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.

Access to the published version:
http://dx.doi.org/10.1080/02643290802003000

Predicting generalisation in the training of irregular word spelling: Treating lexical spelling deficits in a child

Saskia Kohnen*, Lyndsey Nickels*, Max Coltheart* & Ruth Brunsdon#

Macquarie Centre for Cognitive Science

Macquarie University, Sydney

#The Children's Hospital at Westmead, Syndey

Short title: Generalisation in irregular word spelling

Address for correspondence:

Saskia Kohnen,

Macquarie Centre for Cognitive Science (MACCS),

Macquarie University, Sydney,

NSW 2109, Australia.

Tel: +61 2 9850 4439

Fax: +61 2 9850 6059

Email: skohnen@maccs.mq.edu.au

Abstract

This paper presents a single case study investigating the mechanisms underlying generalisation of treatment benefits to untrained words in spelling. Brunsdon, Coltheart and Nickels (2005) observed that untreated words that improved tended to be those whose errors were closest to being correct prior to treatment. These words also tended to be high in written frequency. The present study employed the same treatment techniques used by Brunsdon et al. with KM, a developmental surface dysgraphic. During a first treatment the characteristics of words whose spelling improved without specific training were identified. These characteristics were then used in a second treatment to test whether it was possible to predict generalisation. The results showed that treatment generalisation to untreated irregular words was best predicted by neighbourhood size and frequency. We suggest that the processes underlying treatment generalisation are based on the interaction between the orthographic lexicon and the graphemic buffer. Clinical implications are discussed.

Acknowledgements

We would like to thank KM and her family for their cheerful and enthusiastic participation. Thankyou also to KM's teachers and the control children. We are grateful to Alan Taylor for statistical support. Helpful comments on an earlier draft were provided by M.-J. Tainturier and an anonymous reviewer. This paper was prepared while Saskia Kohnen was supported by an International Macquarie University Research Scholarship (iMURS). Lyndsey Nickels was funded by an NHMRC Senior Research Fellowship and Max Coltheart by Australian Research Council Special Research Centre and Federation Fellowship grants.

Introduction

Cognitive neuropsychological research concerned with the spelling system has usually focussed on theoretical aspects of dysgraphia (e.g. Badecker, Hillis, & Caramazza, 1990; Caramazza & Miceli, 1990; McCloskey, Badecker, Goodman-Schulman, & Aliminosa, 1994; McCloskey, Macaruso, & Rapp, 2006; Romani, Ward, & Olson, 1999; Tainturier & Rapp, 2004; Temple, 1985; 1986; 1997). While there is less research evaluating treatments for dysgraphia, some well designed training studies with adults have been published (e.g. Beeson, 1999; Behrmann, 1987; De Partz, Seron, & Van der Linden, 1992; Rapp & Kane, 2002; Schmalzl & Nickels, 2006). In contrast, there are currently only two cognitive neuropsychological treatment studies of spelling disorders in childhood (Brunsdon, Coltheart, & Nickels, 2005; Stadie & van de Vijver, 2003). Such treatment studies can be useful not only for empirically evaluating clinical treatment methods but can also provide a novel way to investigate theoretical proposals concerning the spelling process (e.g. Brunsdon et al., 2005; Raymer, Cudworth, & Haley, 2003; Sage & Ellis, 2006; Weekes & Coltheart, 1996). This paper reports a treatment study with a 9 year old surface dysgraphic child and is theoretically based on cognitive neuropsychological dual route models of spelling (Barry, 1994; Patterson & Shewell, 1987).

(Figure 1 about here)

Dual route models of spelling propose two main procedures by which spelling can be achieved, a lexical/semantic route and a sublexical route (see Figure 1) (Barry, 1994; Patterson & Shewell, 1987). The sublexical route converts phonemes (sounds) into graphemes (letters), using learned knowledge about the correspondences between sounds and letters. Thus, words with predictable phoneme-to-grapheme correspondences (regular words, such as <CAT>, <FOOD>, <BLADE>) can be spelled accurately using the sublexical processing route. Spelling unfamiliar words and nonwords (e.g. <GOP>, <NIVE>, <ZOOD>) can be successful via the sublexical route, too. However, in order to correctly spell words which have less predictable spellings, such as <HAVE>, <GONE> or <FRUIT> (henceforth referred to as irregular words), word specific knowledge is required about the letters that make up the spellings of these words and the order of these letters. This information is stored in a long term memory component, the orthographic lexicon. The lexical and the sublexical routes converge at the grapheme level. For example, when spelling the word ‘gone’, the abstract letter nodes at the grapheme level <G>, <O>, <N>, <E> receive activation from the orthographic lexicon. There is evidence for a feedback loop between the orthographic lexicon and the grapheme level (see Miceli & Capasso, 2006 for a review of the evidence), such that letters that are active in turn activate the orthographic representations that contain them (see Figure 2)[1]. The grapheme level, also known as the graphemic output buffer, operates like a short term memory component where letters are held active until letter shapes for writing or letter names for oral spelling have been accessed. Note that throughout this article, the written form will be presented in <>.

(Figure 2 about here)

---

[1] A very convincing case was presented by McCloskey, Macaruso and Rapp (2006). Their patient’s lexical errors (e.g., <SOLD> written as <COLD>) were best explained by activation from the grapheme level feeding back to the lexical level. First, it was shown that the number of lexical errors the patient made was greater than predicted by chance, i.e. the lexical errors were ‘true’ lexical errors which resulted from the selection of a lexical entry (e.g. <COMPUTE> was activated when asked to spell <CARPET>). Secondly, the lexical errors systematically included letter intrusions from previous spelling responses. Persisting activation at the grapheme level from previous responses was fed back into the lexical level and resulted in the selection of incorrect lexemes.

Impairment to any of the components of the spelling process will result in difficulties in spelling (dysgraphia). However, depending on the locus of the impairment, the nature of the problem will differ. Here we are concerned with what is known as surface dysgraphia, which presents as a difficulty in spelling irregular words while spelling of regular words and nonwords is relatively preserved. Surface dysgraphia arises from processing difficulties along the lexical spelling route, while the sublexical pathway is less impaired or unimpaired. Surface dysgraphia has been described in adults after brain injury (e.g. Beauvois & Derouesné, 1981; Behrmann & Bub, 1992; Goodman & Caramazza, 1986) as well as in children without neurological disorders (e.g. Temple, 1986; 1997). Lexical spelling deficits can also occur in combination with deficits along the sublexical route of spelling.

The lexical processing difficulties that give rise to the symptoms of surface dysgraphia can arise at three different points in the spelling system. Access to the orthographic lexicon may be disrupted (i.e. link 1 in Figure 1). Representations within the lexicon itself may be of poor quality (i.e. incomplete or absent) or the transfer of information from the orthographic lexicon to the graphemic buffer may be dysfunctional (i.e. link 2 in Figure 1). For example, patient A could have perfectly intact representations in the orthographic lexicon but a disruption in accessing this information (i.e. only link 1 is disrupted). Patient B may have intact access to the orthographic lexicon but the representations are not fully specified. Patient C may have intact access procedures as well as intact representations but there is a deficit in transferring information to the graphemic buffer (i.e. only link 2 is disrupted). All of these different impairments in isolation, as well as in combination, lead to difficulties in spelling of irregular words. Currently, it is hard to distinguish between these different functional impairments as most cognitive models of spelling require further specification in order to determine what behavioural differences might be expected as a function of these different loci of abnormality.

Treatment of lexical spelling deficits

Treatment of lexical spelling deficits usually aims to strengthen lexical representations within the orthographic lexicon (or access to these lexical representations) or to facilitate semantic to orthographic activation (e.g. Beeson, 1999; Behrmann, 1987; Brunsdon et al., 2005; De Partz et al., 1992; Raymer et al., 2003; Schmalzl & Nickels, 2006; Stadie & van de Vijver, 2003). Training is usually successful in that the trained words are spelled significantly better after intervention than before (e.g. Aliminosa, McCloskey, Goodman-Schulman, & Sokol, 1993; Beeson, Hirsch, & Rewega, 2002; Hillis & Caramazza, 1987). This item-specific training effect could arise from repeated attempts to activate a word's representation during training (Rapp & Kane, 2002). Another possibility is that, due to training, new lexical representations are formed or incomplete ones are completed.

