Too harts, won sole: Using Dysgraphia treatment to address homophone representation

Polly Barr[1,2], Britta Biedermann[3,2], Marie-Joseph Tainturier[4], Saskia Kohnen[2,5], Lyndsey Nickels[2,5]

[1] *The Brain and Mind Centre, University of Sydney, Sydney, Australia*

[2] *ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, Australia*

[3] *School of Occupational Therapy, Social Work and Speech Pathology, Curtin University, Perth, Australia*

[4] *School of Psychology, Bangor University, Bangor, Wales*

[5] *Department of Cognitive Science, Macquarie University, Sydney, Australia*

Corresponding Author:

Dr Polly Barr, *The Brain and Mind Centre, University of Sydney, Sydney, NSW 2000 Australia*

Email: polly.barr@sydney.edu.au

Word count: 7825

Abstract word count: 195

Abstract

Previous spoken homophone treatment in aphasia found generalisation to untreated homophones and interpreted this as evidence for shared phonological word form representations. Previous written treatment of non-homophones has attributed generalisation to orthographic neighbours of treated items to feedback from graphemes to similarly spelled orthographic word forms. This feedback mechanism offers an alternative explanation for generalisation found in treatment of spoken homophones.

The aim of this study was to investigate the mechanism underpinning generalisation (if any) from treatment of written homophones. To investigate this question a participant with acquired dysgraphia and impaired access to orthographic output representations undertook written spelling treatment. Generalisation to untreated items with varying degrees of orthographic overlap was investigated. Three experimental sets included homographs (e.g. bank-bank), heterographs (e.g., sail-sale), and direct orthographic neighbours (e.g., bath-path). Treatment improved written picture naming of treated items. Generalisation was limited to direct neighbours. Further investigation of generalisation found that items with a greater number of close neighbours in the treated set showed greater generalisation. This suggests that feedback from graphemes to orthographic word forms is the driving force of generalisation. The lack of homograph generalisation suggests homographs do not share a representation in the orthographic lexicon.

Key words: Homophone, Language production, Spelling, Treatment, Dysgraphia

Introduction

There is general agreement that each word in our language requires a stored representation of meaning (semantics), and phonological/orthographic form (e.g., Caramazza, 1997; Dell, 1986; Levelt, Roelofs, & Meyer, 1999). However, there are circumstances when a word's form completely overlaps with that of another word (e.g., *cricket* the insect, and *cricket* the sport). These types of words (homophones) clearly have separate meanings and hence separate semantic representations but perhaps their form need not be separately represented? The nature of homophone representation has been investigated over the past 20 years using various different techniques (e.g., frequency effects on reaction times, neuropsychological treatment, error rates) yet it is still unclear how homophones are represented in the lexicon. This study explores the orthographic representation and written production of homophones, which treatment factors may influence generalisation and the nature of orthographic processing.

Representations in the lexicon

There are currently no models that focus solely on the orthographic representation of homophones. In contrast, there are three competing models of homophone representation in spoken production. These can therefore assist conceptualisation of written homophone production. Figure 1 depicts these three theories of spoken homophone production, whereby in (1a) homophones share a lexical representation as in, for example, Levelt et al.'s Two Stage model (1999; Figure 1a). According to (1b) homophones have independent representations with strictly feed forward activation (Figure 1b; Caramazza, Costa, Miozzo, & Bi, 2001), while in (1c) homophones have independent representations with interactive activation between the lexical representations and segment level (Figure 1c; Middleton, Chen & Verkuilen's 2015 adaptation of Schwartz, Dell, Martin, Gahl, & Sobel, 2006). Due to similarities between spoken and written production (Damian, Dorjee, &

Stadthagen-Gonzalez, 2011) it seems plausible that phonological and orthographic architecture are similar. Therefore, we can assume that the orthographic representation of homophones is one of the figures in Figure 1, however the lexical representation would be orthographic (not phonological) and the segment level would be graphemes (not phonemes). Previous psycholinguistic studies have offered competing evidence for the different types of models shown in Figure 1 (e.g., Caramazza et al., 2001; Dell, 1990; Jescheniak & Levelt, 1994), however here we focus on neuropsychological evidence.

Figure 1 about here

Treatment studies of spoken homophone production

One method of investigating homophone representation and the nature of orthographic processing involves exploring generalisation of the effects of language therapy with people with aphasia. Previous studies have found that treating one homophone (e.g., *knight*) improves spoken naming of the untreated homophone partner (e.g., *night*). Following Blanken's (1989) landmark study, Biedermann and colleagues replicated the pattern of homophone generalisation in both German (one person) and English people with aphasia (two people) with lexical-access impairments. Biedermann and Nickels (2008b) found this effect for both homographic homophones (e.g., *cricket* [insect]/*cricket* [sport]) and heterographic homophones (e.g., *knight/night*), with no difference in the extent of generalisation explained by orthographic similarity. As there was no generalisation to phonologically related items, the authors interpreted generalisation as being due to *shared* representations at the phonological word form level (support for Figure 1a).

Biran et al. (2013) replicated this work in Hebrew using a phonological cueing therapy with two people with aphasia who also had word retrieval deficits. Both participants showed improvement in naming treated homographic homophones (e.g., *mapa*; tablecloth) and their untreated homophone partners (e.g., *mapa*; map), but not phonologically related controls (e.g., *maca;* Matzah).

In sum, to date, every phonological treatment study that has addressed spoken homophone production has found generalisation of treatment to untreated homophones. Furthermore, the effects have been interpreted as evidence for shared representations at the phonological word form (Biedermann et al., 2002; Biedermann & Nickels, 2008a, 2008b; Biran et al., 2013). These findings support Levelt et al.'s (1999) Two-Stage Model of spoken language production (Figure 1a) where homophones share a phonological word form.

However, an alternative explanation for effects described above, is that homophone generalisation is caused by feedback from phonemes to independent representations (as in Figure 1c). Treatment of one word form would result in increased activation of the target's phonemes. Activation from these phonemes could feedback and activate other word forms that also contain these phonemes, therefore the word form with the greatest amount of phonological overlap would show the most generalisation, with homophones showing greatest generalisation given their 100% overlap. However, Biedermann and Nickels (2008b) rejected this feedback account as they found no evidence for differing degrees of generalisation to phonologically related non-homophones dependent on number of shared phonemes (indeed, there was no generalisation to phonologically similar items regardless of their phonological similarity; see also Biedermann et al., 2018).

The architecture shown in Figure 1b does not predict generalisation to untreated items, and therefore it cannot explain the homophone-specific generalisation found in the phonological lexicon (Biedermann et al., 2002; Biedermann & Nickels, 2008a, 2008b; Biran et al., 2013). Consequently, the major debate that remains is as to whether homophone generalisation is caused by a shared word form (Figure 1a) or by feedback to independent word form representations (Figure 1c). This disentangling of homophone representation is one motivation for studying generalisation in the orthographic lexicon.

Treatment studies of written production

It is well established that there are separate and independent (but related) orthographic and phonological output lexicons (e.g., Shelton and Weinrich, 1997). However, as noted above, most research investigating the production of homophones has involved the spoken modality- indeed few models address the representation of written homophones at all (but see Caramazza, 1997). Therefore, in order to gain a comprehensive understanding of homophone representation in the language system, it is critical to investigate whether the orthographic lexicon is structured similarly to the phonological lexicon. If homophones share a phonological word form, do they also share an orthographic word form? To our knowledge, there has been only one previous study investigating orthographic treatment of written homophone naming. Behrmann (1987) conducted a treatment study with a participant (CCM) who had impaired access to the orthographic output lexicon, resulting in poor irregular spelling with phonologically plausible errors and homophone confusions. Before treatment, CCM scored 49% on homophone spelling to dictation (with a disambiguating sentence) and a large proportion (57%) of errors were the homophone partner (e.g. writing ‘sail for ‘sale’). It was hypothesised that CCM had difficulty retrieving the homophone spelling on the basis of semantic information (accessing the orthographic output lexicon from semantics). CCM was treated using a series of tasks that involved contrasting the different spellings of homophones, using pictures and written sentence completion tasks (e.g., she was shown a picture of a knight, with the written word *knight*, which was orthographically contrasted with the word *night*, and then completed a written sentence requiring the spelling of the target *knight*). The treatment improved homophone spelling of the treated homophones and this generalised to better spelling of untreated irregular words but not homophone partners. Behrmann (1987) suggested this was due to improved orthographic lexical access (for the treated homophones) as well as an improved visual-checking mechanism (generalisation to irregular words). This was supported by the pattern of responses to untreated homophone partners: the percentage of non-word errors was reduced (from 36% to 18%) and a larger percentage of errors were the (treated) homophone partner (from 57% to 82%).

This is contradictory to previous (spoken) homophone treatment findings that found homophone generalisation (e.g. Biedermann et al., 2002; Biedermann & Nickels, 2008a, 2008b; Biran et al., 2013). However, Behrmann's (1987) written treatment investigated written orthographic production of items that were orthographically different (i.e., heterographic homophones, e.g., flour/flower). In contrast, the spoken homophone treatments focused on phonological generalisation to items shared phonology (regardless of orthography). It remains to be investigated whether treatment of words that share orthography and phonology (homographic homophones), results in generalisation in written naming. Further, some methodological weakness in Behrmann's (1987) treatment design could also have led to reduced homophone generalisation. In particular, items selected for treatment were spelled incorrectly at baseline whereas the majority of correctly spelled items were assigned to the untreated group (30/50). Therefore, the untreated items had less room to improve to show generalisation (and the treatment effects may have also been magnified by ‘regression to the mean’, Howard, Best & Nickels, 2015).

There are other instances of differing patterns of generalisation across phonological and orthographic treatment modalities. For example, while there has been a lack of generalisation to phonological neighbours (e.g. cricket-ticket) in some spoken treatment (e.g., Biedermann & Nickels, 2008b), generalisation to orthographic neighbours (e.g. clock-block) has been found in some written treatment (e.g., Harris et al., 2012).

Harris et al. (2012) and Sage and Ellis (2006) both conducted treatment studies with individuals with aphasia who presented with a graphemic buffer impairment (the level at which activation of graphemes is maintained before processing for output). Both studies found that treatment generalised to untreated non-homophonic direct neighbours (pairs of words with one grapheme different e.g. *clock-block*). The authors suggested this generalisation could have been caused by the treatment increasing activation of the orthographic representations of the treated items (e.g. clock), which in turn activated the graphemes stored in the graphemic buffer (e.g., c-l-o-c-k). These activated graphemes would have fed activation back to the word forms that shared these letters (e.g., the graphemes l-o-c-k- would feed activation back to ‘block’) resulting in subsequent improved production of both the target and neighbours. This suggests an orthographic network similar to the one shown in Figure 1c. However, Krajenbrink, Nickels, and Kohnen (2017) failed to replicate this direct neighbour generalisation in two similar cases with a graphemic buffer profile of impairment. One possible explanation was that, in these individuals, perhaps there was reduced or absent feedback from the graphemic buffer to the lexicon (Krajenbrink et al., 2017).

