary, which were treated as covariates. Analysis code with full details of our model fitting procedure are available on the Open Science Framework (see above). Initially, "maximal" models were fitted with all relevant random intercepts and slopes (Barr, Levy, Scheepers & Tily, 2013). However, these models failed to converge. Random slopes were removed one by one until acceptable convergence was achieved; the final models contained only random intercepts for participants and items. Contrasts were specified comparing each training group to the control group, and each training group to the two other trained groups (e.g., LSay to LSpell, LSay to MPC, and MPC to LSpell) for both trained and untrained items.

Results

Baseline differences between groups

A series of one-way ANOVAs indicated that there were no significant differences between the four groups on baseline Letter Sound Test [$F(3, 81) = 0.13$, $p = 0.94$], word reading [$F(3, 81) = 0.14$, $p = 0.94$], or vocabulary scores [$F(3, 81) = 1.50$, $p = 0.22$]. There were also no significant differences between the experimental groups in terms of number of lessons attended [F(2, 60) = 1.15, *p* = 0.32]. Chi-squared tests indicated no significant differences between the groups in terms of number of females in each group [$\chi^2(3) = 2.20$, p = 0.53], the number of students speaking English at home [$\chi^2(3) = 0.67$, p = 0.88], or the number of children in each group with a developmental diagnosis [$\chi^2(3) = 5.87$, p = 0.12] (see Table 1).

Reading task

Raw score means and confidence intervals for the trained and untrained items are shown in Figure 3. Full results of the analysis model are shown in Appendix A. Contrasts revealed that for the trained items, all trained groups were significantly more likely to read the items correctly than controls (LSay vs Control: β = 1.65, *SE* = 0.68, *z* = 2.43, *p* = 0.02; LSpell vs Control: β = 3.34, *SE* = 0.63, *z* = 5.27, *p* <0.001; MPC vs Control: β = 3.48, *SE* = 0.63, *z* = 5.54, *p* <0.001). In other words, there were clear training effects for all three training conditions compared to business-as-usual controls. There was no significant difference between the LSpell and MPC groups (β = 0.14, *SE* = 0.50, *z* = 0.29, *p* = 0.77), but children in the LSpell and MPC groups were significantly more likely to read the items correctly than children in the LSay group (LSpell vs LSay: β = -1.69, *SE* = 0.59, *z* = 2.88, *p* = 0.004; MPC vs LSay: β = -1.84, *SE* = 0.55, *z* = 3.32, *p* <0.001). In other words, LSpell and MPC instruction were equally effective, and both were more effective than LSay training (see Figure 3).

For the untrained items, the results for the trained groups were not significantly different from controls (LSay vs Control: β = 0.31, *SE* = 0.70, *z* = 0.44, *p* = 0.66; LSpell vs Control: β = 0.40, *SE* = 0.66, *z* = 0.62, *p* = 0.54; MPC vs Control: β = 0.14, *SE* = 0.68, *z* = 0.20, *p* = 0.84). There were also no significant differences between any of the training groups (LSay vs LSpell: β = -0.09, *SE* = 0.69, *z* = 0.13, *p* = 0.89; LSay vs MPC: β = 0.18, *SE* = 0.69, *z* = 0.26, *p* = 0.80; LSpell vs MPC: β = -0.30, *SE* = 0.66, *z* = 0.41, *p* = 0.68). In other words, there was no evidence that any of the training methods had an effect on the untrained words (see Figure 3).

[Insert Figure 3 about here]

Orthographic choice task

As above, results of the full analysis model are shown in Appendix A and Figure 4 summarises the raw scores across conditions. Contrasts revealed that for the trained items, there was no significant difference between the LSay group and controls (β = 0.18, *SE* = 0.40, *z* = 0.46, *p* = 0.65), but both the LSpell and MPC groups were significantly more likely to choose the correct response than controls (LSpell vs Control: β = 0.82, *SE* = 0.38, *z* = 2.13, *p* = 0.03; MPC vs Control: β = 0.88, *SE* = 0.38, *z* = 2.35, *p* = 0.02). In other words, an effect of training was apparent for the LSpell and MPC groups, but not for the LSay group.

The difference between the LSay and LSpell groups was not significant (β = -0.63, *SE* = 0.40, *z* = 1.58, *p* = 0.11), nor was the difference between the LSay and MPC groups (β = -0.70, *SE* = 0.38, *z* = 1.83, *p* = 0.07), or the difference between the MPC and LSpell groups (β = 0.07, *SE* = 0.38, *z* = 0.19, *p* = 0.85).

For the untrained items, there were no significant differences between trained groups and controls (LSay vs Control: β = 0.52, *SE* = 0.34, *z* = 1.54, *p* = 0.12; LSpell vs Control: β = 0.47, *SE* = 0.33, *z* = 1.42, *p* = 0.16; MPC vs Control: β = -0.03, *SE* = 0.33, *z* = 0.11, *p* = 0.92). Likewise, there were no significant differences between the training groups (LSay vs LSpell: β = 0.05, *SE* = 0.34, *z* = 0.15, *p* = 0.88; LSay vs MPC: β = 0.55, *SE* = 0.33, *z* = 1.67, *p* = 0.09; LSpell vs MPC: β = -0.50, *SE* = 0.34, *z* = 1.50, *p* = 0.13).