Improved spelling is not necessarily restricted to training words. There are reports in the literature of improvement in irregular words that were not specifically targeted for training (e.g. Behrmann, 1987; Brunsdon et al., 2005; Raymer et al., 2003; Weekes & Coltheart, 1996). This is called treatment generalisation. Given that there are many words in English and that treatment is usually an expensive resource, treatment generalisation is not only of theoretical but also of clinical importance. However, it is as yet unclear which processes underlie treatment generalisation.

Treatment generalisation has been observed as a result of enrichment of semantic representations. Hillis (1991) treated HG who had a semantic impairment. After treatment, HG had improved not only her written picture naming for trained items but also for untrained items within the same category. In this case, treatment generalisation arose as an effect of enhanced semantic processing.

Successful treatment of graphemic buffer deficits can also lead to improved spelling of untreated items (Rapp, 2005; Rapp & Kane, 2002; Sage & Ellis, 2006).

Rapp and Kane (2002) propose different possible underlying mechanisms that may be responsible for generalisation mediated at the level of the graphemic buffer: repeated exposure to words may improve the buffer’s capacity to maintain representations; scanning speed in the buffer may be increased; the speed of transfer to the letter-shape level may be increased. It is also possible that training leads to strengthening of orthographic representations, which may be less affected by the buffer damage than weaker representations (Rapp & Kane, 2002).

It has been assumed that treatment of orthographic lexicon deficits is likely to yield benefits only to words that were specifically trained (Rapp, 2005; Rapp & Kane, 2002). However, there are reports of patients with lexical spelling deficits who showed effects of treatment generalisation – that is, untrained words improved after training (e.g. Behrmann, 1987; Brunsdon et al., 2005). As these patients did not have graphemic buffer impairments, it is unlikely that generalisation effects were due to enhanced functioning of the buffer. It is just such reports of treatment generalisation in individuals with lexical spelling deficits but functional graphemic buffers that are relevant to the current study. These cases provide evidence that there are generalisation effects that are mediated by the orthographic lexicon. What mechanisms might be responsible for these lexical generalisation effects?

Weekes and Coltheart (1996) explore possible mechanisms underlying treatment generalisation with reference to the improvement of trained and untrained words in <u>reading</u>. They propose that there are two underlying effects, one specific and one more general. The specific effect concerns the improvement of trained items and is argued to be due to restoration of (or improved access to) word representations in the orthographic lexicon. Treatment, however, may also lead to a general improvement of the access procedure for all lexical entries[2]. It is this overall enhanced access mechanism that has been argued to result in the benefit for untreated words (Brunsdon et al., 2005; Weekes & Coltheart, 1996).

Not all words benefit equally from treatment generalisation. One factor that might be influential in acquired cases is whether a word was premorbidly present in the patient's language processing system. Weekes and Coltheart (1996) explored this by asking their patient, NW, to define spoken words, the assumption being that a correct definition indicated the premorbid presence of that word in the patient's language processing system. They reported that untrained premorbidly present words improved from 20-30% correct prior to treatment to 40-65% correct after treatment. In contrast, untrained words premorbidly not present hardly changed at all: 0-15% were read correctly prior to treatment and 20-25% after treatment.

It could be that treatment generalisation varies according to the quality of representation of untreated words: the better the representation, the more likely is the word to benefit from the improved retrieval mechanism. While Weekes and Coltheart (1996) used the ability to define a word as a measure of whether there was a representation for a word, attempting to measure the quality of the lexical representation was not part of their study.

The study that, to our knowledge, has investigated the cognitive mechanisms underlying treatment generalisation in most detail (Brunsdon et al., 2005) was undertaken with a developmental surface dysgraphic, MC. Apart from a significant

---

[2] The exact nature of this improvement is not specified and would depend on the particular architecture of the theory used. However, one possibility, in a connectionist model, might be that there is a general increase in connection strength across all the links to the lexicon (see Dell et al., 1997, for an example of a connectionist model where this could apply, albeit in a different domain).

improvement for spelling of trained words, MC also showed a generalisation effect for untreated words. Brunsdon et al. argued that a surface dysgraphic child may gradually build up (partially) correct orthographic representations for words. However, this process may be slowed compared to a child with normal acquisition. They proposed that, in developmental surface dysgraphia, there might also be a continuing problem with the retrieval of information from the orthographic lexicon.

Brunsdon et al. investigated the idea that the likelihood of generalisation varies depending on the quality of pre-treatment representations. The quality of lexical representations was measured by comparing the target spelling to the error response MC made prior to training. Brunsdon et al. used five different methods of aligning the target spelling and the error response to estimate the degree of match between the two items. The five coding schemes used to do so were referred to as 'match schemes' and were originally developed with reference to models of visual word recognition (Davis & Bowers, 2006). Each match scheme will be described briefly below, examples are provided in Appendix A. A detailed description and evaluation of the different schemes can be found in Davis and Bowers (2006). Match 1 simply aligns the target spelling and the error response at the leftmost letter. Match 2 and 3 are vowel centred match schemes. The first vowel is used as an anchor. For match 2, surrounding consonants are aligned from left to right, while for match 3, the remaining consonants are aligned from right to left. For match 4, error and target are aligned at the exterior letters, positions for middle letters are defined relative to the last and initial position, i.e. one position forward from the initial letter or one position backward from final position. Match 5 is a form of spatial coding with position independent letter codes. Letter identity is not tied to letter position. Order of letters is coded by relative activity, with the first letter receiving most activation and successive letters receiving progressively less activation. One advantage of spatial coding in the context of this research is that it recognises the similarity between target spellings and error responses for transposed letters (e.g. <FRIEND> and <FREIND> and <YACHT> and <YATCH>). For all match schemes, the similarity of the error to the target spelling is expressed as a numerical value. For example, a value such as 3/5 can be read as: 3 letters of the error response are exactly the same as the target spelling which is 5 letters long.

Overall, Brunsdon et al. (2005) found that words which improved without training had pre-treatment error responses that were closer to the correct spelling (or closer to a match value of 1.0) than those words that did not improve. Brunsdon et al. were cautious to suggest that one particular coding scheme was superior in predicting generalisation to any other since the coding schemes were highly correlated. However, schemes that put slightly more emphasis on exterior letters, namely match schemes 4 and 5, were consistently the strongest predictors of generalisation. Brunsdon et al. also found that untrained words that generalised were more likely to have correct exterior letters prior to training than those words that did not improve without specific training (75% vs. 46% correct exterior letters).

While most emphasis was placed on the match schemes, Brunsdon et al.'s exploration of the data also included word characteristics of the untrained words, such as length, frequency of occurrence and neighbourhood size. The best statistical model for generalisation included high written and spoken frequency and high values on match 4. Put in simpler terms, words of high frequency whose exterior letters were correct before training were the words most likely to improve without specific training.

Brunsdon et al. also analysed how the errors changed over the course of the treatment study for untrained words. Significant increases occurred on the vowel centred match schemes 2 and 3. This means that most changes happened in the spelling of the vowels. Vowels are the most likely part of the word to contain inconsistent or irregular spellings (Venezky, 1970).

To summarise, treatment generalisation can be due to improvements at the level of the orthographic lexicon. Weekes and Coltheart suggested that training leads to an enhancement of lexical retrieval, which benefits words that do not receive specific treatment. How much untreated words benefit from improved access mechanisms may vary as a function of the quality of the pre-treatment lexical representation of that word. Brunsdon et al. used measures of similarity between pre-treatment error response and target spelling as a proxy for the quality of a lexical representation.

Case Study

The two studies reported in this paper used a cognitive neuropsychological framework to guide treatment of lexical spelling deficits in the case of dysgraphia in childhood. The studies had both clinical and theoretical aims.