Generalisation from treated to untreated spelling has also been found in cases of developmental dysgraphia with no graphemic buffer profile impairment (e.g., Brunsdon, Coltheart, & Nickels, 2005; Kohnen, Nickels, Coltheart, & Brunsdon, 2008). Kohnen et al. (2008) found that this generalisation was dependent on orthographic neighbourhood. When a word is spelled similarly to many of other words it has a large orthographic neighbourhood (e.g., line has 22 ‘neighbours’, lime, lane, lint, pine etc), whereas a word has a small orthographic neighbourhood if very few words differ in spelling by one letter (e.g., skull has only one neighbour; skill). Similar to the studies with acquired dysgraphia, Kohnen et al. (2008) suggested that untreated words with large orthographic neighbourhoods are more likely to improve due to increased feedback from graphemes to orthographic word forms that are repeatedly activated due to the large number of graphemes shared with other words, including, most likely, the treated words.

This feedback mechanism, as the cause of generalisation to direct neighbours in written production, is analogous to the alternative (feedback) explanation for generalisation to untreated homophones in spoken production mentioned earlier (Biedermann & Nickels, 2008b, 2008a). However, if it were the case that generalisation in the spoken modality was due to feedback, there should have also been generalisation to phonological neighbours (which was not found), as was found to orthographic neighbours (e.g., Harris et al., 2012).

In sum, it is still unclear whether the effects of generalisation found for phonological homophone treatment are due to feedback or shared-word forms, nor is the nature of the orthographic representation of homophones clear. Therefore, we carried out a treatment study that aimed to further investigate whether any generalisation in orthographic homophone treatment is due to improved shared lexical entries, or feedback to separate entries and how this differs across heterographic and homographic homophone spelling. We thus treated three groups of stimuli (homographs, heterographs, non-homophonic controls) and investigated generalisation to five untreated groups (homograph partners, heterograph partners, direct neighbours of the controls, unrelated high orthographic neighbourhood words and unrelated low orthographic words).

If the mechanism for generalisation is feedback from graphemes, then we would expect to find generalisation to homographs as well as (less) generalisation to items that have high (but not 100%) orthographic overlap (heterographs and direct neighbours), as in Figure 1c. If we find generalisation only for homographic homophones, this will suggest they share a word form representation, and there is no generalisation due to feedback from graphemes as in Figure 1a. If no generalisation is found, this is consistent with homographs having separate representations with strictly feed forward activation as in Figure 1b.

Case History

The participant in this study, CWS, was a 67-year-old right-handed, high school educated, former builder from North Wales. He learnt both Welsh and English before the age of six and still used both regularly. He reported that both pre- and post-stroke, he was equally proficient in English and Welsh (i.e., he had no 'dominant' language), however, as this treatment investigates English, only his English naming performance is reported (see Roberts, 2013, for a comprehensive report of his bilingual language abilities). CWS suffered a right frontal infarct in 1997 (18 years prior to this experiment). This resulted in left-sided hemiplegia and crossed-aphasia (aphasia due to right hemisphere damage despite right-handedness) resulting in agrammatic, non-fluent speech. Table 1 shows CWS's language performance on a range of standardised (on English monolinguals) tests. CWS's spoken and written comprehension remains intact along with word and non-word repetition. CWS performed within the control range for spoken object and action naming (Druks & Masterson, 2000). His visual word recognition was just below ceiling and within control performance range. Regular and irregular reading aloud were intact, although he showed severely impaired non-word reading, a symptom pattern that is consistent with phonological dyslexia (see Tainturier, Roberts, & Leek, 2011, for detailed analysis of his reading).

Table 1 about here

Table 2 about here

CWS's spelling performance is reported in further detail in Table 2. Although CWS was impaired in all aspects of spelling, he had a significantly larger deficit when spelling irregular words compared to regular words and non-words (Chi squared, $X^2(1, N = 80) = 12.99$, $p<.001$, $X^2(1, N=80)= 10.90$, $p=.001$). During written picture naming tasks (e.g. PALPA 53), despite being asked not to name or spell aloud, CWS would attempt to spell by breaking a word down into phonemes (which he could do accurately) and spelling one phoneme at a time. This pattern of performance suggests attempted use of a sub-lexical strategy secondary to damage to the orthographic output lexicon or access to this lexicon. This strategy results in better performance on regular word spelling compared to irregular word spelling and phonologically plausible (regularisation) errors. Table 3 shows examples of errors taken from Roberts (2013), indeed 39% of CWS's spelling errors were phonologically plausible. However, CWS also produced a large number of phonologically implausible errors (33%). To summarise, CWS presented with a mixed dysgraphia profile. He had a clear lexical impairment, however the presence of a length effect and impaired nonword spelling also suggests possible additional graphemic buffer and/or sub-lexical impairments (for a detailed analysis of this see Roberts; 2013).

Table 3 about here

CWS's orthographic word form level impairment made him a suitable candidate to investigate generalisation of homophone treatment at the orthographic word form level.

Intervention Study

A copy and recall treatment (CART) (e.g. Beeson, 1999) in the presence of the picture was conducted. This task was chosen to ensure that the treatment did not improve spelling via a focus on phoneme to grapheme conversion but instead increased the accessibility of the orthographic representations from the semantic system (as this was the representation we were investigating). Previous treatment studies have found CART to be an effective strategy to strengthen orthographic representations (Beeson, Hirsch, & Rewega, 2002).

Method

Stimuli.

Stimuli were picturable nouns presented as photographs 300 x 300 pixels in size, displayed in the centre of a computer screen with written descriptions underneath. Descriptions were designed to clarify picture identity to facilitate written naming. They were not ‘stand-alone’ descriptions and did not contain any semantic competitors of the target (e.g. the definition used with knife was ‘used for eating'). All stimuli (picture with the description) had over 70% name agreement when named by 10 control participants (mean age 29.20 years).

Stimuli belonged to one of eight experimental subsets: 1) Homographic homophones (e.g., *cricket* [insect]), (2) homographic partners of 1 (e.g., *cricket* [sports]), (3) heterographic homophones (e.g., *sale*), (4) heterographic partners of 3 (e.g., *sail*), (5) non-homophonic controls with direct neighbours (e.g., *bath*), (6) direct orthographic neighbours of 5 (e.g., *path*), (7) non-homophonic control words with high orthographic neighbourhoods (e.g., *line*) (8), and non-homophonic control words with very low orthographic neighbourhoods (e.g., *church*). The direct neighbours in subsets 5 and 6 consisted of words with one grapheme substituted (they differed by one grapheme in the same position e.g., *cake-cave*; Coltheart, Davelaar, Jonasson, & Besner, 1977). They did not include additions or subtractions. Homophone and words with direct neighbours (subsets 1-6) were randomly assigned to two sets (per condition i.e. 1 or 2, 3 or 4, 5 or 6). These sets were then adjusted to ensure matching on the variables presented in Table 4 before randomly being assigned to treated or untreated conditions. Values for all the psycholinguistic variables except frequency and regularity were obtained from N-watch (Davis, 2005). CELEX (Baayen, Piepenbrock, & Gulikers, 1996) was used to obtain frequency counts per million, then log10 transformations were performed. An item was listed as irregular if it had at least one grapheme that was classed as not the most frequent use by Perry, Ziegler and Coltheart (2002) or rare by Fry (2004).

Table 4 about here

The treated and untreated subsets were matched for accuracy across the three baseline sessions (see Table 4). The treated subsets were matched with their paired untreated subset (i.e., subsets, 1 with 2, 3 with 4, 5 with 6) for all the psycholinguistic variables shown in Table 4 (paired-sample t-tests, *p*=> .05; e.g., subset 1 had the same frequency as subset 2). The untreated orthographic neighbourhood sets (subsets 7, untreated high orthographic neighbourhood, and 8, untreated low orthographic neighbourhood) were also matched to subset 5 (treated non-homophone, and therefore also 6, untreated direct neighbours of subset 5) on all variables except the orthographic neighbourhood variables for subset 8. As orthographic neighbourhood was manipulated in subsets 7 and 8, the untreated low orthographic neighbourhood subset (subset 8), was significantly different to subset 7 (and therefore 5 and 6) for any variables associated with orthographic neighbourhood (regularity and phonological neighbourhood frequency). Subset 7 also had significantly lower phonological neighbourhood density. None of the subsets had any direct substitution neighbours in another subset apart from subset 5 that, by design, had exclusively direct neighbours in subset 6. It was impossible to match number of items (and variables) across pairs of sets, given the nature of the experimental items (i.e., the need for stimuli to be picturable homophones and picturable non-homophones with picturable direct neighbours).

Procedure.

All subsets were presented for picture naming over three pre-treatment baseline tests and three post-tests. Only stimuli from subsets 1, 3 and 5 were treated. As indicated in the timeline shown in Figure 2, both written and spoken word production were tested during pre-treatment baselines, once within the treatment phase and at post-tests. All assessments were two weeks apart, except for the final two post-tests. The experimenter read out the description as the picture was presented. In each session items were presented in a different randomised order.

Spoken naming familiarisation.

Two days prior to baseline testing, two sessions of spoken naming familiarisation were conducted, two days apart from each other. This was to ensure that CWS was familiar with the picture names and to rule out any incorrect written responses due to ambiguous pictures and to ensure that the phonology of each item was equally available. The spoken familiarization phase consisted of presentation of the stimulus picture and spoken name for CWS to repeat. The stimuli were split into two equal sets. During Session 1, Set 1 was presented first with the correct name for CWS to repeat. The same items were subsequently presented for CWS for uncued spoken naming, if he named an item correctly, he was given feedback (e.g., 'well done, that is correct'). If he produced the wrong name or no response, the correct name was given for repetition (he was always able to repeat the item correctly). Set 2 was then presented using the same procedure. Session 2 consisted of the same procedure as Session 1 with Set 2 items presented. Overall, in total, each item was repeated and named four times (twice in each session).

Assessment sessions.

Because of the large number of items, each assessment was split into two sessions. Each set contained 84 experimental items and 16 filler items (which were included for data collection for a separate study). During the first session of Baseline 1, CWS wrote the names of Set 1, and completed spoken naming of the pictures of Set 2. During the second session of Baseline 1 CWS was asked to write the names of Set 2 and complete spoken naming of Set 1. Whether Set 1 or Set 2 was given for written naming in the first or second session of an assessment alternated across time points. The two sessions assessing both modalities of one set were at least one day apart. As we were primarily interested in investigating the effect of treatment on written naming performance, and only secondarily in the effect of spelling treatment on spoken naming (which was close to ceiling), all of our baselines and post-tests assessed written naming before spoken naming. Each set was given in a different randomised order at each presentation time point.