[Insert Figure 4 about here]

It is important to note that although the MPC and LSpell groups performed better than controls on the trained items, their performance was at approximately chance levels (see

Figure 4). In contrast, the Control and LSay groups were performing at below chance levels, suggesting that they may have been systematically choosing incorrect responses. In order to explore this possibility, we calculated the proportion of correct, regularized and one-letter-different responses for each group (see Table 3).

[Insert Table 3 about here]

All groups were more likely to choose the regularized option than the other two response options. However, this tendency was less pronounced for the children in the LSpell and MPC groups. In order to conduct exploratory Chi square tests comparing response proportions, we collapsed together the results of the two groups involving explicit attention to orthography (the LSpell and MPC groups) and the two groups that did not (the Control and LSay groups). The combined LSpell and MPC groups showed a greater tendency to choose correct responses for the trained words than the combined LSay and control groups, $\chi^2(1) = 17.46$, $p < 0.001$, but a comparison of incorrect responses indicated that the combined LSpell and MPC groups also showed greater tendency to choose one-letter-different distractors compared to combined LSay and control groups, $\chi^2(1) = 9.91$, $p = 0.002$. Thus, in the LSpell and MPC groups, training seemed to shift children towards a response strategy that involved more attention to the orthography of the trained items, but it was not enough to ensure that children were reliably able to choose the correct responses.

Discussion

We compared the effectiveness of two different methods of sight word instruction (Look and Say and Look and Spell) to a method of mispronunciation correction instruction for teaching irregular words to beginning readers. Using a randomized design, we were interested in effects on words that were directly trained, and whether the different methods would have different consequences for generalization to untrained words. Overall, we predicted that Mispronunciation Correction (MPC) would be the most effective method of instruction because it involved both explicitly processing the links between graphemes and phonemes, and actively connecting partially decoded responses to existing spoken vocabulary knowledge. We predicted that Look and Spell (LSpell), which involved active processing of the word's orthography during writing but no explicit attention to decoding or active connection to existing vocabulary knowledge, would be less effective than MPC but more effective than Look and Say (LSay), which did not involve any explicit processing of orthography. We also predicted that all three methods would be more effective than a business-as-usual control.

In terms of effects on trained words, our predictions were partially supported. All three instruction methods were more effective than a business-as-usual control condition for improving children's ability to read the trained words aloud. Children in the LSpell and MPC groups were more likely to read the trained items correctly than children in the LSay condition, but there was no evidence of a difference between the LSpell and MPC groups. Thus, active processing of the words' orthographic forms appeared to be important, but explicitly decoding the words and connecting partially decoded forms to spoken vocabulary (as in the MPC condition) did not seem to provide additional benefits compared to writing the words, at least after the short period of instruction provided in our study.

On the orthographic choice outcome measure, children in the LSay condition were no more likely to choose the correct response than children in the control group. Children in the LSpell and MPC groups were significantly more likely than controls to choose the correct responses for trained words, but again, there was no significant difference between these two groups. In other words, performance in the LSpell and MPC groups was superior to the performance of the control group. Active processing of the orthography of the trained words appeared to be crucial for children's ability to choose the correct orthographic forms.

Although children in the LSpell and MPC groups were more likely to choose correct responses for the trained words, they only selected the correct responses approximately one third of the time. In fact, children in *all* groups were more likely to choose the regularized distractors (e.g., *jellus*) than the correct forms of the words (*jealous*). This is perhaps unsurprising given that the children were receiving explicit instruction in regular grapheme-phoneme correspondences as part of their classroom instruction. In the context of this overall tendency to favour regularized distractors, the LSpell and MPC children were more likely than the other groups to choose the correct responses, but they were also more likely than the other groups to choose a one-letter different distractor (e.g., *jealus)*. Therefore, the LSpell and MPC children did show some evidence of a shift towards a response strategy involving attention to the orthography of the trained words, but they were not always able to select the correct response.

At first glance, this echoes Zipke's (2016) finding that participants who received mispronunciation correction training made more attempts to correct their own pronunciations, but those attempts did not lead to more correct responses. However, Zipke assessed participants on words that were not seen during training, meaning that she was assessing application of a general strategy, not orthographic learning. Together, the results of our reading and orthographic choice tasks suggest that children in the MPC and LSpell conditions had begun to develop orthographic representations for the trained words, but that these representations were not sufficiently detailed or stable to allow for reliably correct recognition of the words' spellings in the presence of relatively close distractors.

Previous work has shown that competent readers in Grade 1 can learn the orthography of new words after approximately four massed attempts at reading aloud with corrective feedback (Ehri & Saltmarsh, 1995; Reitsma, 1983). Thus, there was reason to believe that three spaced sessions, during which the written form was visible for at least 10 minutes, should have resulted in a significant amount of learning, especially considering the additional processing undertaken in the LSpell and MPC conditions. However, in future, it would be interesting to see whether results would improve after an additional training session.

Turning to the untrained words, there was no evidence of generalization in any of the training conditions. This may also be due to the relatively short duration of our training programs. In a study using a variant of “Look, Say, Cover, Write, Check” instruction where generalization was observed (Kohnen et al., 2008), a child with difficulties spelling irregular words received 10 sessions of instruction over two weeks. In a study of mispronunciation correction where generalization to matched untrained words was observed (Dyson et al., 2017), children received eight training sessions over four weeks, and discussed word meanings, wrote and matched words, and completed rhyming tasks, in addition to correcting mispronunciations. Thus, our training may not have been sufficient to allow for the development of the orthographic knowledge or strategic processes necessary to result in improvement on untrained words.