The clinical aim was to improve lexical spelling. We chose a treatment design that would allow us to investigate whether treatment was effective and long lasting as well as to examine which tasks, that were not necessarily part of the training focus, would benefit from treatment.

The theoretical aim was to investigate the factors underlying treatment generalisation. Specifically, we set out to see if it would be possible to predict treatment generalisation for the spelling of irregular words. We also investigated possible generalisation of the spelling treatment to irregular word reading, and spelling of homophones.

Case History

KM, a student at a mainstream primary school in suburban Sydney, was 9 years and six months old at the beginning of this study. At the age of 4, and prior to any formal instruction of reading or writing, she sustained a severe traumatic head injury as a result of a car accident. A CT scan showed multiple skull fractures, a small right parietal subarachnoid haemorrhage and a small contusion in the left parietal lobe. KM had posttraumatic amnesia for five weeks after the incident. In the years following the accident, KM was regularly seen for medical and neuropsychological review. KM consistently presented with significant right-sided hemiplegia, attention deficits, and emotional and behavioural difficulties. Her fine motor control was poor. Overall cognitive abilities (such as intellectual skills, memory, attention, visual perception, new learning and problem solving) were average (as assessed using the Differential Ability Scales (DAS, Elliot, 1990) and the Wechsler Preschool and Primary Scale of Intelligence (WPSSI-III, Wechsler, 2002). There were concerns regarding KM's reading and spelling.

(Table 1 about here)

The current study is the second of two treatment studies conducted with KM. Details of the first study can be found in Kohnen, Nickels, Brunsdon and Coltheart (2008). The initial assessment of KM's spoken and written language skills was conducted prior to the first study and one year prior to the current study. In conversation, KM's language was free of syntactic or morphological errors. KM made occasional phonological and phonetic/articulatory errors (interchanging /s/ - /θ/, /z/ - /ð/, /s/ and /□/). Her receptive vocabulary was normal but her expressive vocabulary was more than a standard deviation below the mean for her age, i.e. in the impaired range. KM's reading and spelling performance was characterised by sublexical as well as lexical difficulties (for further details of the initial testing see Kohnen et al., 2008). Following the initial assessment, KM was given treatment to improve her impaired sublexical spelling processes. KM was taught two spelling rules that she had not acquired spontaneously. The treatment goals were achieved and we assessed KM again in order to see whether the lexical reading and spelling difficulties were still present. The results of this assessment (i.e. pre lexical treatment assessment) will be presented below, together with a summary of earlier results where required. Raw scores as well as normative data for all assessments can be found in Table 1.

Phonological Skills

KM had no difficulties with the segmentation of sounds in words or nonwords (Psycholinguistic Assessment of Language Processing in Aphasia (PALPA), subtests 16 & 17, Kay, Lesser, & Coltheart, 1992) and her auditory rhyme judgment was flawless (PALPA 15). On the Children's Test of Nonword Repetition (Gathercole & Baddeley, 1996), KM's scores were slightly lower than the means for the oldest control group consisting of children aged 8-8;11. KM may have mild difficulties with the repetition of multisyllabic nonwords, which could be attributed to a deficit at the level of the phonological output buffer.

Naming and Semantics

KM's expressive vocabulary as measured on the Assessment of Comprehension and Expression 6-11 (ACE6-11) Picture Naming Test (Adams, Cooke, Crutchley, Hesketh, & Reeves, 2001) was low but within normal limits. Her performance was almost errorless on the somewhat easier PALPA spoken picture naming task (PALPA 53). On a written picture naming task using the same items as the spoken picture naming task, KM made 21 spelling errors. Half of these errors included at least some orthographic knowledge of the target spelling (e.g. <SCISSORS> spelled as <SCISERS>), the other half were entirely phonologically plausible spellings (e.g. <IRON> spelled as <IEN>), there was also a single semantic error (<SCREW> spelled as <NAIL>). KM obtained an average score on the Peabody Picture Vocabulary Test (PPVT, Dunn & Dunn, 1997), indicating a receptive spoken vocabulary within the normal range. Her spoken word picture matching (PALPA 47) and written word picture matching (PALPA 48) was almost flawless. Overall, KM's spoken word vocabulary and semantics were unimpaired.

Single Letter Processing

KM could both name and sound out all lower case and upper case letters accurately. She was also error free in writing letters when given their sounds.

Visual Lexical Decision

We tested KM's visual lexical decision with regular and irregular words, nonwords and pseudohomophones (e.g. <JALE>) (PALPA 27) and used modified t-test statistics (Crawford & Garthwaite, 2002; Crawford & Howell, 1998) to compare KM's performance to that of a control group of five children who were in her class at school. KM's accuracy was similar to that of her classmates on regular ($t = 0$ $p = 1.0$) and irregular words ($t = 1.66$ $p = .158$) as well as nonwords ($t = 1.07$ $p = .166$). However, KM accepted significantly more pseudohomophones than the normal control group ($t = 2.84$ $p = .036$).

Single Word Reading

KM's reading was measured on the Castles and Coltheart word and nonword lists (Appendix A in Coltheart & Leahy, 1996). When using Crawford's (Crawford & Garthwaite, 2002; Crawford & Howell, 1998) modified t-tests to compare KM's reading to the normative data reported by Edward and Hogben (1999), KM's reading of regular and irregular words and nonwords was not significantly below the performance of her age group (regular words, t = .-962 $p = .338$; irregular words, $t = -1.48$ $p = .141$; nonwords, $t = -1.05$ $p = .287$). Overall, KM's single word (and nonword) reading was within the normal range.

Single Word Spelling

KM's spelling was tested using the word and nonword items from Robinson and Weekes (1995). The lists are matched for length, frequency and neighbourhood size. Neighbourhood size is defined as the number of words which differ from the target item by only one letter in the same serial position (Coltheart, Davelaar, Jonasson, & Besner, 1977). We compared KM's performance to nine age-matched normal controls. KM's spelling was normal for regular words (t = -1.21 p = .260) and nonwords (t = -.193 p = .852). KM's spelling of irregular words was significantly worse than that of her age-matched controls (t = -2.4 p = .043).

When attempting to spell irregular words, KM predominantly produced phonologically plausible spellings (59%, 13/21; e.g. <DOVE> as <DUV>) suggesting use of a sublexical procedure. 14% of the errors were homophone confusions (e.g. <SCENT> as <SENT>). All homophone confusion errors were such that the regular word was produced instead of the irregular target. It is difficult to determine the underlying processing deficit for these errors. KM might not have had a fully specified orthographic lexical entry for the irregular homophone (e.g. <SCENT>) and thus computed the spelling response using sublexical procedures. 23% (5/21) of KM's errors on the irregular words could not be classified (e.g. <TOMB> as <TOWM>). These errors could partially stem from poor knowledge of phoneme-to-grapheme correspondences. For instance, <OW> in the example above is not a frequent way to write /u:/. Overall, most errors could be attributed to a lexical processing difficulty and attempted use of a sublexical procedure.

KM only showed an effect of length for spelling of irregular words, but not for spelling of nonwords or regular words (nonwords, Jonckheere trend test z = .62 p = .543; regular words: z = 1.62 p = .106; irregular words: z = .2.9 p = .004). As the regular and irregular word lists we used do not differ significantly in neighbourhood size or frequency (Robinson & Weekes, 1995), these variables cannot account for the difference reported above (see Sage & Ellis, 2004). None of the spelling errors made on the word and nonword lists take the typical form of errors suggestive of a graphemic buffer deficit, i.e. (simple) substitutions, additions, omissions and movement errors (Miceli, Silveri, & Caramazza, 1987), without also being phonologically plausible. Given these results, there is no clear evidence for a graphemic buffer deficit in KM.