Figure 2 about here

Treatment consisted of six sessions over two weeks before an interim assessment and then another six sessions over two weeks. Overall, CWS received 12 treatment sessions over four weeks. Each treatment session contained 60 experimental items (i.e. the homograph, heterograph and direct neighbour treatment sets) as well as 15 filler items and took roughly 1 and half hours. CWS also completed six homework sessions, one after every second treatment session. The first post-test occurred two days after the last treatment session. Due to CWS' unplanned hospitalisation, post-test 2 and 3 were conducted later than planned at three- and four-weeks post treatment. Consequently, analysis was conducted only including one post-test, although all three post-tests are depicted in the figures below.

Treatment.

In order to promote lexical (rather than sublexical) spelling, treatment was based on the copy and recall (CART) approach from Beeson (1999). CWS was presented with a stimulus picture and correct spelling of the target name and asked to copy the word while the word stayed in sight. This immediate copying was excellent, and this stage of the treatment was error-free. The experimenter then covered both the presented correct spelling and CWS's immediate copy and counted to ten aloud. After this ten second delay CWS was asked to 'try and spell the name of the picture again'. The correct spelling was then presented, and CWS was asked to judge whether he had correctly spelled the target item. If he (correctly) confirmed that his delayed copy was correctly spelled, feedback was given, and the next item was presented. If he misspelled the item and realised this was the case, he was asked what part he thought was wrong. Then the target word was presented and contrasted to his incorrect spelling, before he copied it once again. On the very few occasions that he incorrectly judged his spelling as correct, the experimenter presented the target item and pointed out the contrast, before he copied the correct target. On the rare occasion he was unable to produce anything from delayed copying, the target was presented again to copy. Therefore, each item was written correctly twice in each treatment session. Homework also consisted of immediate copying and delayed recall. CWS was given a booklet with one picture per page. On one page the written word was presented below the stimulus picture for copying. This would be followed by another stimulus item for copying, and then the first picture presented for delayed recall after this intervening item, followed by the second item for delayed recall. As the homework was conducted independently no feedback was given.

Analysis.

We examined the effects of treatment on performance using Weighted Statistics (Howard, Best, & Nickels, 2015). We initially established if there was overall improvement over the course of the study by conducting a trend (WEST-Trend) analysis. If there was significant improvement, then we also conducted a Rate of Change (WEST-ROC) analysis to investigate if this trend could be attributed to improvement during the treatment phase. Only if both these tests showed significant positive change did we conclude there was treatment-specific improvement (Howard et al., 2015). For written naming, we analysed both whole word accuracy and letter accuracy using letter accuracy scoring adapted from Buchwald and Rapp (2009). Each letter attracted a score of one if it was correctly produced in the correct position. Between 0.25-1 points were deducted for each letter position that was either transposed, migrated, substituted, missed or included an additional letter[1]. We present the average letter accuracy in the results (following e.g. Buchwald & Rapp, 2009; Krajenbrink et al., 2017 ; Sage & Ellis, 2006), however we also conducted analysis on the sum of the letter scores (see Appendix A; Figure A1 and Table A1). We were concerned that the average letter analysis would diminish improvement in longer words. For example, one more letter correct in a three-letter word represents 33% improvement, but improvement by one letter in a four-letter word results in 25% increase. However, as shown in Appendix A, there was no difference in the results between the analysis using the average and that using the sum letter scoring.

Results

We only report the spelling analysis below, as our primary interest is on written naming. The results and analysis of spoken naming are presented in Appendix B (Figure B1 and Table B1). In brief however, spoken naming was close to ceiling (as was intended by the familiarisation phase), and there was no statistically significant treatment-related change.

Due to the unplanned extended period between the last treatment session and post-tests two and three, analysis with only one post-test is reported below (see Table 5). However, analysis including all three post-tests is presented in Appendix C. All the effects were in the same direction in both analyses; however, improvement was not maintained for all items (as shown in Figures 3-6).

---

[1] For example, when scoring KNIFE spelled as *neafh* the E has migrated from fifth position to the second, so it was scored 0.5. When dealing with multiple errors in a single response, following Buchwald and Rapp (2009) and Krajenbrink et al., (2017), the 'visible' transformation is scored:, we did not assume migration of substituted letters, nor penalise the same position twice. For example, spelling WORD as *whod*, was scored by penalising the addition of H between W and O (i.e. 0.5 score for W and the O) and deletion of R (0 points for R). We would NOT, for example, assume that O and R transpose [(0.75 points each), and the R is also replaced with H (reducing the 0.75 score to 0) as there is no way of knowing this. Instead only what is seen (that O has moved and there is an additional H) is scored.

Figure 3 about here

There was significant improvement due to treatment both for word and letter accuracy for the treated homographs (Figures 3a & b; and Table 5). However, there was no treatment-related improvement for the untreated homographs, for either word or letter accuracy analyses.

Figure 4 about here

The same pattern was found for the heterographs; significant improvement for the treated items (both whole word, and letter analyses; Figure 4 and Table 5) but no generalisation to the untreated items.

Figure 5 about here

The treated non-homophonic controls also improved due to treatment in both whole word and letter analyses (as shown in Figure 5 and Table 5). In the whole word analysis, the untreated direct neighbours of controls did show significant improvement, but this was not replicated across the letter level analysis.

Figure 6 about here

There was no treatment-related improvement for whole word or letter accuracy for either of the untreated orthographic neighbourhood control sets (Figure 6 and Table 5).

Table 5 about here

Discussion

The aim of this study was to investigate whether homophones have shared or separate representations in the orthographic output lexicon by investigating generalisation of treatment-related improvement to items with differing degrees of orthographic overlap. We found that written naming of treated items improved in both whole word and letter analyses and there was some generalisation to untreated items (in the whole word analysis). If the effect of treatment was due to improved knowledge or application of phoneme-grapheme-conversion rules or improved functioning of the graphemic buffer, generalisation to all items that were not treated would have been expected. This did not occur, ruling out these non-lexical mechanisms underlying treatment-related improvement.

One means by which treatment can improve written naming is through strengthening of orthographic representations (e.g. Rapp & Kane, 2002). This strengthening makes these representations more accessible following treatment (Krajenbrink et al., 2017). This predicts improvement of any orthographic representation that was activated during treatment. An alternative account of treatment-related improvement proposes that prior activation (i.e. during treatment) can improve access to the lexical representation by strengthening the links to these representations from semantics (e.g., Howard, Hickin, Redmond, Clark, & Best, 2006; Howard, Nickels, Coltheart & Cole-Virtue, 2006; Wheeldon & Monsell 1992). We have no means of distinguishing between these two accounts as the mechanism for improvement in treated items for CWS, or indeed whether both mechanisms are at play. However, some insights may be gained from the pattern of generalisation (see later).

The only untreated items that showed improvement were direct orthographic neighbours of treated items (which were non-homophones), and this was restricted to the whole word analysis. Sage and Ellis (2006) suggest that generalisation to direct orthographic neighbours originates from treatment increasing activation of the treated word form (e.g., *bath*) which activates the letters (e.g., *B, A, T, H*). During treatment, letters that are shared with a neighbour (e.g., for the neighbour *path,* the letters *A, T, H*), feed their activation back to the orthographic word form of this neighbour (e.g., *path*). This in turn activates the neighbour’s word-form representation and improves subsequent access to this item.

However, if feedback from shared letters was the mechanism underpinning generalisation to direct neighbours in the present study, then treating homographs should also result in generalisation to their untreated homophone partners (as homographs share 100% of letters with their treated partners). This was not the case. Consequently, in the next section we present the results of supplementary analyses that aimed to explore potential reasons for this apparent contradiction: why might feedback result in generalisation to items that do not overlap by 100% of their letters but not to items that do?

Further Analyses: Exploring generalisation to untreated items

In common with other similar investigations (e.g. Sage & Ellis, 2006, Harris et al., 2012), we conceptualised orthographic overlap as overlap between matched pairs of items, then subsequently predicted the extent of expected generalisation based on this measure. However, any untreated item will not just receive feedback from its treated 'mate' but also from any treated item with which it has orthographic overlap. In other words, the amount of priming for an untreated lexical item following treatment relates to the number of times that item is activated over the treatment set as a whole, not just from its experimentally assigned partner. Hence, *coal* will not only be activated (and primed) when its experimentally assigned direct neighbour partner *coat* is treated but also when *cone, cow* and *court* are treated (but presumably to a lesser degree).

Consequently, to further understand what was driving generalisation in this study, we carried out additional analyses examining whether there was an influence of the similarity to other items that were treated (number of treated neighbours).

Method

To investigate the effect of similarity to all treated items on generalisation we investigated two measures of orthographic overlap. 'Orthographic Overlap' was the average of the orthographic similarity between each untreated word and all the items in the treated set using the ‘ends-first’ spatial coding system within the match calculator application (Davis, 2007). For a more detailed explanation of the coding and algorithms used see Davis and Bowers (2006), however, in brief, the match calculator outputs a weighting for each untreated word compared to each treated word (one being an exact match e.g. bank-bank, and zero having no letters in common e.g. drain-shoe). The second measure, ‘Number of Treated Near Neighbours’ was the number of treated items that were zero, one and two Levenshtein distance neighbours of the untreated items[2] using the vwr package (Keuleers, 2013) in R. Levenshtein distance is the number of single-character steps (addition, subtraction or substitution) it takes to transform one word into another. Therefore, for the word *bank* (money)*, bank* (river) is a zero Levenshtein distance neighbour, whereas *band* is a one Levenshtein distance neighbour (substitute the *k* with *d*) and *bat* is a two Levenshtein distance neighbour (remove the *k* and substitute the *n* with *t*).

When comparing the untreated direct neighbour and untreated homograph sets, the direct neighbour set had significantly higher average Number of Treated Near Neighbours than the homograph set (two-sample t-test, t(44)=1.98, *p*=.05, 4.33 vs 3.045 respectively). However, this was not true for Orthographic Overlap (t(44)=-0.04, *p*=.96, 0.11 vs 0.11 respectively).

The dependent variables, the amount of treatment-related improvement, were the WEST-ROC coefficients for whole word accuracy and average letter accuracy across the word for every untreated item. Correlations were performed between these dependent variables and the two new variables (orthographic overlap, and number of treated near neighbours) and written lemma frequency, length (number of letters), regularity, number of orthographic neighbours, average orthographic neighbour frequency, number of

---

[2] We had controlled for the effect of orthographic neighbourhood by including a high and low orthographic control set. However, this did not consider orthographic neighbourhood *within* our set of treated items.

phonological neighbours and average phonological neighbourhood frequency. All continuous variables were centred. Any variables that significantly correlated with improvement (n=108, *p*≤ .05, *r* ≥0.2) were entered into a stepwise linear regression using the package gdata in R (Warnes et al., 2014) alongside any variables these (correlated) variables also correlated with. No items that were correlated above .80 were included within one regression model, to avoid potential multicollinearity (Hutcheson & Sofroniou, 1999).