A further reason for a lack of observed generalization, and for the relatively small effects of training, may be the fact that the words we taught were somewhat advanced for children in their third term of formal schooling. This was out of necessity – we had to choose words that were concrete, easily represented by a picture, matched to other words of similar length and frequency, and unfamiliar to the majority of participants in their written form. This ruled out many of the shortest and most frequent irregular words, which tend to be function words (e.g., *they*, *are*).

Although our study was the first to directly compare these methods of irregular word reading instruction under tightly controlled experimental conditions, there were some limitations. Firstly, our sample size was comparatively small, as it was challenging to recruit schools who were willing to allow their Kindergarten students to participate in a multi-day study during class time. However, it is worth noting that our training groups were roughly the size of an average Kindergarten classroom, and our participants represented a broad range of reading ability. A second limitation of our study was the fact that we did not conduct a follow-up assessment in order to determine whether knowledge was retained over time. Again, this was a practical decision made to reduce the load on participating schools, but it is a crucial question for future studies to explore.

Although mispronunciation correction instruction is traditionally delivered in the context of written or spoken sentences (e.g., Dyson et al., 2017; Zipke, 2016), our version of mispronunciation correction instruction took place at the single word level, without the presence of meaningful sentence context. This was crucial in order to equate instruction as much as possible across conditions and to allow a direct comparison with LSay and LSpell. It was also appropriate for our participants, who had limited reading experience and may have struggled to make use of written sentence context. We speculate that the presence of sentence context may help children to learn to apply mispronunciation correction as a general strategy, but it may be less important for orthographic learning of specific words. Studies have shown that young children tend to learn the orthographic forms of words better when they read them in isolation rather than in context (e.g., Ehri & Roberts, 1979; Ehri & Wilce, 1980; Landi et al., 2006; Miles & Ehri, 2017; Stuart et al., 2000).

A lack of sentence context at the point of assessment may be a reason why we did not observe generalization in the mispronunciation correction condition – children were asked to read both trained and untrained words in isolation, and could not use context to support mispronunciation correction attempts, although they could in theory access the correct spoken form using their existing vocabulary knowledge (e.g., see Tunmer & Chapman, 2012). There is clearly a need for further studies exploring the role that sentence context plays in mispronunciation correction, both in terms of item-level orthographic learning and in terms of general strategy learning.

While our results point to a key role for active processing of orthography in irregular word reading instruction, they do not clarify whether different methods of processing have different consequences for orthographic learning. We did not find clear evidence of an advantage for our MPC condition (which involved explicit processing of grapheme-phoneme correspondences and linking of a partially decoded form to a correct phonological form) over the LSpell condition (which involved explicitly retrieving and writing the letters in the words but no explicit decoding). However, it is possible that differences may emerge with larger sample sizes, after longer periods of training, or for different words, and we cannot rule out the fact that some students may have been engaging in some degree of implicit grapheme-phoneme processing in the LSpell condition. It would be interesting to compare MPC to LSpell instruction over a longer period of time, measuring the rate of learning as well as longer-term retention, to see whether differences emerge. It would also be interesting to determine whether a combination of decoding, matching decoded forms to correct spoken forms, *and* spelling would be more effective than either MPC or LSpell alone for orthographic learning.

In summary, our results clearly support the view that irregular word reading instruction should include active processing of a word's orthography during instruction. Our results provide evidence that Look and Spell and mispronunciation correction instruction are equally effective for teaching irregular words, at least after short-term instruction. Finally, our results provide clear directions for future research. A more detailed understanding of the active ingredients of irregular word reading instruction is crucial for maximizing the efficiency and effectiveness of early reading instruction, so that all children have the chance to become fluent independent readers.

Disclosure statement

There are no relevant financial or non-financial competing interests to declare.

Data availability statement

Data and analysis code for this study are available at https://osf.io/cj7xr/

Acknowledgements

This work was supported by a grant from the Australian Research Council (grant number DP200100311) to Anne Castles, Kate Nation, Lisi Beyersmann and Erik Reichle. We would like to thank the participating children and their parents, class teachers and senior staff at participating primary schools, and Sarah Coen and Ashna Arora for their assistance with data collection and data entry.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Bowers, J. S., & Bowers, P. N. (2017). Beyond Phonics: The Case for Teaching Children the Logic of the English Spelling System. *Educational Psychologist, 52*(2), 124-141. https://doi.org/10.1080/00461520.2017.1288571

Broom, Y. M., & Doctor, E. A. (1995). Developmental surface dyslexia: a case study of the efficacy of a remediation programme. *Cognitive Neuropsychology, 12*(1), 69-110. https://doi.org/10.1080/02643299508251992

Brunsdon, R., Coltheart, M., & Nickels, L. (2005, Mar). Treatment of irregular word spelling in developmental surface dysgraphia. *Cognitive Neuropsychology, 22*(2), 213-251. https://doi.org/10.1080/02643290442000077

Colenbrander, D., Wang, H. C., Arrow, T., & Castles, A. (2020). Teaching irregular words: What we know, what we don't know, and where we can go from here. *The Educational and Developmental Psychologist, 37*(2), 97-104. https://doi.org/10.1017/edp.2020.11