Summary of assessment, results and interpretation

KM had difficulties in visual lexical decision, with errors particularly for pseudohomophones. Her reading of regular and irregular words and nonwords was normal, with irregular word reading in the low average range. KM's spelling of regular words and nonwords was also unimpaired. However, her spelling of irregular words was below age expectation. KM's errors suggested an over-reliance on the sublexical spelling route. Overall, KM's spelling difficulties can be attributed to a processing difficulty along the lexical spelling route, without an additional graphemic buffer impairment. We cannot determine whether KM had poor access to lexical representations, poor quality representations within the lexicon or whether there was a problem in transferring lexical information to the graphemic buffer. While at the initial assessment KM was classified as a mixed dysgraphic, after treatment of the sublexical spelling route (Kohnen et al., 2008), she had become a surface dysgraphic.

Treatment Study 1

Rationale and Goals

The clinical goal for study 1 was to improve irregular word spelling. The theoretical goal was to ascertain whether Brunsdon et al.'s finding that treatment of irregular word spelling showed generalisation to untreated items could be replicated in another child with surface dysgraphia. If so, we wanted to identify the factors influencing the generalisation of treatment to irregular word spelling.

We were also interested in generalisation across tasks. Overall, there are mixed findings in the literature regarding the generalisation between reading and spelling: In most previous treatment studies targeting reading, there was no generalisation to spelling (e.g. Scott & Byng, 1989; Weekes & Coltheart, 1996). However, in one study (Brunsdon, Hannan, Coltheart, & Nickels, 2002) spelling did improve after the treatment of irregular word reading. Generalisation in the reverse direction was found for MC (Brunsdon et al., 2005): reading for irregular words improved after training irregular word spelling. Therefore, in order to monitor possible changes in KM's reading of the items used for the spelling study, KM was also asked to read all items at every assessment point.

Generalisation effects do not usually extend to homophones. For example, Behrmann (1987) found improvement of spelling for treated but not untreated homophones. Similarly, Hillis (1993) found improvement for reading, spelling and comprehension of treated but not untreated homophones. Brunsdon et al. (2005) also found no changes for untreated homophone reading or spelling. We included spelling of homophone pairs in this study in order to monitor whether treatment would impact on KM's spelling of homophones.

Stimuli

A list of monosyllabic words that were irregular for reading was obtained by removing plurals, possessives, homophones, and regular past tense words from the monosyllables of the CELEX database (Baayen, Piepenbrock, & Van Rijn, 1993). Irregularity was defined in reference to a set of grapheme-to-phoneme correspondences listed by Rastle and Coltheart (1999).

Only words that KM indicated she knew, when she was presented with the spoken word, were used. The resulting 291 words were spelled and read twice by KM (baseline 1.1 and baseline 1.2) in order to determine whether her spelling and reading was improving prior to treatment (Franklin, 1997). This was not the case as both KM's reading and spelling performance were stable over the two initial baselines (reading, McNemar's Chi Square = .000 p = 1.0; spelling: Chi Square = 1.227 p = .721).

(Table 2 about here)

Spelling. KM spelled 35.4% (103/291) of the words correctly at baseline 1.1 and 38.1% (111/291) at baseline 1.2. See Table 2 for details of the error analysis. KM's spelling errors were mostly phonological approximations of the correct spelling suggesting that KM attempted to spell the words using sublexical spelling procedures.

In order to have an overall measure of closeness between the spelling errors and the target spellings at the letter level we performed the following analysis: Each letter in a response that also occurred in the target spelling was scored as 1, incorrect letters as 0. The scores were summed and divided by the number of letters in the target spelling. For example, <TOWM> for <TOMB> would receive a score of 3/4, <DUV> for <DOVE> 2/4. At baseline 1.1, 73.9% of letters included in the errors were correct. At baseline 1.2, this was the case for 74.2% of the letters. The high number of correct letters in the incorrect spellings indicates that overall the errors KM made were quite similar to the target (correct) spellings.

(Table 3 about here)

Reading. At both baseline 1.1 and baseline 1.2, KM read 65.3% of the words correctly. About half of KM's errors on both baselines were phonological approximations to the correct word, indicating sublexical decoding of letters into sounds. A high proportion of the reading errors were whole word substitutions. For details of the error analysis see Table 3.

The 291 words were then divided into three sets. One set consisted of 73 words which were spelled correctly on both baselines. The remaining words were spelled incorrectly on at least one of the baselines. Of these, 42 words were used for training. The training words were matched to the remaining 176 words on written CELEX frequency, spoken CELEX frequency[3] (Baayen et al., 1993), length in letters and number of orthographic neighbours, defined as the number of words which differ from the target item by only one letter in the same serial position (Coltheart et al., 1977). We counterbalanced how many times items were spelled and read correctly and on which baseline.

We also determined match values for each item using the five match schemes (Davis & Bowers, 2006) giving measures of the similarity of KM's spelling errors to their targets. For items that were misspelled on both baselines but in two different ways, we calculated match values for both misspellings. For each match scheme we compared the group means for misspellings on baseline 1.1 to the misspellings of baseline 1.2. There was no significant difference between baseline 1.1 and 1.2 (related t-test p = .458 – .894). We then randomly selected for each item either the misspelling from baseline 1.1 or from baseline 1.2. Means for each match scheme of this mixed group were compared to means of baseline 1.1 and means of baseline 1.2 means. Again, there were no significant differences indicating that the misspellings chosen were representative in their similarity to the target spellings at

[3] Note that data used to determine frequency and neighbourhood are based on adult language. There are few norms available for children.

both baselines. The set of training items and the set of untrained control items were matched on the five match values.

Since Brunsdon et al. (2005) found that the correctness of exterior letters was a factor determining generalisation, we added another binary variable, called ‘exterior letters correct’. For this variable we simply determined whether the initial and final letters of the misspellings were correct or not.

A final binary variable we added was called ‘phonologically plausible error’. Misspellings that sounded like the target when read out loud were assigned to the category ‘phonologically plausible error’. The rationale behind this variable was to distinguish between misspellings that were produced sublexically, which were more likely to be ‘phonologically plausible errors’ than those that were orthographically based errors. In addition, unlike the other measures, this procedure does not involve complicated and time-consuming computations and may therefore be suitable for use in clinical practice if found to be a relevant factor in generalisation. The training and control sets were matched on the mean number of phonologically plausible errors.

(Table 4 about here)

Treatment Procedure

Table 4 shows the overall treatment design. Treatment was conducted over a period of two weeks. There was one training session with the first author and nine consecutive practice sessions with the Special Aide Teacher at school (who had been trained by the first author). KM’s training was followed by a post treatment assessment and a follow up assessment. At each assessment point, the same tests were administered: reading and spelling of the 291 irregular words, reading, definition and spelling of a set of homophones. Assessments were conducted over at least 2 sessions to avoid item repetition. One half of the items was spelled in the first session and read in the second session and reverse for the second half of the items.

Treatment method

The treatment method is exactly as used by Brunsdon et al. (2005) for their set 1 training. The target items were printed on flash cards in lower case (size 48 Arial font).

In the initial training session, KM was shown a flash card with the correct spelling on it. The word was read out aloud by the first author. KM was asked to copy the word. The word was then removed from sight and after a 10 second delay (the first author counted to ten), KM had to write down the word again. No instructions were given as to how she should try to remember the spelling. Finally, KM was asked to spell the word to dictation. If on any of these steps KM made an error, she was shown the flash card again. KM copied 41/42 words correctly initially with the target word in sight and 40/42 words were initially correct in delayed copying, 39/42 words were spelled correctly initially when required to write to dictation, i.e. only 4 words required repetition of any step in the initial session

Practice sessions followed the initial training session and were discontinued once KM had spelled at least 90% of the words correctly on 2 consecutive sessions. During practice sessions, the flash cards were shuffled into a random order and then read out aloud to KM. She was asked to repeat the word to ensure correct perception and then write it down. KM was given feedback as to whether her response was correct or not. If incorrect, KM was shown the flash card with the correct spelling for 5 seconds. She was then asked to write it from memory before moving on to the next word.