Results

Word accuracy

On the basis of the correlations (see Appendix D) the first model for the whole word accuracy regression included the effect of the variables length, regularity, number of orthographic neighbours, number of phonological neighbours, orthographic overlap and number of treated near neighbours on treatment-related improvement for whole word accuracy (WEST-ROC; see Appendix E). Table 6 shows that after the backwards step-wise regression, the only significant variable to predict WEST-ROC word accuracy scores was the number of treated near neighbours. Figure 6 shows that as the number of treated near neighbours increased, so did the amount of treatment-related improvement indexed by the WEST-ROC coefficients.

Table 6 about here

Figure 7 about here

Letter scoring

As there were no significant correlations between improvement in letter accuracy and any variable (see Appendix D), regressions were not performed for the average letter scoring.

Discussion

The results of the regression analyses showed that, for word accuracy, only the number of near neighbours in the treated set predicted generalisation for untreated items. As direct neighbours had significantly more near neighbours than homographs in the treated set, this likely accounts for why only generalisation to direct neighbours was significant. These results suggest that when, during treatment, treated items activate their corresponding letters, these letters feedback activation to other items in the orthographic lexicon that share some of those letters. This results in the representations of those items also being strengthened and more accessible. This mechanism of feedback from shared letters leading to improved access to untreated orthographic representations is in line with the interpretations of generalisation in previous studies (e.g., Kohnen et al., 2008; Sage & Ellis, 2006). Moreover, this suggests that one mechanism underpinning improved spelling of treated items is also improved accessibility of lexical representations, although it is not possible to rule out additional treatment-related strengthening of semantic-lexical links.

General Discussion

This study examined homophone representation in the orthographic output lexicon; specifically, whether homophones share a word form representation. This was investigated by treating written picture naming using copy and recall treatment (in the presence of a picture) and investigating generalisation to untreated written picture naming of items with varying degrees of orthographic overlap. All treated items improved, however generalisation only occurred in whole word analysis to direct neighbours of the treated items and not to untreated homophone partners. Further analyses revealed that generalisation was determined by the degree of orthographic overlap (number of near neighbours) with whole treated set rather than simply the degree of orthographic overlap with the treatment partner. In other words, an item was originally categorised by its similarity to its experimentally assigned 'treatment partner' (e.g. homophones, *bank-bank* share all of the same letters whereas direct neighbours *car-ear* only share 2/3 letters). However, our results suggest that whether an untreated item will benefit from generalisation of treatment depends not so much on this similarity, but rather on how much orthographic overlap it has with all of the treated items regardless of experimental set (i.e. all 58 of the treated heterographs, homographs and controls). For example, the untreated item *chest* had 100% orthographic overlap with its experimentally assigned partner *chest* but had only 1 other item in the whole treated set (treated heterographs, homographs and controls combined) that was a 'near orthographic neighbour'. In contrast, although the untreated item *car* only had 66% overlap with its experimentally assigned treated partner *ear,* it had 11 near orthographic neighbours in the whole treated set.

What do the results mean for homophone representation?

We found no generalisation from improved spelling of treated homophones to their untreated partners. Instead, generalisation was dependent on how many near neighbours were in the treated set, suggesting feedback from graphemes that increased accessibility of word forms resulted in generalisation. This implies that homophones do not share an orthographic word form, as if they did, improved accessibility of the word form would predict both homophone partners (treated and untreated) should improve following treatment.

This is not in line with previous work investigating homophone representations in spoken production (Biedermann et al., 2002; Biedermann & Nickels, 2008a, 2008a; Biran et al., 2013; Jescheniak & Levelt, 1994) which concluded that there was a shared homophone representation in the phonological lexicon. However, there is some research to suggest that having one close neighbour (i.e. a homophone) can be detrimental to production, whereas larger numbers of more distant neighbours (as in this study) can be beneficial (Mirman, Kittredge, & Dell, 2010). It is unclear how this theory for spoken word production (using depth and width of phonological 'basins' as a method of word selection) would apply to generalisation of orthographic therapy. Nevertheless, it does imply that having several near neighbours can facilitate production more than having one identical word as a 'neighbour'.

Why would homophone retrieval generalise in the phonological but not the orthographic lexicon?

Previous treatment of homophones in spoken production has resulted in generalisation (Biedermann et al., 2002; Biedermann & Nickels, 2008a, 2008a; Biran et al., 2013), however, the present study did not find generalisation in written homophone production. This could be due to one of three reasons.

Firstly, perhaps the phonological and orthographic lexicons are not organised identically as we originally hypothesised: it is possible that (homographic) homophones share a representation in the phonological output lexicon but have separate representations in the orthographic output lexicon. At first it seems inconsistent to have different architectures for phonology and orthography, however, Best, Herbert, Hickin, Osborne and Howard (2002) point out that this seems more plausible when considering the considerable differences between acquiring spoken and written language (e.g. age, method and ease of acquisition). Nevertheless, there seems no logical reason as to why homographic homophones would share phonological but not orthographic representations, in contrast heterographic homophones require separate representations in the orthographic lexicon (but not in the phonological lexicon), due to their different spellings.

Secondly it is possible that the differences between written homophone production in the current study and spoken homophone production in the previous studies are in fact due to differences in the participants' responses to treatment. In order to fully test this hypothesis, both a phonological and orthographic treatment should be carried out with the same participant. If indeed homophones are represented as shared representations in the spoken modality, but independent representations in the written modality, then generalisation should occur for spoken naming treatment but not for written naming treatment. However, this design was not an option for CWS as his spoken naming was close to ceiling.

One crucial difference between CWS and the participants that undertook the phonological homophone treatment in Biedermann et al.'s (2002; Biedermann & Nickels, 2008b; 2008a) studies is the number of languages spoken by each participant. While the latter studies only included monolingual speakers with aphasia, CWS was an early bilingual, highly proficient in Welsh and English. Bilinguals are known to have smaller vocabulary sizes within each of their languages compared to monolingual speakers (e.g., Bialystok & Feng, 2009). Therefore, it is conceivable that a homophone representation is more likely to be a non-homophone for a late bilingual speaker who might not know both homophone meanings (e.g., due to their different frequencies), than a monolingual who has a larger vocabulary and is familiar with both word forms. However, this seems unlikely for CWS who was an early bilingual, growing up in a Welsh dominant household, but exposed frequently to English in the community from an early age and using both languages daily throughout his life. Nonetheless, it would be worthwhile investigating both orthographic and phonological homophone treatment within a monolingual participant to rule out that the possible lack of generalisation was not caused by some undetected non-native-like differences in the English lexicon[3].

Thirdly, it is possible that previous homophone generalisation in spoken production was due to feedback from phonemes to independent word representations (as in Figure 1c earlier). However, unlike the current study, previous spoken homophone generalisation studies may have included more near neighbours of homophones in the treated set overall, and relatively few for the phonologically related controls, hence, resulting in exclusively homophone generalisation. Currently, no study has investigated spoken homophone treatment generalisation as a function of overlapping phonemes with *all* treated items, but our research suggests that this is vital.

Unfortunately, we are unable to currently determine which of the above possibilities caused the difference in homophone generalisation depending on the production modality. However, our results suggest that homophones do not share a representation in the orthographic lexicon.

Why is there a difference between whole word and letter accuracy?

As we only found generalisation in the whole word analysis, it seems logical that predictors of generalisation were only found in the whole word correlations and regressions. However, letter analysis has often been suggested to be a more sensitive measure of change (e.g. Krajenbrink et al., 2017), so why was no significant generalisation found using letter scoring when there was with whole word scoring? It is possible that CWS was already spelling the majority of the letters correctly (compared to whole words) resulting in less

---

[3] Gvion, Biran, Sharabi, and Gil (2015) conducted a phonological homophone treatment with a bilingual participant, however, as this participant suffered from phonological output buffer impairment (not phonological word form impairment), homophone generalisation was not predicted. In fact, no treatment effects at all were found in this individual. Therefore, this particular case is uninformative in terms of homophone representations and whether being bilingual can influence homophone representation and generalisation.

room for improvement. Indeed, CWS's scoring on the letter analysis was higher (averaged over the three baselines was 64.07%) than the word analysis (16.67%) prior to treatment. Perhaps one letter improvement in a few items was enough to result in spelling more items 100% correctly, therefore improving whole word accuracy (to 37.50%- an improvement of 20.83 %), but this was not enough to significantly improve letter accuracy (to 71.74%; an improvement of only 9.08%). Indeed, of the 62% of the words CWS spelled incorrectly (across all baselines) had one letter wrong (i.e., the majority of errors were one letter incorrect) 23% of these were spelled correctly at post-test. This subset of words with one letter wrong, had an average word score of 0% correct and an average letter score of 74% correct before treatment. At post-test one 23% of these words are spelled correctly (23% improvement) with a letter accuracy score of 80% (6% improvement). This shows how a small letter improvement (6%) can result in a larger whole word improvement (23%) explaining the difference between whole word and letter analysis. Clearly, it is not always the case that letter scoring is a more sensitive measure of improvement. This also supports Konhen et al 's (2008) finding that items that are more accurate before treatment were most likely to show generalisation.

Clinical implications

This study was conducted first and foremost to provide insights into theories of language production (homophone representation and models of written language production). However, it also carries clinical implications. First, it replicates the previous literature demonstrating that the Copy and Recall (CART) method can, in the short term, improve spelling (e.g., Beeson, 1999; Beeson et al., 2002) and in this case with a participant with impaired access to the orthographic lexicon. Secondly, our findings support previous research that suggests that those items that are most likely to generalise are those from larger neighbourhoods (Brunsdon et al., 2005; Kohnen et al., 2008). However, while replication is required, our research suggests that in order to maximise generalisation, untreated items should be selected that have many neighbours within the treatment set

Conclusion

To conclude, this single case study shows that copy-and-recall-treatment can (at least transiently) improve spelling of treated items in a participant with impaired access to the orthographic lexicon. These effects generalised to items which have many near neighbours in the treated set. While replication is required across a case series, our results have implications for theories of orthographic processing and potential mechanisms underlying treatment effects. The lack of homophone generalisation and the fact that generalisation was predicted by number of near neighbours in the treatment set, suggests that the orthographic lexicon is organised with separate homophone representations, but that feedback from graphemes to word form representations that share graphemes is possible (as implemented in models with interactive activation, e.g., Dell, Lawler, Harris, & Gordon, 2004; Middleton et al., 2015, Figure 1c). If enough items are treated that share graphemes with untreated items, this can result in generalisation of treatment via feedback from these shared graphemes. More research is needed to determine if feedback also drives generalisation in spoken word production.