Conrad, N. J., Kennedy, K., Saoud, W., Scallion, L., & Hanusiak, L. (2018). Establishing word representations through reading and spelling: comparing degree of orthographic learning. *Journal of Research in Reading, 42*(1), 162-177. https://doi.org/10.1111/1467-9817.12256

Dyson, H., Best, W., Solity, J., & Hulme, C. (2017). Training Mispronunciation Correction and Word Meanings Improves Children's Ability to Learn to Read Words. *Scientific Studies of Reading, 21*(5), 392-407. https://doi.org/10.1080/10888438.2017.1315424

Ehri, L. C. (2005). Learning to read words: Theories, findings and issues. *Scientific Studies of Reading, 9*(2), 167-188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C. (2014, Jan). Orthographic Mapping in the Acquisition of Sight Word Reading, Spelling Memory, and Vocabulary Learning. *Scientific Studies of Reading, 18*(1), 5-21. https://doi.org/10.1080/10888438.2013.819356

Ehri, L. C., & Roberts, K. T. (1979). Do beginners learn printed words better in contexts or in isolation? *Child Development, 50*(3), 675-685. https://doi.org/https://doi.org/10.2307/1128932

Ehri, L. C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing, 7*, 295-326. https://doi.org/10.1007/BF03162082

Ehri, L. C., & Wilce, L. S. (1980). Do beginners learn to read function words better in sentences or in lists? *Reading Research Quarterly, 15*(4), 451-476. https://doi.org/https://doi.org/10.2307/747274

Elbro, C., de Jong, P. F., Houter, D., & Nielsen, A.-M. (2012). From Spelling Pronunciation to Lexical Access: A Second Step in Word Decoding? *Scientific Studies of Reading, 16*(4), 341-359. https://doi.org/10.1080/10888438.2011.568556

Flocabulary. (2020). *Wordlists for Grades K-8.* Retrieved from https://www.flocabulary.com/wordlists/

Galuschka, K., Ise, E., Krick, K., & Schulte-Korne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomized controlled trials. *PloS One, 9*(2), e89900. https://doi.org/10.1371/journal.pone.0089900

Hilte, M., & Reitsma, P. (2006). Spelling pronunciation and visual preview both facilitate learning to spell irregular words. *Annals of Dyslexia, 56*, 301-318. https://doi.org/ 10.1007/s11881-006-0013-3

Joshi, R. M., Treiman, R., Carreker, S., & Moats, L. C. (2008). How words cast their spell: Spelling is an integral part of learning the language, not a matter of memorization. *American Educator, Winter 2008-2009*, 6-16.

Kearns, D. M., Rogers, H. J., Koriakin, T., & Al Ghanem, R. (2016). Semantic and Phonological Ability to Adjust Recoding: A Unique Correlate of Word Reading Skill? *Scientific Studies of Reading, 20*(6), 455-470. https://doi.org/10.1080/10888438.2016.1217865

Kohnen, S., Nickels, L., Brunsdon, R., & Coltheart, M. (2008, Feb). Patterns of generalisation after treating sub-lexical spelling deficits in a child with mixed dysgraphia. *Journal of Research in Reading, 31*(1), 157-177. https://doi.org/10.1111/j.1467-9817.2007.00366.x

Kohnen, S., Nickels, L., Coltheart, M., & Brunsdon, R. (2008, May). Predicting generalization in the training of irregular-word spelling: treating lexical spelling deficits in a child. *Cogn Neuropsychol, 25*(3), 343-375. https://doi.org/10.1080/02643290802003000

Kuperman, V., Stadthagen-Gonzales, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behaviour Research Methods, 44*, 978-990. https://doi.org/10.3758/s13428-012-0210-4

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

Landi, N., Perfetti, C. A., Bolger, D. J., Dunlap, S., & Foorman, B. R. (2006, Jun). The role of discourse context in developing word form representations: a paradoxical relation between reading and learning. *Journal of Experimental Child Psychology, 94*(2), 114-133. https://doi.org/10.1016/j.jecp.2005.12.004

Larsen, L., Kohnen, S., Nickels, L., & McArthur, G. (2015). The Letter-Sound Test (LeST): a reliable and valid comprehensive measure of grapheme–phoneme knowledge. *Australian Journal of Learning Difficulties, 20*(2), 129-142. https://doi.org/https://doi.org/10.1080/19404158.2015.1037323

Lenth, R. V. (2021). *emmeans: Estimated marginal means, aka least-squares means.* In (Version 1.5.4) [R package]. https://CRAN.R-project.org/package=emmeans

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic Easy-Access Resource for Phonological and Orthographic Neighbourhood Densities. *PloS One, 7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children's printed word database: continuities and changes over time in children's early reading vocabulary. *British Journal of Psychology, 101*(Pt 2), 221-242. https://doi.org/10.1348/000712608X371744

McArthur, G., Castles, A., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., & Banales, E. (2015, Jul-Aug). Sight Word and Phonics Training in Children With Dyslexia. *Journal of Learning Disabilities, 48*(4), 391-407. https://doi.org/10.1177/0022219413504996

McArthur, G., Kohnen, S., Jones, K., Eve, P., Banales, E., Larsen, L., & Castles, A. (2015). Replicability of sight word training and phonics training in poor readers: a randomised controlled trial. *PeerJ, 3*. https://doi.org/10.7717/peerj.922