Analysis

Firstly, we checked for each item group whether whole word accuracy differed over the two baselines using McNemar tests. In order to measure possible improvement following training, the number of words correct at post test was compared to the number of whole words correct at the highest baseline using McNemar tests. We chose to use the highest baseline as a measure of pre-treatment performance because this is a more conservative estimation of the effects of treatment. Given that a non significant result yielded by comparison to the highest baseline cannot necessarily be interpreted as a lack of improvement, we also compared whole word accuracy over several assessment points using Cochran's Q. If not specifically mentioned otherwise, two tailed tests are reported.

(Figure 3 about here)

Results

In the following paragraphs the results for reading and spelling of the trained and untrained irregular words will be reported. Results regarding the spelling of homophones will be presented after training study 2.

Irregular word spelling. Results for irregular word spelling for training phase 1 are shown in Figure 3. There were two baselines for study 1 referred to as baseline 1.1 and baseline 1.2. There were two post training assessments, one immediately after training (post test 1) and the other one 12 weeks after the training had been discontinued (follow up 1).

Baseline performance for the training items was stable (McNemar's test, exact $p = 1.0$). Directly after training, KM's spelling performance had improved significantly compared to the highest baseline (McNemar's test, exact $p < .001$, one tailed). The decrease from the immediate post test (PT1) to the follow up assessment was significant (McNemar's test, exact $p < .001$). However, KM's performance at follow up 1 was still significantly better than at baseline 1.2 (McNemar's test, exact $p < .001$).

KM did not improve significantly on the untrained words from baseline 1.1 to baseline 1.2 (McNemar's test, exact $p = .410$). Generalisation to untrained exception words was evident in KM's spelling directly after treatment. The improvement from the highest baseline to the immediate post test was highly significant (McNemar's test, exact $p < .001$). The slight decrease from directly after the treatment to follow up was not significant (McNemar's test, exact $p = .430$). KM's performance at follow up 1 was still significantly higher than at baseline 1.2 (McNemar's test, exact $p < .001$).

Directly after training, there was a significant decrease in spelling of words that were correct at both baselines (McNemar's test, exact $p = .016$). While KM's level of accuracy for these words at follow up 1 was not significantly different to post test 1 (McNemar's test, exact $p = 1.0$) it was still a significant decrease from baseline (McNemar's test, exact $p = .008$).

To sum up, there was a significant improvement for spelling of treated and untreated words, which was long lasting. Words spelled correctly at baseline decreased in spelling accuracy.

Analyses of treatment generalisation. We conducted several analyses to investigate possible factors influencing generalisation. Comparisons were made between the untrained words that improved without specific treatment and those that did not. Firstly, we investigated the intercorrelations between the variables we hypothesized could be predictors of generalisation. These variables were four word characteristics; length in letters, neighbourhood size, written frequency and spoken frequency, and 8 characteristics of the pre-treatment errors - the five match values measuring similarity of targets and errors, a composite score of all match values, 'external letters correct', 'phonologically plausible error', and a broader similarity score measuring whether the target and the response diverged by 1, or 2 or more letters ('similarity 0-2').

(Table 5 about here)

The correlation matrix is shown in Table 5. Like Brunsdon et al. (2005), we found that all match values were highly significantly intercorrelated. There were some significant correlations between the match values and some of the other error measures. Significant correlations between match values and word characteristics were rare. As expected, written and spoken frequency were highly correlated as well as neighbourhood size and length.

We also investigated which of these variables were significantly correlated with the binary outcome variable ‘generalisation’ at two different assessment points: directly after the treatment (PT1) and more than two months later (FU1). The correlations can be found in Table 5. Most variables that were correlated with generalisation at post test 1 correlated even more strongly with generalisation at follow up 1. The variable ‘phonologically plausible error’ did not correlate with generalisation at PT1 or FU1 and was hence dropped from further analyses.

In order to determine which variables predicted generalisation when the intercorrelations between variables were controlled for, we then conducted a series of logistic regressions. The first regression included all variables that were significantly correlated with generalisation at post test 1 (i.e. written frequency, spoken frequency, neighbourhood size, match 1, match 3, match 4, composite match score, similarity 0-2). In this analysis, written frequency, neighbourhood size and match 3 were all significant predictors of generalisation (written frequency, $B = .533$ $S.E. = .233$ $p = .022$; neighbourhood size, $B = .090$ $S.E. = .41$ $p = .03$; match 3, $B = 2.195$ $S.E. = 1.106$ $p = .047$). In addition, because of the significant correlations between the different error characteristic variables, we conducted a supplementary series of analyses which included the word characteristics (written frequency, spoken frequency, and neighbourhood size) but only one of the error characteristic variables that correlated significantly with generalisation per analysis (e.g. match 1, match 3, etc.). None of the error characteristics reached significance in these analyses, with the exception of match 3 (B = 2.195 S.E. = 1.106 p = .047). Thus the best statistical model to predict generalisation directly after training included neighbourhood size, written frequency and match 3 as a measure of target-error similarity. When using match 3 to compare a spelling error to its target spelling the two items are aligned at the vowel. The items in study 1 that were most likely to show generalisation were those that were high on all three of these measures.

The same analyses were undertaken for generalisation at the follow up assessment after the end of training phase 1 (FU1). In this analysis, written frequency, neighbourhood size and match schemes 4, 5 and the composite match score were all significant predictors of generalisation (written frequency, B = .709 S.E. = .264 p = .007; neighbourhood size, B = .204 S.E. = .052 p = .000; match 4, B = 4.505 S.E. = 2.019 p = .026; match 5, B = 8.645 S.E. = 3.723 p = .020; composite match score, B = -13.613 S.E. = 5.676 p = .016). Again, we conducted supplementary regressions with the word characteristics (i.e. length, written frequency, spoken frequency and neighbourhood size) and a single error characteristic variable in each analysis. In these analyses, when only a single error characteristic was entered in the analysis, match 3, 4, 5 and the composite match score were each significant predictors (match 3, B = 3.107 S.E. = 1.249 p = .016; match 4, B = 2.925 S.E. = 1.066 p = .006; match 5, B = 4.226 S.E. = 1.574 p = .007; composite match score, B = 2.990 S.E. = 1.330 p = .025). Given that match 3, 4 and 5 were all significant predictors by themselves and because of the high intercorrelations between these variables, we would be reluctant to make strong claims as to which of these variables was the ‘true’ predictor. Neighbourhood size and written frequency were always highly significant in all of these analyses. Thus, the best statistical model for predicting generalisation at follow up included neighbourhood size, written frequency, and additionally a measure of similarity of target error - match 3, 4, 5, and/or composite match score.

Summary and discussion of treatment generalisation

Overall, the best predictors of generalisation in all analyses were written frequency and neighbourhood size. A measure of target-error similarity was also found to be a significant predictor (match 3, 4, and 5 and the composite match score). In comparison, Brunsdon et al. (2005) found the best statistical model to include written frequency and match 4. Match values 1, 3, 4, and 5 were all strong predictors of generalisation in Brunsdon et al.'s analyses.

Irregular word reading

(Figure 4 about here)

The data for irregular word reading can be found in Figure 4.

Training words. KM's reading of the training words was stable at baseline (McNemar's test, exact $p = 1.0$). Directly after training (PT1) KM's reading had improved. While this improvement only approached significance (2 tailed) compared to the highest baseline (McNemar's test, exact $p = .092$), there was a significant improvement over the two baselines and post test 1 (Cochran's $Q(2) = 8.59$ $p = .014$). The improvement from post test to follow up was not significant (McNemar's test, exact $p = .219$). However, KM's performance at follow up 1 was significantly better than before training (baseline 1.2) (McNemar's test, exact $p < .001$).

Untrained words. KM's reading performance for the untrained words was stable at baseline (McNemar's test, exact $p = .360$). The improvement from the higher baseline 1.1 to post test 1 was not significant (McNemar's test, exact $p = .212$). Performance over the baselines and post test 1 was only just significant (Cochran's $Q(2) = 6.0$ $p = .050$). At follow up 1, KM's reading improvement between baseline 1.1

and follow up approached significance (McNemar's test, exact p = .055). The increase in reading accuracy for the untrained words over the four assessment points (baseline 1.1, baseline 1.2, post test 1, follow up 1) was significant (Cochran's Q(3) = 10.87 p = .012).