Acknowledgement

The authors would like to thank CWS for his participation in this research. During the preparation of this paper, LN was funded by an Australian Research Council Future Fellowship [FT120100102 (2012-2018) and PB was funded by Macquarie University's international HDR scholarship].

Declarations of interest

The authors have no potential conflict of interest.

References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). Celex2. *Linguistic Data Consortium, Philadelphia*.

Beeson, P. M. (1999). Treating acquired writing impairment: strengthening graphemic representations. *Aphasiology*, *13*(9–11), 767–785. https://doi.org/10.1080/026870399401867

Beeson, P. M., Hirsch, F. M., & Rewega, M. A. (2002). Successful single-word writing treatment: Experimental analyses of four cases. *Aphasiology*, *16*(4–6), 473–491. https://doi.org/10.1080/02687030244000167

Behrmann, M. (1987). The rites of righting writing: Homophone remediation in acquired dysgraphia. *Cognitive Neuropsychology*, *4*(3), 365–384. https://doi.org/10.1080/02643298708252044

Best, W., Herbert, R., Hickin, J., Osborne, F., & Howard, D. (2002). Phonological and orthographic facilitation of word-retrieval in aphasia: Immediate and delayed effects. *Aphasiology*, *16*(1–2), 151–168. https://doi.org/10.1080/02687040143000483

Bialystok, E., & Feng, X. (2009). Language proficiency and executive control in proactive interference: Evidence from monolingual and bilingual children and adults. *Brain and Language*, *109*(2–3), 93–100. https://doi.org/10.1016/j.bandl.2008.09.001

Biedermann, B., Blanken, G., & Nickels, L. (2002). The representation of homophones: Evidence from remediation. *Aphasiology*, *16*(10–11), 1115–1136. https://doi.org/10.1080/02687030244000545

Biedermann, B., Cholin, J., Jorschick, A., Croot, K., Hameau, S., & Nickels, L. (2018). Is the homophone advantage influenced by post-lexical effects? *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*. https://doi.org/10.1016/j.cortex.2018.01.014

Biedermann, B., & Nickels, L. (2008a). Homographic and heterographic homophones in speech production: Does orthography matter? *Cortex*, *44*(6), 683–697. https://doi.org/10.1016/j.cortex.2006.12.001

Biedermann, B., & Nickels, L. (2008b). The representation of homophones: More evidence from the remediation of anomia. *Cortex*, *44*(3), 276–293. https://doi.org/10.1016/j.cortex.2006.07.004

Biran, M., Gvion, A., Sharabi, L., & Gil, M. (2013). The Representation of Homophones in the Phonological Lexicon: Additional Evidence from Hebrew. *Procedia - Social and Behavioral Sciences*, *94*, 86–87. https://doi.org/10.1016/j.sbspro.2013.09.041

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

Caramazza, A. (1997). How Many Levels of Processing Are There in Lexical Access? *Cognitive Neuropsychology*, *14*(1), 177–208. https://doi.org/10.1080/026432997381664

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1430–1450. https://doi.org/10.1037/0278-7393.27.6.1430

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI*. Lawrence Erlbaum Associates.

Davis, C. (2007). Match calculator. Retrieved 18 April 2018, from http://www.pc.rhul.ac.uk/staff/c.davis/utilities/matchcalc/index.htm

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*(1), 65–70. https://doi.org/10.3758/BF03206399

Davis, C. J., & Bowers, J. S. (2006). Contrasting five different theories of letter position coding: Evidence from orthographic similarity effects. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 535–557. https://doi.org/10.1037/0096-1523.32.3.535

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S. (1990). Effects of Frequency and Vocabulary Type on Phonological Speech Errors. *Language and Cognitive Processes*, *5*(4), 313–349. https://doi.org/10.1080/01690969008407066

Dell, G. S., Lawler, E. N., Harris, H. D., & Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*, *21*(2–4), 125–145. https://doi.org/10.1080/02643290342000320

Druks, J., & Masterson, J. (2000). An action and object naming battery. Hove: Psychology Press.

Fry, E. (2004). Phonics: A Large Phoneme - Grapheme Frequency Count Revised. *Journal of Literacy Research*, *36*(1), 85–98. https://doi.org/10.1207/s15548430jlr3601_5

Gvion, A., Biran, M., Sharabi, L., & Gil, M. (2015). *Learning from naming treatment on the phonological representation of homophones*. Presented at the American Speech Hearing Association (ASHA) convention, Colorado, USA.

Harris, L., Olson, A., & Humphreys, G. (2012). Rehabilitation of spelling in a participant with a graphemic buffer impairment: The role of orthographic neighbourhood in remediating the serial position effect. *Neuropsychological Rehabilitation*, *22*(6), 890–919. https://doi.org/10.1080/09602011.2012.709872

Howard, D. (2000). Cognitive neuropsychology and aphasia therapy. In *Acquired Neurogenic Communication Disorders: A Clinical Perspective*. Retrieved from http://www.wiley.com/WileyCDA/WileyTitle/productCd-1861561113.html

Howard, D., Best, W., & Nickels, L. (2015). Optimising the design of intervention studies: critiques and ways forward. *Aphasiology*, *29*(5), 526–562. https://doi.org/10.1080/02687038.2014.985884

Howard, D., Hickin, J., Redmond, T., Clark, P., & Best, W. (2006). Re-Visiting “Semantic Facilitation” of Word Retrieval for People with Aphasia: Facilitation Yes But Semantic No. *Cortex*, *42*(6), 946–962. https://doi.org/10.1016/S0010-9452(08)70439-8

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Sage.

Jescheniak, J. D., & Levelt, M., J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843. https://doi.org/10.1037/0278-7393.20.4.824

Jescheniak, J. D., Meyer, A. S., & Levelt, W. J. M. (2003). Specific-word frequency is not all that counts in speech production: Comments on Caramazza, Costa, et al. (2001) and new experimental data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 432–438. https://doi.org/10.1037/0278-7393.29.3.432

Keuleers, E. (2013). *vwr: Useful functions for visual word recognition research. R package version 0.3. 0*.

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

Krajenbrink, T., Nickels, L., & Kohnen, S. (2017). No effect of orthographic neighbourhood in treatment with two cases of acquired dysgraphia. *Aphasiology*, *31*(5), 595–628. https://doi.org/10.1080/02687038.2016.1218435

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1–38.

Middleton, E. L., Chen, Q., & Verkuilen, J. (2015). Friends and foes in the lexicon: Homophone naming in aphasia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 77–94. https://doi.org/10.1037/a0037778

Mirman, D., Kittredge, A., K., & Dell, G. S. (2010). Effects of near and distant phonological neighbors on picture naming. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Perry, C., Ziegler, J. C., & Coltheart, M. (2002). A dissociation between orthographic awareness and spelling production. *Applied Psycholinguistics*, *23*(1), 43–73. https://doi.org/10.1017/S0142716402000036

Roberts, J. (2013). *Cross-Lingisitic treatment generalisation in Welsh-English Bilingual Aphasia* (Unpublished doctoral dissertation). Bangor University, UK.

Sage, K., & Ellis, A. (2006). Using orthographic neighbours to treat a case of graphemic buffer disorder. *Aphasiology*, *20*(9), 851–870. https://doi.org/10.1080/02687030600738945

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, *54*(2), 228–264. https://doi.org/10.1016/j.jml.2005.10.001

Shelton, J., R., & Weinrich, M. (1997). Further Evidence of a Dissociation between Output Phonological and Orthographic Lexicons: A Case Study. *Cognitive Neuropsychology*, *14*(1), 105–129. https://doi.org/10.1080/026432997381637

Tainturier, M.-J., Roberts, J., & Leek, E. C. (2011). Do reading processes differ in transparent versus opaque orthographies? A study of acquired dyslexia in Welsh/English bilinguals. *Cognitive Neuropsychology*, *28*(8), 546–563. https://doi.org/10.1080/02643294.2012.698986

Warnes, G. R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., … Rogers, J. (2014). gdata: Various R programming tools for data manipulation. *R Package Version*, *2*(3), 35.

Appendices

Appendix A

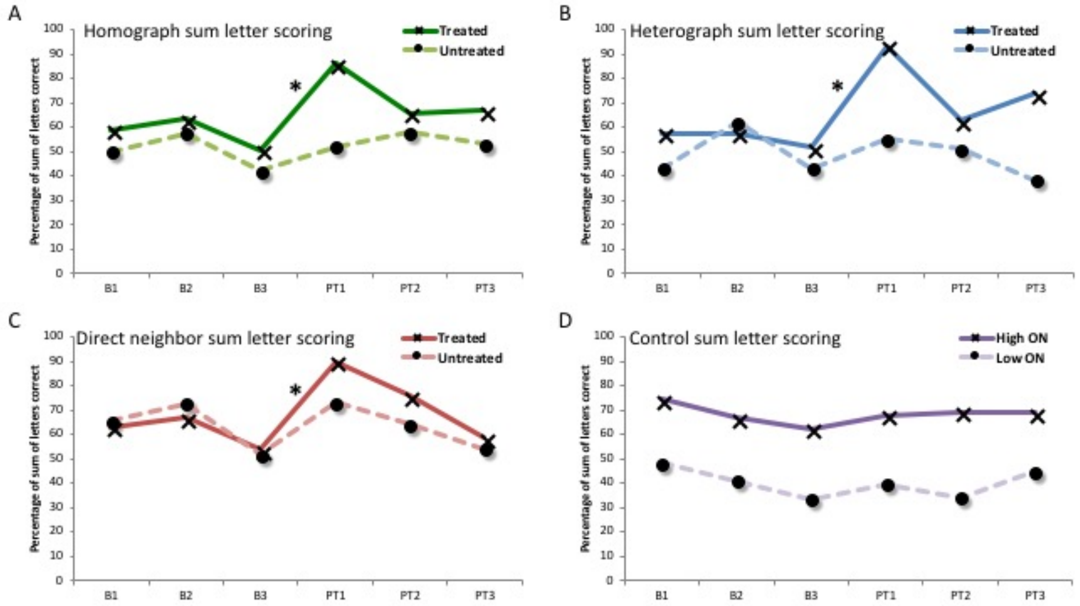

Figure A1. Graphs to show written naming of the a) the homographs (subsets 1&2), b) the heterographs (subsets 3 &4) c) non-homophonic controls (5&6) and d) the orthographic controls (7&8) for the summed letter accuracy

Table A1. Results of the WEST t-tests and their significance for treatment across the three baselines and first post-test for the summed letter accuracy.