McArthur, G., Sheehan, Y., Badcock, N. A., Francis, D. A., Wang, H. C., Kohnen, S., Banales, E., Anandakumar, T., Marinus, E., & Castles, A. (2018). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews*(11), Article CD009115. https://doi.org/10.1002/14651858.CD009115.pub3

Miles, K. P., & Ehri, L. C. (2017). Learning to Read Words on Flashcards: Effects of Sentence Contexts and Word Class in Native and Nonnative English-Speaking Kindergartners. *Early Childhood Research Quarterly, 41*, 103-113. https://doi.org/10.1016/j.ecresq.2017.06.001

Miles, K. P., Rubin, G. B., & Gonzalez-Frey, S. (2018). Rethinking Sight Words. *The Reading Teacher, 71*(6), 715-726. https://doi.org/10.1002/trtr.1658

Moats, L. C. (2005). How spelling supports reading, and why it is more regular and predictable than you think. *American Educator, Winter 2005/2006*, 12-22.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Ace of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology A, 50*(3), 528-559. https://doi.org/10.1080/027249897392017

Nation, K. (2017). Nurturing a lexical legacy: reading experience is critical for the development of word reading skill. *npj Science of Learning, 2*(1). https://doi.org/10.1038/s41539-017-0004-7

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: Reports of the subgroups* (NIH Publication No. 00-4769). National

Institute of Child Health and Human Development. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

Ouelette, G. (2010). Orthographic learning in learning to spell: The roles of semantics and type of practice. *Journal of Experimental Child Psychology, 107*(1), 50-58. https://doi.org/10.1016/j.jecp.2010.04.009

Perfetti, C.A. (1997). The psycholinguistics of spelling and reading. In C.A. Perfetti, L. Rieben & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages,* (pp. 21–38). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Potter, D. (2012). Sight words [Blog post]. http://blendphonics.org/?page_id = 765

Reitsma, P. (1983). Printed word learning in beginning readers. *Journal of Experimental Child Psychology, 36*, 321-339. https://doi.org/10.1016/0022-0965(83)90036-X

Savage, R., Georgiou, G., Parrila, R., & Maiorino, K. (2018). Preventative Reading Interventions Teaching Direct Mapping of Graphemes in Texts and Set-for-Variability Aid At-Risk Learners. *Scientific Studies of Reading, 22*(3), 225-247. https://doi.org/10.1080/10888438.2018.1427753

Shahar-Yames, D., & Share, D. L. (2008). Spelling as a self-teaching mechanism in orthographic learning. *Journal of Research in Reading, 31*(1), 22-39. https://doi.org/10.1111/j.1467-9817.2007.00359.x

Shapiro, L. R., & Solity, J. (2016). Differing effects of two synthetic phonics programmes on early reading development. *British Journal of Educational Psychology, 86*(2), 182-203. https://doi.org/10.1111/bjep.12097

Share, D. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition, 55*(2), 151-218. https://doi.org/10.1016/0010-0277(94)00645-2

Solity, J., & Vousden, J. (2009). Real books vs reading schemes: a new perspective from instructional psychology. *Educational Psychology, 29*(4), 469-511. https://doi.org/10.1080/01443410903103657

Steacy, L. M., Elleman, A. M., Lovett, M. W., & Compton, D. L. (2016). Exploring Differential Effects Across Two Decoding Treatments on Item-Level Transfer in Children with Significant Word Reading Difficulties: A New Approach for Testing Intervention Elements. *Scientific Studies of Reading, 20*(4), 283-295. https://doi.org/10.1080/10888438.2016.1178267

Steacy, L. M., Wade-Woolley, L., Rueckl, J. G., Pugh, K. R., Elliot, J. D., & Compton, D. L. (2019). The role of Set for Variability in irregular word reading: Word and child predictors in typically developing readers and students at-risk for reading disabilities. *Scientific Studies of Reading, 23*, 523-532. https://doi.org/10.1080/10888438.2019.1620749

Stuart, M., Masterson, J., & Dixon, M. (2000). Spongelike acquisition of sight vocabulary in beginning readers? *Journal of Research in Reading, 23*(1), 12-27. https://doi.org/https://doi.org/10.1111/1467-9817.00099

Terzopoulos, A. R., Niolaki, G. Z., & Masterson, J. (2018). Intervention for a lexical reading and spelling difficulty in two Greek-speaking primary age children. *Neuropsychological Rehabilitation*, 1-22. https://doi.org/10.1080/09602011.2018.1467330

Tunmer, W. E., & Chapman, J. W. (2012). Does Set for Variability mediate the influence of vocabulary knowledge on the development of word recognition skills? *Scientific Studies of Reading, 16*, 122-140. https://doi.org/10.1080/10888438.2010.542527

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of*

Experimental Psychology, 67, 1176-1190. https://doi.org/10.1080/17470218.2013.850521

Zipke, M. (2016). The importance of flexibility of pronunciation in learning to decode: A training study in set for variability. *First Language, 36, 71-86.* https://doi.org/10.1177/0142723716639495

Appendix A

Unconditional model – reading

Fixed Effects	Estimate	SE	*z*	*p*
Intercept	-3.82	0.53	-7.23	<0.0001
Random Effects	Variance		SD	
Intercept (Participant)	8.52		2.92	
Intercept (Item)	1.74		1.32	
Model Fit	AIC	Pseudo-R^2 (Conditional)		
	936.4	0.76		