In summary, there were some improvements for the reading of trained as well as untrained irregular words during study 1, even though reading had not been the focus of the training.

Treatment Study 2

Frequency and neighbourhood size were the most consistent predictors of spelling generalisation in study 1. In a second study, we investigated whether it would be possible to predict generalisation for different groups of untreated words based on the results of study 1. We compared whether words that were higher in frequency and neighbourhood size showed greater generalisation than words that were lower on these measures. The similarity between target and error at pre-test (match values) was also predictive of generalisation in study 1 but this was a weaker effect. Hence in study 2 we examined whether there was any benefit over and above that of frequency and neighbourhood size by contrasting sets that were of equally high frequency and neighbourhood size but varied in target-error similarity (match values).

We also included reading of all irregular words as well as spelling of homophones to monitor possible changes in KM's performance in these tasks across the study.

Stimuli

We reassessed spelling for all 291 words that were included in study 1. Since there were no significant differences for spelling between follow up 1 and this assessment (McNemar's test, exact p = .268), we used follow up 1 and the new assessment as baselines for the second study. These two assessment points are now referred to as the baselines for study 2: baseline 2.1 and baseline 2.2.

(Figure 5 about here)

In study 2, we included 138 of the original 291 words. These 138 words were untreated control items in study 1 and had been misspelled at least once on the two baselines for study 2. Spelling accuracy did not change between baseline 2.1 and baseline 2.2 for these items (McNemar's test, exact p=.728). The item structure can be seen in Figure 5. For each of these words we computed the values of similarity between target and error using match schemes 1-5. For each match scheme we compared the group means for misspellings on baseline 2.1 to the misspellings of baseline 2.2. While there was no significant difference between the baselines (related t-test p = .081 – .187), baseline 2.2 match scores were higher and the differences between the baselines were much closer to being significant when compared to study 1, where p values ranged from .458 - .894. Thus, for words that had two different error responses on the two baselines for study 2, we used the last misspelling, i.e. the misspelling made at baseline 2.2, as this was likely to be the most accurate spelling and hence provide the most conservative measure.

The 138 words were split into 2 groups which differed significantly in the two best predictors of generalisation: neighbourhood size and written frequency. Since written and spoken frequency are so highly correlated, the two groups differed significantly not only in written but also in spoken frequency. Thus, one group of words was high in neighbourhood size and frequency while the other group was low in these variables, the two best predictors of generalisation. Each of these two groups was split in two again, so that there were four sets of words (see Figure 5). All four sets were matched on length. The group of words with low values on the best predictors was divided into two sets, one of which was to be trained. We will refer to these sets as 'low value training set' and 'low value control set'. The 'low value training set' and 'low value control set' were matched for frequency, neighbourhood size, length and all five match schemes. Both low value item sets were significantly lower in spoken frequency, written frequency and neighbourhood size than each of the high value sets (see Table 6 for means). It is important to keep in mind that we only used words that KM knew, hence the overall frequency of the words was relatively high (mean log written frequency=2.57, with a range of 1.2 – 4.78),

The group of words that had high values on the best predictors of generalisation was divided into two sets that were matched on neighbourhood size and frequency. In addition, one of the high value sets was significantly higher on match schemes 3, 4 and 5 than the other high value set and both the low value sets. The two high value sets will be referred to as 'high value neighbourhood size and frequency set' and 'high value neighbourhood size, frequency and matches set'. The 'high value neighbourhood size and frequency set' does not differ significantly on the match schemes from the 'low value control set' and the 'low value training set'.

Hypotheses

We predicted that significantly more of the treated words would be spelled correctly after treatment than before. The results of study 1 led us to predict that 'low value control set' words, which were of low frequency and neighbourhood size and untreated would not improve. In contrast, significantly more correct word spellings were expected for both untrained high value sets. We were interested in whether 'high value neighbourhood size, frequency and matches set' would improve even more than 'high value neighbourhood size and frequency set' because the former set had additionally high values on match schemes 3, 4, and 5.

Treatment method

The treatment method was exactly as in study 1. KM had one initial training session with the first author and 6 consecutive training sessions with her Special Aide Teacher at school until she had spelled 90% of the words correctly in two consecutive sessions.

(Figure 6 about here)

Results for irregular word spelling

Results are shown in Figure 6. Spelling performance for all words was stable over the two baselines for study 2 (McNemar's test, exact p = .268). We also conducted a direct post test (post test 2) in the week after training and a follow up (follow up 2), which took place 12 weeks after the end of the training. Performance after training was compared to the highest baseline, which was baseline 2.2.

Training words. Spelling performance for the 'low value training set' did not change over the two baselines (McNemar's test, exact p = 1.0). The increase in spelling accuracy for trained words from directly before training (baseline 2.2) to directly after training (post test 2) was highly significant (McNemar's test, exact p < .001). While the decrease in spelling accuracy from post test 2 to follow up 2 was significant (McNemar's test, exact p = .013), KM still spelled the trained words significantly better at follow up 2 than before training (BL2.2-FU2, McNemar's test, exact p = .004).

Low value control set. Spelling performance for the 'low value control set' did not increase significantly prior to training (McNemar's test, exact p = .180). There was no change in accuracy directly after training (BL2.2-PT2, McNemar's test, exact p = 1.0) or at follow up (BL2.2-FU2, McNemar's test, exact p = .804).

While baseline performance was stable, there was a slight but non-significant increase from baseline 2.1 to baseline 2.2 for low value control words. Spelling accuracy at post test 2 was at a similar level as at baseline 2.2. Thus there was some increase in spelling performance over the three assessment points. However, the conservative comparison of the highest pre-test to the post test may deflate this effect. We thus decided to compare spelling performance over the two baselines 2.1 and 2.2 and the immediate post test. However, there remained no significant change even over the three assessment points (Cochran's $Q(2) = 4.78$ $p = .092$). We can therefore be confident in saying that there is no generalisation for the low value control set (see Figure 6).

High value neighbourhood size and frequency set. Spelling performance was stable at baseline (McNemar's test, exact $p = .774$). Accuracy increased significantly from baseline 2.2 to post test 2 (McNemar's test, exact $p = .007$). There was no significant change from post test 2 to follow up 2 (McNemar's test, exact $p = 1.0$). At follow up 2, KM spelled significantly more words correctly than at baseline 2.2 (McNemar's test, exact $p = .031$).

High value neighbourhood size, frequency and matches set. There was no significant increase in spelling performance at baseline (McNemar's test, exact $p = .791$). The increase in spelling accuracy from baseline 2.2 to post test 2 was significant (McNemar's test, exact $p = .008$). KM maintained high spelling accuracy from post test to follow up (McNemar's test, exact $p = .791$). At follow up 2, spelling accuracy for the high value neighbourhood size, frequency and matches set was significantly higher than at the highest baseline (BL2.2-FU2, McNemar's test, exact $p = .019$).

Comparing treatment effects. In order to compare the relative effects of treatment on the different sets, we compared 'change scores' for the different item sets using Wilcoxon two samples test (Howard, 2004). We were especially interested if there was any difference in the amount of improvement as a result of treatment generalisation between the ‘high value neighbourhood size and frequency set’ compared to the ‘high value neighbourhood size, frequency and matches set’. Change scores were computed as follows: post test spelling score minus the average of the two pre training spelling scores.

The improvement on trained items was significantly higher than improvement for any of the control groups (‘low value training set’ vs. ‘low value control set’, Wilcoxon two samples, $z = 4.48$ $p = .001$; ‘low value training set’ vs. ‘high value neighbourhood size and frequency set’: $z = 2.95$ $p = .003$; ‘low value training set’ vs. ‘high value neighbourhood size, frequency and matches set’: $z = 3.51$ $p = .001$).