	Sum letter accuracy			
Subset (degrees of freedom)	West trend	West-ROC	TSI	Consistent across sum/average letter
Homographs: 1. Treated (1,21)	1.88*	3.82**	✓	yes
2.Untreated (1,21)	-0.42		✗	yes
Heterographs: 3. Treated (1,13)	2.92**	4.38**	✓	yes
4.Untreated (1,13)	1.01		✗	yes
Non-homophone: 5. Treated (1,23)	2.91**	4.73**	✓	yes
6.Direct N (1,23)	0.14		✗	yes
Controls: 7.High ON (1,25)	-1.24		✗	yes
8. Low ON (1,22)	-1.65		✗	yes

$^{*}p$<.05 $^{**}p$<.001

TSI= Treatment specific improvement

✓= both West-trend and West-roc were significant suggesting significant improvement over the course of the study and during the treatment phase signifying treatment related improvement

✗= any change over the study could not be attributed to the treatment

Consistent across sum/average letter: is the same TSI shown if the analysis is conducted with the sum of letters correct per word or the average letter percentage correct per word.

Appendix B

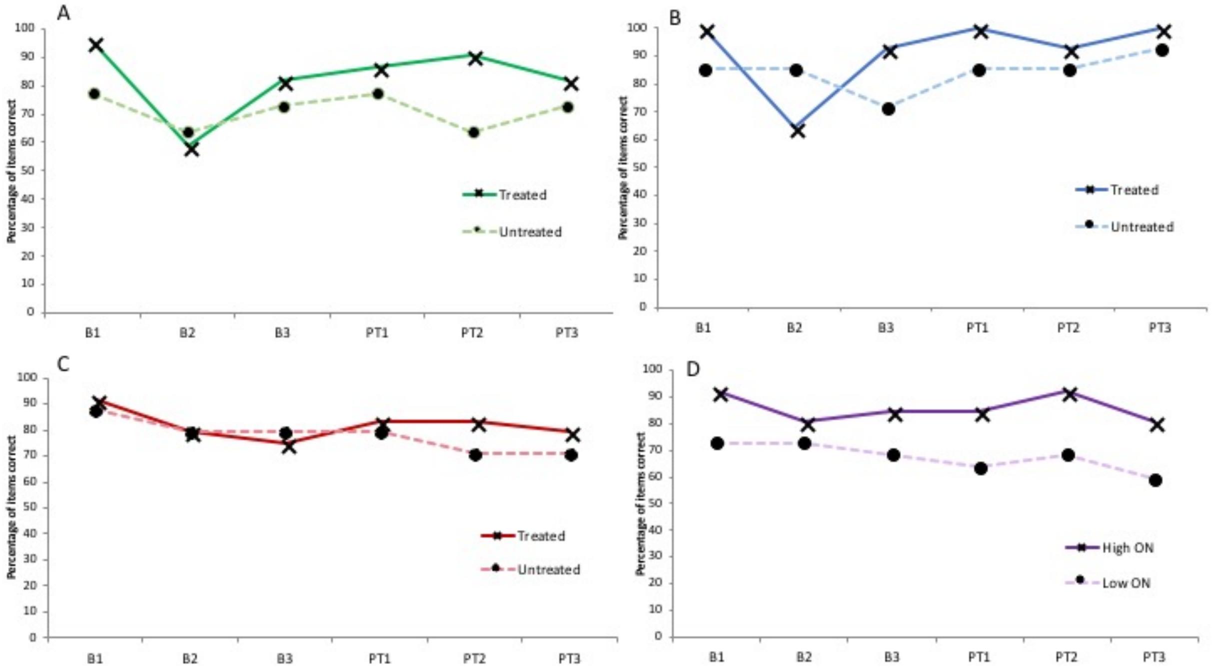

Figure B2. Graphs to show the spoken production of the a) the homographs (subsets 1&2), b) the heterographs (subsets 3 &4) c) non-homophonic controls (5&6) and d) the orthographic controls (7&8)

Table B2. Results of the WEST one sample t-tests and their significance for treatment for spoken naming.

	One Post-test			Three Post-test			Consistent across one/three post-tests
Subset (degrees of freedom)	West trend (t)	West-ROC (t)	TSI?	West trend (t)	West-ROC (t)	TSI?	
Homographs: 1. Treated (1,21)	-0.22		✗	0.49		✗	yes
2.Untreated (1,21)	0.33		✗	-0.25		✗	yes
Heterographs: 3. Treated (1,13)	2.28*	10.20**	✓	2.25*	1.68	✗	no
4.Untreated (1,13)	-0.41		✗	-0.78		✗	yes
Non-homophone: 5. Treated (1,23)	-1.57		✗	-0.97		✗	no
6.Direct N (1,23)	-0.95		✗	-1.52		✗	yes
Controls: 7.High ON (1,25)	-0.84		✗	-0.59		✗	yes
8. Low ON (1,22)	-1.23		✗	-1.18		✗	yes

$^{*}p<.05$ $^{**}p<.001$ (one-tailed)

TSI= Treatment specific improvement

✓= both West-trend and West-roc were significant suggesting significant improvement over the course of the study and during the treatment phase signifying treatment related improvement

✗= any change over the study could not be attributed to the treatment

Consistent across one/three post-tests: is the same TSI result shown in analysis that include one or three post-tests.

Appendix C

Table C1. Results of the WEST one sample t-tests and their significance for treatment across the three baselines and three post-tests for written naming.

	Whole word accuracy			Average letter accuracy			Consistent across word/letter
Subset (degrees of freedom)	West trend (t)	West-ROC (t)	TSI?	West trend (t)	West-ROC (t)	TSI?	
Homographs: 1. Treated (1,21)	2.53**	4.62**	✓	1.72		✓	no
2.Untreated (1,21)	1.11		✗	0.29		✗	yes
Heterographs: 3. Treated (1,13)	2.74*	4.60**	✓	1.41		✓	no
4.Untreated (1,13)	-0.12		✗	-1.00		✗	yes
Non-homophone: 5. Treated (1,23)	2.87 **	6.15 **	✓	0.77		✓	no
6.Direct N (1,23)	1.30		✓	-1.72		✗	yes
Controls: 7.High ON (1,25)	0.04		✗	-0.49		✗	yes
8. Low ON (1,22)	-0.38		✗	-1.07		✗	yes

**p*<.05 ***p*<.001 (one-tailed)

TSI= Treatment specific improvement

✓= both West-trend and West-roc were significant suggesting significant improvement over the course of the study and during the treatment phase signifying treatment related improvement

✗= any change over the study could not be attributed to the treatment

Consistent across word/letter: is the same TSI result shown across both word and letter accuracy

Appendix D

Table A4. The Pearson’s correlation r value and significance $^{+}$ for all the psycholinguistic variables and the orthographic overlap, number of neighbours in the treated set and treatment effect for the accuracy analysis. Items in italics indicate which variables significantly correlated with generalisation (WEST-ROC) and their co-variables and therefore were included in the regression

	Frequency	Length	Regularity	OrthN-freq	OrthN	PhonoN-freq	PhonN	Orth-O	Neighbours
Length	0.02								
Regularity	-0.07	-0.26*							
OrthN-freq	0.01	-0.67**	0.35**						
OrthN	-0.11	-0.3*	0.1	0.14					
PhonN-freq	-0.03	-0.49**	0.22*	0.66**	0.05				
PhonN	0.1	-0.19	-0.02	0.01	0.65**	0.02			
Orth-O	-0.22	0.11	0.26*	0.28*	-0.09	0.35**	-0.06		
Neighbour	0.02	*-0.51***	*0.27**	*0.61***	0.18	*0.44***	0.15	*0.29**	
WEST-ROC - Whole word	0.1	-0.08	0.02	0.15	-0.04	0.07	0.00	0.06	*0.20**
WEST-ROC – Average letter	-0.04	0.13	0.03	-0.06	0.06	-0.06	0.01	0.05	-0.04

$^{+}$ Due to negative correlation the reverse end of the p-value was taken

$^{*}r \geq 0.2$ plus $p<.05$

** $r \geq 0.2$ plus $p<.001$

Frequency= written lemma frequency

Length= number of letters

Regularity= whether a word was spelled regularly or not

OrthN-freq= the number of orthographic neighbours

OrthN= the average frequency of the orthographic neighbours

PhonN-freq= the number of phonological neighbours

PhonN= the average frequency of the phonological neighbours

Orth-O= measure of orthographic similarity between each untreated word and every item in treated set using the ‘ends-first’ coding system

Neighbours= the number of items that were zero, one and two Levenshtein distance neighbours of the treated set in the untreated set

Appendix E

First regression model: WEST-ROC (whole word accuracy) ~ length + regularity + orthN-freq + PhonoN-freq + Orth-O + Neighbours

	Sum of sq	RSS	AIC	F value	P value
<none>		325.49	137.15		
length	0.82	328.36	132.09	0.25	0.62
regularity	0.72	328.27	132.06	0.22	0.64
orthN-freq	2.06	329.60	132.50	0.63	0.43
PhonoN-freq	0.41	327.96	131.96	0.13	0.72
Orth-O	0.06	327.60	131.84	0.02	0.89
Neighbours	7.51	335.06	134.27	2.32	0.13

Length= number of letters
Regularity= whether a word was spelled regularly or not
OrthN-freq= the number of orthographic neighbours
PhonN-freq= the number of phonological neighbours
Orth-O= measure of orthographic similarity between each untreated word and every item in treated set using the ‘ends-first’ coding system
Neighbours= the number of items that were zero, one and two Levenshtein distance neighbours of the treated set in the untreated set

List of figure captions

Figure 1: Three architectures for homophone representation in the phonological lexicon a) separate lemmas and shared (Levelt et al., 1999) b) no lemma level and separate modality specific word forms (Caramazza et al., 2001) and c) lemmas but no word form representations and interactive activation (Middleton et al., 2015; Schwartz et al., 2006).
Figure 2: Timeline of baselines (B1, B2, B3), treatment sessions (T1-12) and post-tests (P1, P2, P3).
Figure 3: Accuracy for homographs on whole word accuracy (A), and average letter accuracy (B)

Figure 4: Accuracy for heterographs on whole word accuracy (Panel A), letter percent accuracy (Panel B)

Figure 5: Accuracy for non-homophonic controls and direct neighbours) on whole word accuracy (Panel A), and average letter accuracy (Panel B).

Figure 6: Accuracy for orthographic neighbourhood controls on whole word accuracy (Panel A), and average letter accuracy (Panel B).