Results of final analysis model – reading

Fixed Effects	Estimate	SE	*z*	*p*
Intercept	-5.43	0.66	-8.27	<0.0001
Reading composite score	2.31	0.26	8.99	<0.0001
Vocabulary score	0.30	0.24	1.27	0.20
Group – LSay v control	1.65	0.68	2.43	0.02
Group – LSpell v control	3.34	0.63	5.28	<0.0001
Group – MPC v control	3.48	0.63	5.54	<0.0001
Trained vs untrained items	0.62	0.74	0.84	0.40
Group x item type interaction, LSay	-1.33	0.72	-1.87	0.06
Group x item type interaction, LSpell	-2.94	0.67	-4.39	<0.0001
Group x item type interaction, MPC	-3.35	0.68	-4.96	<0.0001
Random Effects	Variance		SD	
Intercept (Participant)	1.19		1.09	
Intercept (Item)	1.25		1.12	
Model Fit	AIC	Pseudo-R^2 (Conditional)	Pseudo-R^2 (Marginal)	
	807.0	0.75	0.57	

Unconditional model – orthographic choice

Fixed Effects	Estimate	SE	z	p
Intercept	-1.00	0.16	-6.32	<0.0001
Random Effects	Variance		SD	
Intercept (Participant)	0.57		0.76	
Intercept (Item)	0.30		0.54	
Model Fit	AIC	Pseudo-R^2 (Conditional)		
	1947.5	0.21		

Results of final analysis model – orthographic choice

Fixed Effects	Estimate	SE	z	p
Intercept	-1.61	0.33	-4.94	<0.0001
Reading composite score	0.31	0.15	2.03	0.04
Vocabulary score	0.05	0.13	0.38	0.70
Group – LSay v control	0.18	0.40	0.46	0.65
Group – LSpell v control	0.81	0.38	2.13	0.03
Group – MPC v control	0.88	0.38	2.35	0.02
Trained vs untrained items	0.32	0.44	0.72	0.47
Group x item type interaction, LSay	0.34	0.48	0.70	0.49
Group x item type interaction, LSpell	-0.34	0.47	-0.73	0.47
Group x item type interaction, MPC	-0.92	0.47	-1.96	0.05
Random Effects	Variance	SD	Correlation	
Intercept (Participant)	0.97	0.98		
Slope (Participant x item type)	1.08	1.04	-0.78	
Intercept (Item)	0.37	0.61		
Slope (Item x reading composite)	0.20	0.44	-0.18	
Slope (Item x vocabulary score)	0.07	0.26	-0.73 -0.83	
Model Fit	AIC	Pseudo-R^2 (Conditional)	Pseudo-R^2 (Marginal)	
	1892.7	0.33	0.05	

Appendix B

Analysis of trainer effects

We wished to check whether there was any evidence of trainer effects. In the LSpell and MPC conditions, the majority of lessons were delivered by the first author. However, in the LSay condition, three of the groups had at least two lessons delivered by the first author (groups 1-3 below), while two of the groups did not have any lessons delivered by the first author (groups 4 and 5). Therefore, any trainer differences would have been most evident within the LSay condition. We examined the raw scores within the LSay group. This data is summarised below. There was no clear indication that groups receiving training mostly from the first author (1-3) scored higher than groups receiving no training from the first author (4-5). It therefore appears unlikely that trainer effects were playing a role. Rather, it seems more likely that participant outcomes were influenced by existing reading ability – there was a high correlation between pre-test reading scores and post-test reading outcomes in this group (Pearson's r=0.83, $p < 0.001$).

Group	Lessons delivered by first author	Pre-test reading score		Post-test trained reading score		Post-test trained orthographic choice score	
		Mean	Range	Mean	Range	Mean	Range
LSay 1	3	5.6	3-11	2	0-8	1.8	1-3
LSay 2	2	4	2-5	0.25	0-1	1.75	1-3
LSay 3	2	6	0-14	2.8	0-12	5.6	1-11
LSay 4	0	8	3-12	2	0-6	1.67	1-3
LSay 5	0	2	0-4	0	0	2.33	1-4

Figure 1. Study timeline

	Week 1		**Week 2**					**Week 3**	
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9
Trained groups	Reading and GPC knowledge assessment	Vocabulary assessment	Training session 1	Training session 2	Training session 3	Word reading post-test	Orthographic choice post-test	Business as usual	Business as usual
Wait-list control	Reading and GPC knowledge assessment	Vocabulary assessment	Business as usual	Business as usual	Business as usual	Word reading post-test	Orthographic choice post-test, Training session 1	Training session 2	Training session 3