While the change score for the ‘high value neighbourhood size and frequency set’ was significantly higher than for the ‘low value control set’ (Wilcoxon two samples, $z = 2.02$ $p = .022$, one tailed), the difference between the ‘low value control set’ and the ‘high value neighbourhood size, frequency and matches set’ only approached significance (Wilcoxon two samples, $z = 1.51$ $p = .065$, one tailed).

Of most interest for this study was the question of whether there was a difference in the relative improvements for the two high value sets, possibly an advantage for the set of words that was not only high on frequency and neighbourhood size but also on the match values. There was no significant difference in change scores between the two high value control sets (Wilcoxon two samples, $z = 0.46$ $p = .646$), i.e. there was no additional benefit for the set of words with the higher match values.

Summary of results study 2

Training spelling of irregular words with relatively low frequency and neighbourhood size led to long lasting improvements in accuracy. While untrained words with low frequency and low neighbourhood size did not improve without specific training, words with high frequency and high neighbourhood size did improve without specific training. There was no additional benefit for words that also had high values on match schemes 3, 4 and 5.

(Figure 7 about here)

Irregular word reading

For training words in study 2 there were no significant gains in reading. Reading performance at baseline was stable (McNemar's test, exact p = .549) (see Figure 7). KM's reading accuracy was not significantly higher at any assessment point following training than at the highest baseline (BL2.1-PT2, McNemar's test, exact p = .338; BL2.1-FU2: p = .607). Comparing reading performance over 3 or 4 assessment points also did not reveal a significant increase in accuracy (Cochran's Q(2) = 3.89 p = .143; Cochran's Q(3) = 4.74 p = .192).

There were also no significant reading improvements for the 'low value control set' or the 'high value neighbourhood size and frequency set' when reading accuracy after training was compared to the best baseline performance (all comparisons, McNemar's test, exact p>.06) nor when looking at all assessment points (all comparisons, Cochran's Q>2.11 p>.34) (see Figure 7).

While reading accuracy for the 'high value neighbourhood size, frequency and matches set' was not significantly higher at any point after training when compared to the highest baseline, there was a significant increase over the 3 and 4 assessment points (Cochran's Q(2) = 6.13 p = .047; Cochran's Q(3) = 10.2 p = .017).

In summary, there was only a significant increase for reading of the untrained 'high value neighbourhood size, frequency and matches set'. However, note that due to the small number of items in each group and the relatively high accuracy of KM's reading before treatment, changes of statistical significance were difficult to demonstrate.

Homophone spelling

This study focussed on the investigation of mechanisms underlying treatment generalisation. In previous spelling treatment studies, no generalisation was found to untreated homophones (Behrmann, 1987; Brunsdon et al., 2005). It has been argued by both Behrmann and Brunsdon et al. that homophones need word specific spelling training and do not benefit from a general improvement in orthographic processing. Although enhanced processing of the semantic-orthography pathway may lead to improved homophone spelling, the methods used in this study did not focus on the semantic-orthography processing route. We thus did not expect a reduction in errors for homophone spelling.

(Figure 8 about here)

Prior to the first treatment study, KM misspelled 66/96 (baseline 1.1) and 61/96 (baseline 1.2) of the homophones (see Figure 8). At the end of the second treatment study (follow up 2), KM misspelled 54/96 homophones, i.e. 7 errors less than at baseline. This decrease in the number of incorrect spellings is not significant (McNemar's test, exact $p = .099$). There are also no significant changes in error type during the study (homophone confusion errors, McNemar's test, exact $p = .163$; phonologically plausible errors: $p = .860$; other errors: $p = .136$).

Overall, there were no significant improvements in spelling of homophones over the course of the intervention study. While there was a decrease in the number of errors made when spelling homophones, this was not significant. In Behrmann's (1987) study, spelling only improved for homophones that were specifically trained but untreated homophones did not improve. Our results could be taken as further support for Behrmann's (1987) interpretation that homophones require word specific training and do not benefit from treatment generalisation.

Overview of the results of treatment studies 1 and 2

Treatment was successful in that spelling of specifically practiced words improved as a result of training. The largest gains were evident directly after training; spelling performance declined somewhat at the follow up assessments. However, even 12 weeks after training had discontinued, spelling performance was still significantly better than before training. Thus, treatment effects were long lasting. The benefits of this treatment were not restricted to specifically trained words. Improvement was also evident for untrained irregular words. Untrained words that were most likely to benefit from treatment generalisation were of high frequency and high neighbourhood size. The similarity between pre-treatment error and target spelling (as measured by match schemes 3, 4 and 5) did not provide any additional benefit. While irregular words with relatively low frequency and neighbourhood size could improve as a result of specific training, they did not improve without training. Even though reading was not a focus of the training, reading for trained and untrained words improved during study 1. No significant gains were made in homophone spelling. There was also no change in the pattern of errors made when spelling homophones.

Further analyses investigating possible mechanisms underlying generalisation

Were the generalisation effects found in this study mediated at the level of the orthographic lexicon? At the start of the study KM showed an over-reliance on sublexical spelling procedures: she tended to make phonologically plausible errors. It was therefore possible that predictability of spelling might have affected generalisation. Hence it was important to investigate whether the words that generalised were actually more predictable overall compared to the words that did not generalise.

Predictability of exterior letters. Brunsdon et al. (2005) found that, for their case, match schemes which placed relative emphasis on exterior letters were especially predictive of treatment generalisation. For 75% of the words that improved without treatment, the error responses at baseline contained correct initial and final letters. For words that did not improve, only 46% contained correct initial and final letters prior to training. We found a similar difference for KM: 75% of the words that generalised included correct exterior letters in the misspelling at baseline. This is compared to only 52% spelling responses with correct exterior letters in words that did not generalise.

We examined the predictability of these letters. Predictability was based on the Perry list of English phoneme-to-grapheme correspondences (Perry, Ziegler & Coltheart, 2002). If the phoneme-to-grapheme correspondence for both the initial and the final letter within a word corresponded to the most frequent phoneme-to-grapheme correspondence on the Perry list, the value 1 was assigned, otherwise a value of zero was assigned. While only 57% of the exterior letters were predictable in the words that did not generalise, 83% of the exterior letters were predictable for the words that generalised. At the same time, most of the predictable exterior letters were spelled correctly prior to treatment in both the words that generalised (85%) and those that did not generalise (100%). Thus, it seems likely that for KM, whether or not exterior letters were spelled correctly was a function of phoneme-to-grapheme predictability – even before treatment.

Predictability of phoneme-to-grapheme correspondences in generalised words. Given that words which generalised tended to have predictable exterior letters, we investigated whether words that generalised tended to be more predictable overall. In order to investigate whether the overall spelling predictability of a word influenced generalisation, we employed a measure of spelling regularity that was similar to that used by Sage and Ellis (2004). For each phoneme of a word, we checked whether it was represented by the most frequent phoneme-to-grapheme correspondence on the

Perry list (Perry et al., 2002). If a phoneme was represented by its most frequent phoneme-to-grapheme correspondence, that phoneme was given the score 1, otherwise 0. The scores for all phonemes making up a word were then added up and divided by the total number of phonemes in that word. In this system, the word ‘pit’ would get an overall spelling regularity score of 3/3, the word ‘light’ has a score of 2/3. There was no significant difference in spelling regularity scores between words that generalised (mean= .691) and those that did not (mean= .695) in study 1 (two sample t test t(174) = .858 p = .392) or study 2 (t(102) = 0.08 p = .937) (generalised words mean= .678; words that did not generalise mean= .679). This means that the words that KM spelled correctly without specific training were no more predictable overall than the words that did not generalise.