Figure 7. A boxplot to show the relationship between the number of 0, 1 and 2 Levenshtein distance neighbours and treatment related-improvement

Tables

Table1. English Language background assessment for CWS

Task	**Number of items**	**CWS correct %**	**Control Mean %**	**Control Min. %**	**Control SD %**	**N of Controls**
Comprehension						
PALPA 47 oral word-picture matching	40	100%	98	87	2.67	31
PALPA 48 written word-picture matching	40	100%	99	87	1.53	32
Pyramids and Palm Trees Test; three pictures	52	96%	98	94	-	13
Single word Repetition						
Word repetition (Bangor University)	80	100%	-	-	-	-
English Non-word repetition (Bangor University)	40	100%	-	-	-	-
Spoken Naming						
Object Naming battery (list B)	81	91%	-	91	-	40
Action Naming battery (list B)	50	86%	-	86	-	40
Visual Word Recognition						
PALPA 25: Real and non-words lexical decision[++]	120	98%	99	-	0.54	26
PALPA 3 minimal pairs: Written word selection	72	97%	97	-	2.35	23
Reading						
PALPA 19 upper case to lower case letter matching	26	96%	100	96	0.77	26
PALPA 32 grammatical class reading	**80**	95%*	100	-	1.45	32
PALPA 34 lexical morphology and reading	**90**	90%*	-	-		-
English Reading: Regular words (Bangor University)[+]	40	98%	99	95	1.65	20
English Reading: Irregular words (Bangor University)[+]	40	100%	99	95	1.73	20
English Reading: Non-words (Bangor University)[+]	**40**	**43%****	95	83	5.10	20
PALPA 53: Cross Modality Comparisons						
Oral picture naming	**40**	**65%****	99	-	0.87	29
Irregular	20	80%	-	-	-	-
Regular	20	60%	-	-	-	-
Written picture naming	**40**	**15%****	97		3.33	27
Irregular	20	5%	-	-	-	-
Regular	20	25%	-	-	-	-
Spelling to dictation	**40**	**7.5%****	99	-	-	2
Irregular	20	0%*	-	-	-	-
Regular	20	15%*	-	-	-	-
Repetition	40	97.5%	99	-	2.05	28
Irregular	20	100%	-	-	-	-
Regular	20	95%	-	-	-	-

[+] Control scores taken from aged matched control monolingual participants from Bangor University. All other control data is from the appropriate published test.

[++] Average mean and standard deviations across the subsets

Bold represents scores which are impaired (2.5 standard deviations below control mean)

** Scores that are at least two standard deviations below control mean.

* Scores that are thought to be impaired to some degree, but normative data is not available.

Table 2. In-depth spelling-to-dictation assessments

Task	Number of stimuli	CWS % correct	Control Mean %	Control Range %	Control SD %	Control N
Regularity *(Bangor University)*						
Regular words	80[+]	**41**	97	80-100	5.23	20
Irregular words	80[+]	**18**	91	55-100	12.40	20
Non-words	80[+]	**39**	74	53-93	12.07	20
Frequency (Words collapsed across JHU[#] lists)						
High-frequency words	147	21*	-	-	-	-
Low-Frequency words	146	12*	-	-	-	-
Length (JHU list length)						
4-5 Letters	27	30*	99	-	-	5
6 letters	15	7*	92	-	-	5
7+ letters	28	7*	93	-	-	5
Grammatical Category (JHU part-of-speech)						
Nouns	28	4*	-	-	-	-
Verbs	28	4*	-	-	-	-
Adjectives	28	7*	-	-	-	-
Nonwords	34	12*	-	-	-	-
Concreteness (JHU)						
Concrete words	21	19*	98	-	-	5
Abstract words	21	0*	91	-	-	5
Copy						
Direct copy (PALPA 44)	40	98	-	-	-	-
Delayed copy transcoding						
Regular words	20	80*	-	-	-	-
Irregular words	20	45*	-	-	-	-

[#] JHU= John Hopkins University Dysgraphia Battery (Goodman & Caramazza, 1985)

[+] Double administration for CWS, therefore control number of items = 40

Impaired Scores in bold (2.5 standard deviations below control mean)

* Scores that are thought to be impaired to some degree, but normative data is not available.

Table 3. Errors made by CWS in spelling to dictation taken from Roberts (2013)

Error type	**Example**	**Words % (N= 599)**	**Non-words % (N=45)**
Phonologically plausible errors	Into-> INTU	**39**	1
Real word error	Work->WORD	**7**	36
Phonologically implausible nonwords (50% or more letters correct)	Hotel->HOTOL	**33**	7
Phonologically implausible nonwords (less than 50% target letters correct)	Feather->FAFARA	**8**	1
Cross language errors	Nine->NAIN	**12.52**	2.22

Table 4. Matching of experimental subsets on accuracy, log frequency and other psycholinguistic variables.

	Homographs		Heterographs		Non-homophonic controls			
Subset	1	2	3	4	5	6	7	8
	Treated (N=22)	Untreated partners (N=22)	Treated (N= 14)	Untreated partners (N=14)	Treated (N=24)	Untreated Direct neighbours of 5 (N=24)	Untreated High ON (N=26)	Untreated Low ON (N=22)
Written accuracy Baseline 1	9.09	9.09	14.29	7.14	16.67	16.67	25.93	4.55
Written accuracy Baseline 2	13.64	9.09	21.43	21.43	25.00	16.67	22.22	4.55
Written accuracy Baseline 3	0.00	4.55	14.29	14.29	12.50	8.33	23.63	4.55
Spoken accuracy Baseline 1	95.45	77.27	100.00	85.71	83.33	79.17	88.89	72.73
Spoken accuracy Baseline 2	59.09	63.64	64.29	85.71	70.83	70.83	77.78	72.73
Spoken accuracy Baseline 3	81.82	72.73	92.86	71.43	66.67	75.00	81.48	68.81
Frequency: written lemma (log10)	2.06	2.20	2.41	2.62	2.67	2.76	2.86	2.67
Frequency: written word form (log10)	0.57	0.71	1.06	1.19	1.05	0.96	1.15	0.85
Frequency: spoken lemma (log10)	0.69	0.84	1.17	1.29	1.17	1.14	1.25	1.00
Frequency: spoken word form (log10)	1.88	2.01	2.23	2.50	2.48	2.59	2.74	2.49
Syllables	1.14	1.14	1.07	1.07	1.00	1.00	1.00	1.00
Phonemes	3.64	3.64	3.00	3.00	3.09	3.25	3.38	3.19
Letters	4.32	4.32	4.21	4.36	3.95	4.00	3.88	5.05*
Orthographic neighbourhood density	7.41	7.41	7.36	8.57	10.91	10.21	8.62	0.57*
Orthographic neighbourhood freq.	106.61	106.61	62.27	385.25	86.53	129.83	310.67	27.82*
Phonological neighbourhood density	16.05	16.05	21.29	21.29	21.09	20.50	14.92*	11.62*
Phonological neighbourhood freq.	86.14	86.14	451.21	451.21	184.22	306.20	279.22	286.50
Regularity	0.91	0.91	0.57	0.50	0.92	0.92	0.81	0.55*

*Significantly different from matched subset

Orthographic neighbourhood density = number of words with one letter difference

Orthographic neighbourhood freq. = average frequency of all the orthographic neighbours

Phonological neighbourhood density = number of words with one phoneme difference

Phonological neighbourhood freq.= average frequency of all the phonological neighbours

Table 5. Results of the WEST one sample t-tests and their significance for treatment across the three baselines and first post-test.

	Whole word accuracy			Average letter accuracy			Consistent across word/letter
Subset (degrees of freedom)	West trend (t)	West-ROC (t)	TSI?	West trend (t)	West-ROC (t)	TSI?	
Homographs: 1.Treated (1,21)	3.25**	5.25**	✓	2.38*	3.91**	✓	yes
2.Untreated (1,21)	0.62		✗	-0.66		✗	yes
Heterographs: 3.Treated (1,13)	3.38**	5.12**	✓	3.25**	4.41**	✓	yes
4.Untreated (1,13)	1.10		✗	0.94		✗	yes
Nonhomophone: 5.Treated (1,23)	4.26 **	8.26 **	✓	3.20**	5.07**	✓	yes
6.Direct N(1,23)	1.73*	1.83 *	✓	1.00		✗	no
Controls: 7.High ON(1,25)	0.46		✗	-1.19		✗	yes
8.Low ON(1,22)	-1.00		✗	-1.81		✗	yes

**p*<.05 ***p*<.001 (one-tailed)

TSI= Treatment specific improvement

✓= both West-trend and West-roc were significant suggesting significant improvement over the course of the study and during the treatment phase signifying treatment related improvement

✗= any change over the study could not be attributed to the treatment

Consistent across word/letter: is the same TSI result shown across both word and letter accuracy

Table 6. Final regression model for the whole word analysis

Variable	DF	Sum of sq	RSS	AIC	F value	P value
Intercept			330.32	124.74		
Neighbours	1	13.42	343.74	127.04	4.31	.04*

* Significant p-value of <.05

Neighbours= the number of items that were zero, one and two Levenshtein distance neighbours of the treated set in the untreated set