Figure 2. Example training scripts

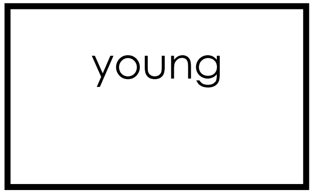
	Slide 1	**Slide 2**
		
Look and Say	**Trainer:** Look at this word. This word is young. Young.	**Trainer:** Say young. **Children:** *Young*. **Trainer:** What is the word: **Children:** *Young*. **Trainer:** Very good! The word is young.
Look and Spell	**Trainer:** Look at this word. This word is young. Young.	**Trainer:** Write young on your whiteboard. Young. **Children:** [write] **Trainer:** Show me what you wrote. OK, great work. Now rub it out. We will have one more go. Have a look at the word on the screen. Write it on your board. **Children:** [write] **Trainer:** Good job everyone. Rub out the word.
MPC	**Trainer:** I'll sound this word out. /y/ /aʊ/ /ŋ/, /yaʊŋ/. You sound it out with me. [points to graphemes as word is sounded out] /y/ /aʊ/ /ŋ/, /yaʊŋ/ **Children:** */y/ /aʊ/ /ŋ/, /yaʊŋ/* **Trainer:** Is /yaʊŋ/ a word you know? **Children:** *No* **Trainer:** So, /yaʊŋ/ is not a real word, but does it sound a bit like a word we know? Just think about it, don't tell me.	**Trainer:** Yes, the word is actually young. Young. Can you say it? **Children:** *Young.* **Trainer:** Good job. Say it again. **Children:** *Young.* **Trainer:** Very good! The word is young. Young.

Figure 3. Proportion correct on the reading outcome measure.

Note: Error bars are 95% confidence intervals.

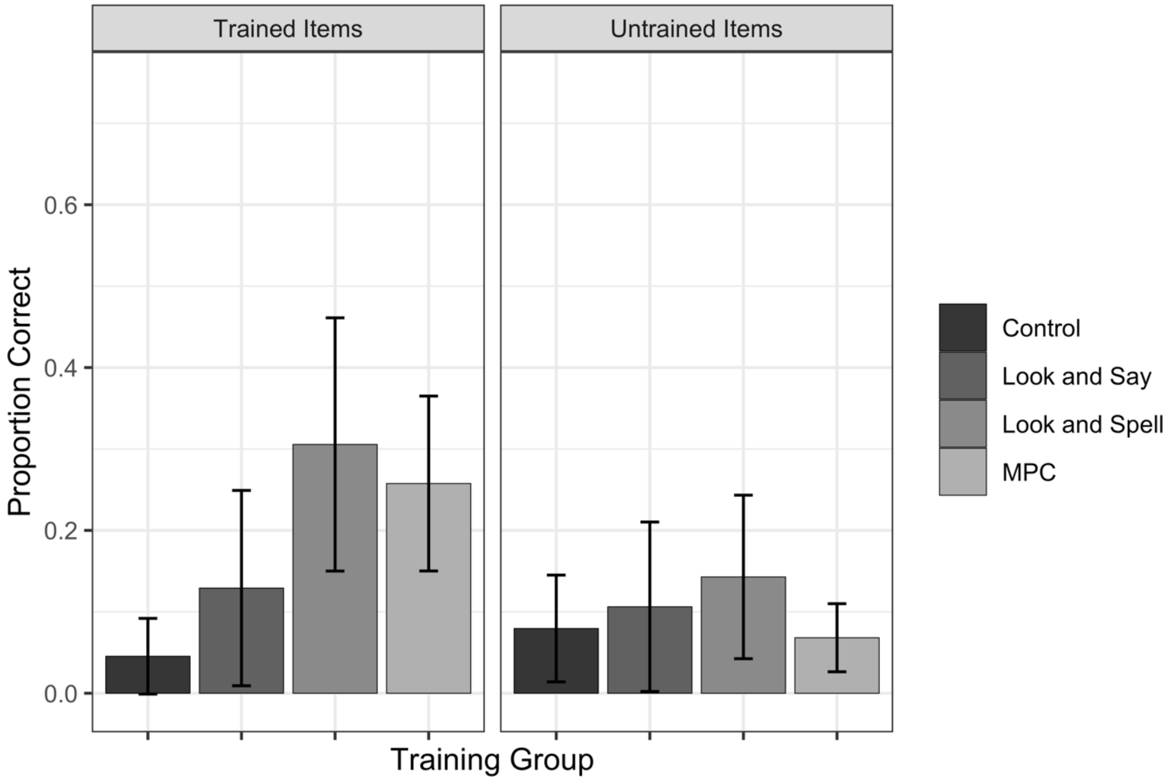


Figure 4. Proportion correct on the orthographic choice outcome measure.

Note: Error bars are 95% confidence intervals.