Figures

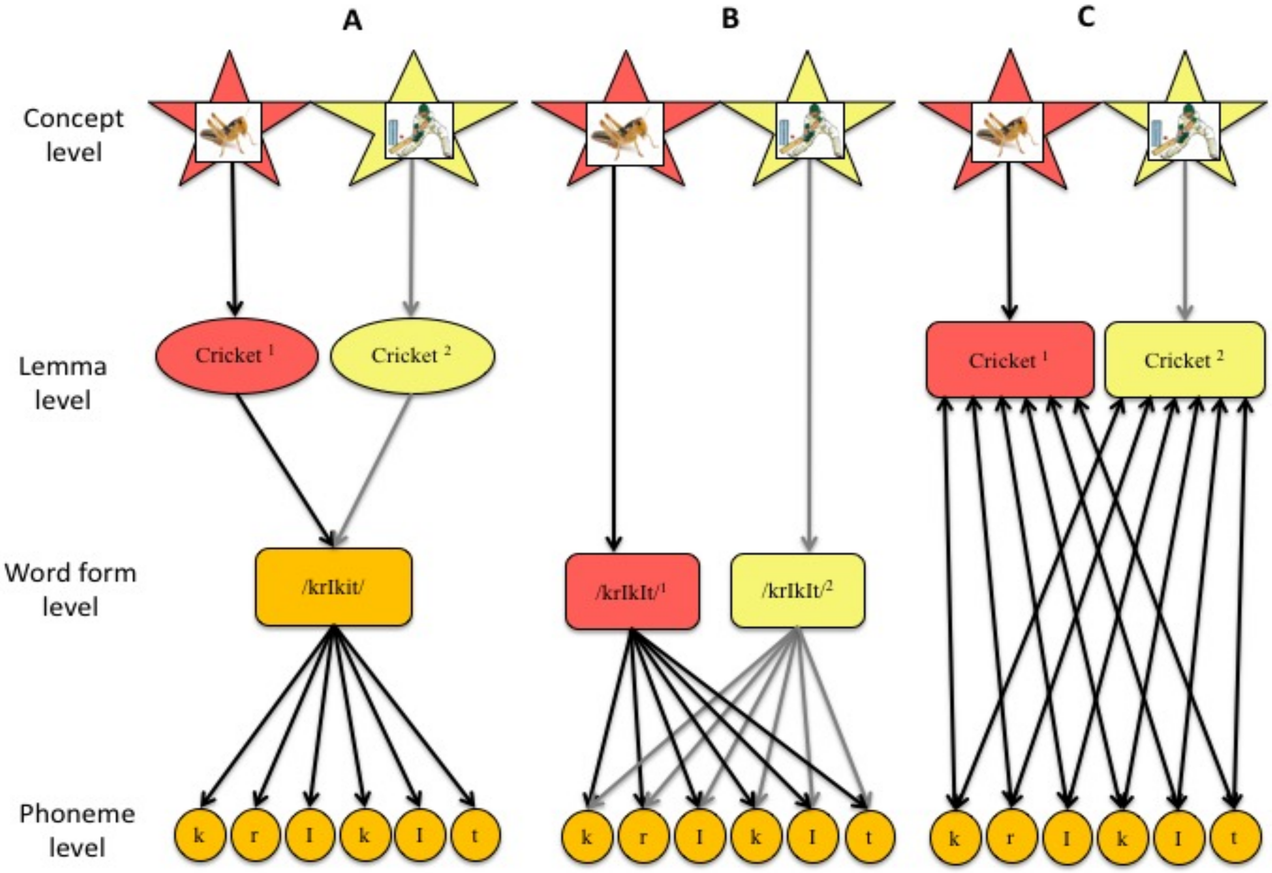

Figure 1

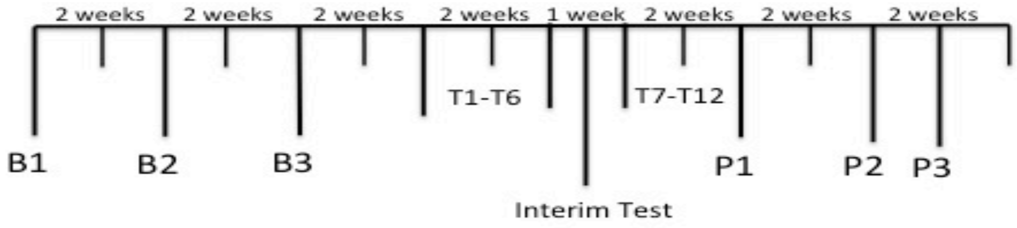

Figure 2

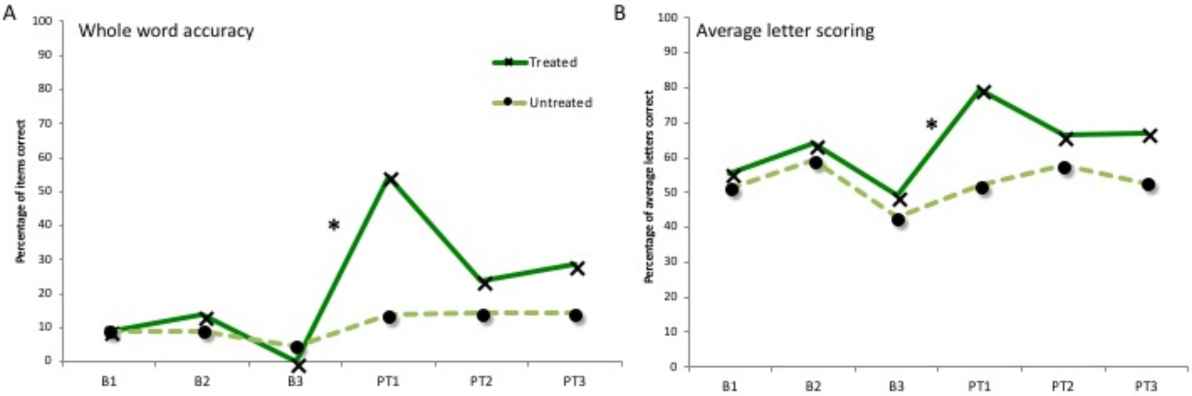

Figure 3

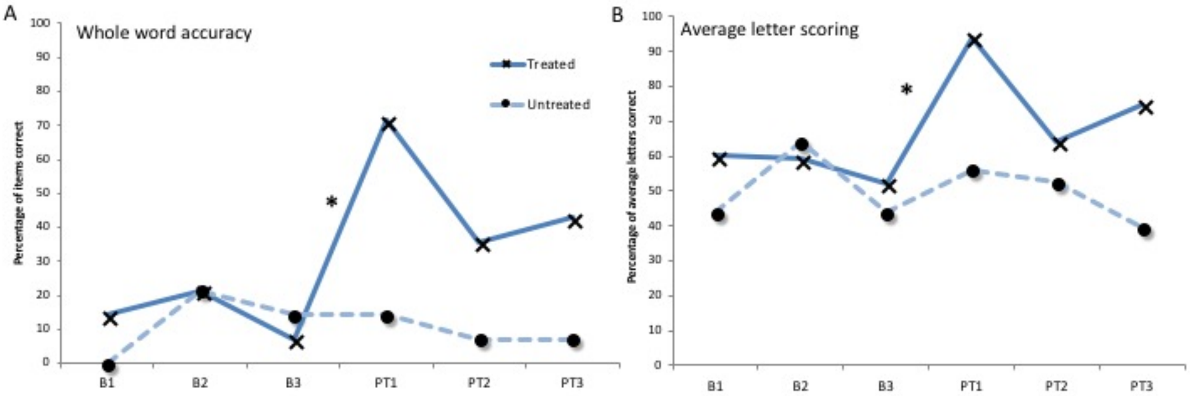

Figure 4

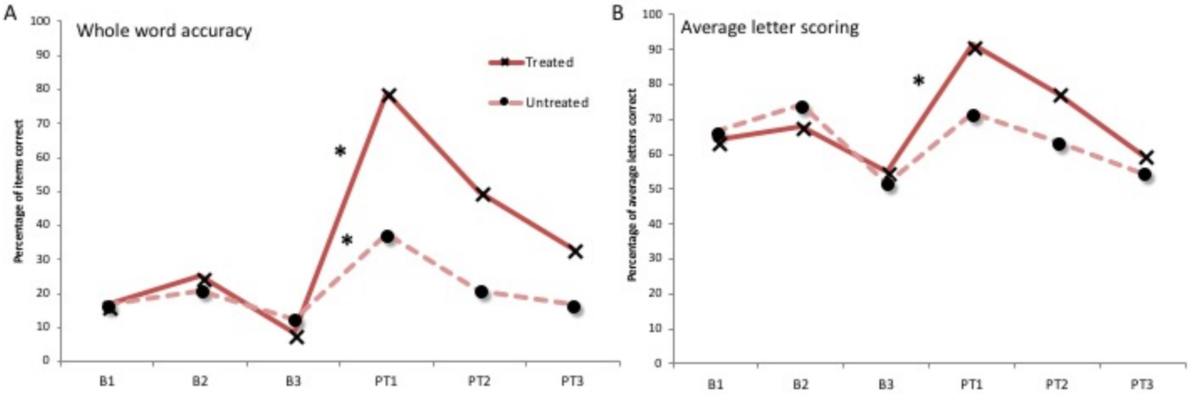

Figure 5

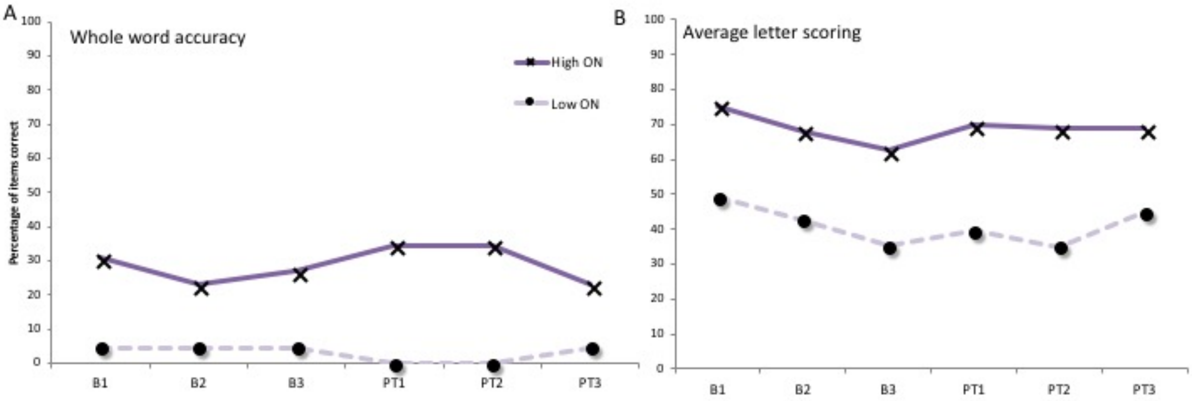

Figure 6

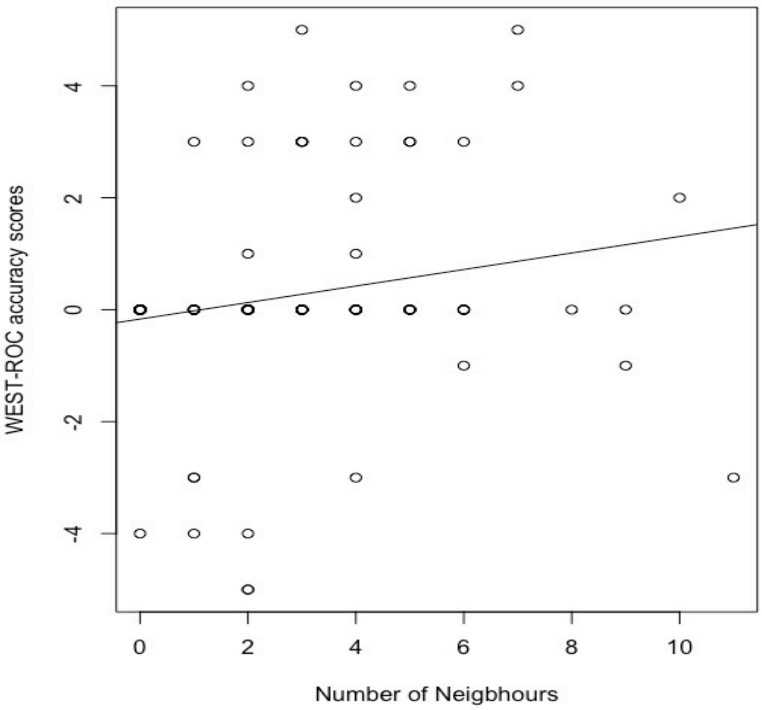

Figure 7