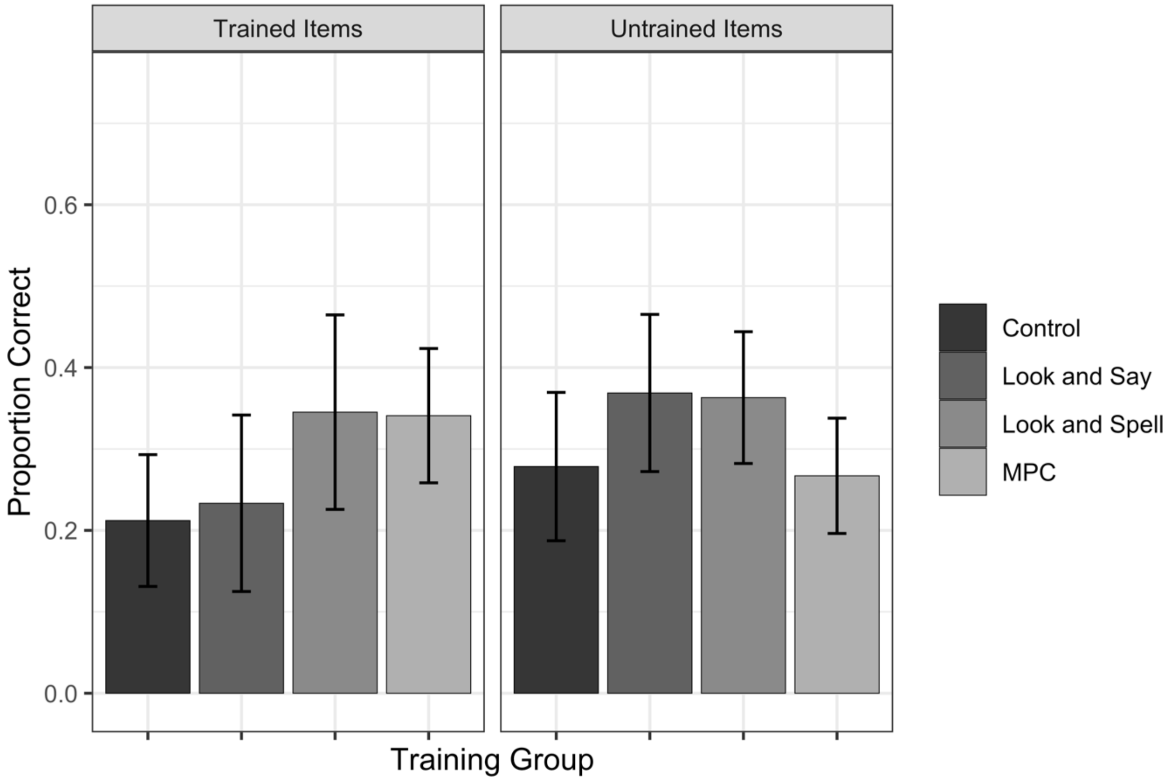


Table 1

Demographic information and background assessment results

	Controls	LSay	LSpell	MPC
N	22	20	21	22
Number female[1]	13	11	8	12
Language other than English[2]	1	1	2	1
Participants with a diagnosis[3]	2	0	2	5
Mean training sessions attended	NA	2.90 (0.31)	2.81 (0.40)	2.96 (0.21)
Mean LeST[4] percentile rank	53.06 (28.69)	48.44 (27.78)	53.58 (38.28)	48.76 (29.26)
Mean pre-test word reading score	5.64 (3.54)	5.20 (3.95)	5.90 (4.70)	5.27 (3.72)
Mean vocabulary score	14.95 (2.24)	15.95 (2.28)	14.48 (2.94)	15.64 (2.38)

Note. Standard deviations are in parentheses. Differences between groups were not statistically significant. LSay = Look and Say. LSpell = Look and Spell. MPC = Mispronunciation Correction. [1]All children identified as male or female. [2]Number of participants speaking a language other than English at home [3]Number of participants with a diagnosis of speech or language difficulties, autism spectrum disorder (ASD), and/or attention deficit hyperactivity disorder (ADHD). [4]Letter Sound Knowledge Test (Larsen et al., 2015)

Table 2

Trained and untrained words

Word Type	Word	Letters	Frequency[1]	Orthographic Neighbours Adult data[2]	Orthographic Neighbours Child data[3]	Irregular GPCs[4]	Syllables
Trained	Stomach	7	4.6	1	0	2	2
	Sugar	5	4.78	1	NA	3	2
	Jealous	7	4.42	1	0	2	2
	Uniform	7	4.12	1	0	1	3
	Heart	5	5.09	9	2	1	1
	Whistle	7	4.29	5	1	2	2
	Tongue	6	4.78	1	0	1	1
	Country	7	5.18	1	0	1	2
	Colour	6	5.12	2	0	2	2
	Listen	6	5.47	3	0	2	2
	Young	5	5.37	0	0	1	1
	Monkey	6	4.9	3	1	2	2
	Mean	6.17	4.84	2.33	0.36	1.67	1.83
Untrained	Sausage	7	5.37	1	0	2	2
	Human	5	5.07	2	0	2	2
	Famous	6	4.86	0	0	2	2
	Diamond	7	4.51	1	0	1	3
	Watch	5	5.55	8	6	1	1
	Spinach	7	4.48	0	0	2	2
	Orange	6	4.96	3	0	2	2
	Answer	6	5.46	1	0	2	2
	Mean	6.13	5.03	2.00	0.75	1.75	2.00
	p	0.91	0.34	0.78	0.63	0.74	0.52

Note. [1]CBBC frequencies from Subtlex UK (van Heuven, Mandera, Keuleers, & Brysbaert, 2014) [2]From CLEARPOND (Marian, Bartolotti, Chabal, & Shook, 2012) [3]From the Children's Printed Word Database (Masterson, Stuart, Dixon, & Lovejoy, 2010) [4]Regularity of a GPC was determined using a list of GPCs extracted from the CELEX database (Baayen, Piepenbrock & Gulikers, 1995).

Table 3
Responses to trained items on the orthographic choice task

	Proportion of Responses (%)			
Response Type	Control Group	Look and Say	Look and Spell	MPC
Correct	21.54	23.63	34.80	34.48
Regularized	56.92	56.96	40.00	41.38
One letter different	21.54	19.41	25.20	24.14

Note. MPC = Mispronunciation correction.