Subregularities. In addition, we analysed whether KM showed generalisation to those words that had similar ‘subregularities’ to the training words. For example, was KM more likely to spell words like <WHOSE> and <WHOM> correctly after training the word <WHO>? If words that generalised contained ‘subregularities’ found in the training words, it could be argued that generalisation was mediated by the sublexical route, i.e. KM acquired these ‘subregularities’ and applied them to other words. Firstly, we analysed which ‘subregularities’ were included in the training words. We then compared how many of these ‘subregularities’ were also included in words that generalised versus those that did not generalise. Words that generalised did not include significantly more of the ‘subregularities’ compared to words that did not generalise (Study 1, Fisher exact z = -.145, p = 1.0; Study 2, z = 1.34 p = .181; see Table 7).

Table 7 about here

In summary, the above sections show that spelling predictability was not a factor that determined generalisation in this study. Thus, generalisation was not mediated by the sublexical spelling route but rather by the lexical spelling route. We will move on to analyse lexical features of the words that generalised.

Neighbourhood size and orthographic similarity

Neighbourhood size was one of the best predictors of generalisation in the present study. Effects of neighbourhood size have been relatively unexplored in treatment studies. However, Sage and Ellis (2006) have recently shown that words with many neighbours can be more resilient to buffer damage and that untreated neighbours of treated words can improve as a result of spelling intervention. To explain better spelling performance for words with many neighbours, Sage and Ellis assumed that there is feedback of activation between the graphemic buffer and the orthographic lexicon (see Folk, Rapp, & Goldrick, 2002; McCloskey et al., 2006; Miceli & Capasso, 2006). They argue that this interactivity led to retention in the buffer being stabilised by additional lexical support from the neighbours of a word. Sage and Ellis (2006) argue that it is the bi-directional flow of activation that led to the spelling improvement for untrained words that were neighbours of training words. While Sage and Ellis do not suggest changes in the lexicon itself, they argue that the additional activation from word neighbours between the lexicon and the buffer assisted their patient's performance on the untrained neighbours. For KM we could argue the following: Strengthening of untrained lexical nodes and their links to the grapheme level might have occurred due to the feedback from the grapheme level to the lexicon. Since more than one lexical node is connected to each letter node, any activated letter will feedback its activation to all those lexical nodes that contain that letter (see Figure 2). Hence the activation of a training word during treatment will result in activation of lexical nodes corresponding to words that share letters with the training word. That is, neighbours of the training word will be activated in the lexicon. The control words in the present study were not chosen to be neighbours of the training words. However, if the above explanation is correct, words that generalised should be orthographically more similar to the training words than those that did not generalise.

In order to test this hypothesis, we measured the degree of orthographic similarity between the untrained words and the trained words[4]. This was achieved by comparing each untrained word to every training word using the five match schemes. There were 42 training words and five match schemes. Thus, for every untrained word, there were 42 x 5 orthographic similarity scores. For example, the degree of orthographic similarity between the untrained word <SPREAD> and the trained word <SPONGE> was .2 (using match 1), .4 (using match 2), .4 (using match 3), .4 (using match 4) and .38 (using match 5). For every untrained word we summed the 42 similarity scores for every match score, so that we had five summed similarity scores for each untrained word. The summed orthographic similarity scores of the generalised words were compared to the summed orthographic similarity scores of words that did not generalise. The scores measuring similarity between the trained and the untrained words will be referred to as orthographic similarity scores1-5.

Orthographic similarity between training words and untrained words in Study 1. The words that generalised at post test 1 were orthographically significantly more similar to the training words than the words that did not improve without training on 'orthographic similarity score 1' (words that generalised: mean = 3.92; words that did

---

[4] This is a different measure from estimating whether untrained words included the same 'subregularities' as trained words. For the 'subregularities' we analysed which parts of the training words were not the most frequent phoneme-grapheme correspondence according to the Perry list, e.g. spelling the sound /S/ as <CH> in 'chef'. We then checked which of these 'subregularities' were present in the control words.

not generalise mean = 3.37) (match 1, independent samples t-test: t = -2.54 p = .015). Match scheme 1, which was used to compute 'orthographic similarity score 1' computes orthographic similarity by aligning two items at the leftmost letters. There was no significant difference on any of the other similarity scores ('orthographic similarity score 2': t = -1.46 p = .142; 'orthographic similarity score 3': t = -1.39 p = .168; 'orthographic similarity score 4': t = -1.67 p = .288; 'orthographic similarity score 5': t = .610 p = .542).

Orthographic similarity between training words and untrained words as measured on 'similarity score 1' also correlated significantly with generalisation at post test 1 (Pearson's correlation: .196 p = .015). No other orthographic similarity score correlated significantly with generalisation at post test 1.

In subsequent regression analyses including 'orthographic similarity score 1' and the variables that best predicted generalisation at post test 1 (i.e. written frequency, neighbourhood size, spelling error vs. target spelling similarity as measured by match 3), orthographic similarity between trained and untrained words was not a significant predictor of generalisation at post test 1. Written frequency (B = .533 S.E. = .233 p = .022), neighbourhood size (B = .090 S.E. = .041 p = .030) and match 3 (B = 2.195 S.E. = 1.106 p = .047) were significant predictors.

The same analyses were performed for words that generalised at follow up 1. Words that generalised at follow up 1 were significantly more orthographically similar to training words than words that did not generalise on orthographic similarity scores 1, 2, 3 and 4 ('orthographic similarity score 1', independent samples t-test: t = -3.55 p = .001; 'orthographic similarity score 2': t = -3.95 p < .001; 'orthographic similarity score 3': t = -3.71 p < .001; 'orthographic similarity score 4': t = -3.36 p = .001;) but not on 'orthographic similarity score 5' (t = -1.23 p = .221). Orthographic similarity between training words and untrained words as measured by similarity scores 1, 2, 3 and 4 also correlated significantly with generalisation at follow up 1 (Pearson's correlations between .264 and .306). We then conducted a series of regression analyses including all variables that correlated significantly with generalisation at follow up 1 (i.e. written frequency, neighbourhood size, spelling error vs. target spelling similarity as measured by match schemes 4 and 5 and the composite match score) and the orthographic similarity between trained and untrained words as measured by similarity scores 1, 2, 3 and 4 in turn. Significant predictors in all of the analyses were written frequency, neighbourhood size and spelling error vs. target spelling similarity as measured by match 5 remained. Orthographic similarity scores 2 (B = .480 S.E. = .155 p = .002), 3 (B = .417 S.E. = .150 p = .005) and 4 (B = .344 S.E. = .164 p = .036) were also significant predictors.

In study 2, there was no significant difference between words that generalised at post test 2 or at follow up 2 and words that did not generalise in terms of orthographic similarity to the training words. Orthographic similarity between untrained words and trained words did not correlate with generalisation at post test 2 or follow up 2.

In summary, the analyses performed (especially for follow up 1) suggest that orthographic similarity of the untrained words to the training words was a factor in generalisation. It thus seems as if orthographic similarity between training words and untrained words and the size of a word's neighbourhood size both play a role in generalisation.

General Discussion

Overview of the results

This treatment study investigated the nature of treatment generalisation in a girl with lexical spelling deficits. We trained irregular word spelling and found improvement for spelling of trained and untrained irregular words. Given that spelling performance of all words was stable at baseline, the improvement found after training can be confidently argued to be a result of the training provided. Supplementary analyses determined that improvement of untrained words was not simply a function of applying the low frequency phoneme-to-grapheme correspondences or 'subregularities' of the training words to untrained words, rather it was mediated by the orthographic lexicon. Untrained words were more likely to improve if they had many orthographic neighbours and occurred frequently in written and spoken language. Another characteristic common to untrained words that generalised was that they were more orthographically similar to the training words than words that did not generalise.

Improvement of training words

Training led to improved spelling of the training words. What change in processing of these words underlies these improvements? For this discussion, we will assume that a lexical entry consists of a lexical node in the orthographic lexicon pointing to letters at the grapheme level, and that there are bi-directional links between the lexicon and the grapheme level (see Figure 2, earlier). We will discuss each element of the training procedure (spelling to dictation, direct copying and delayed copying) in turn and consider the processing involved.

Delayed and direct copying. During copying and delayed copying, KM saw the training word. Seeing the word would have resulted in the activation of abstract letter identities at the input level which in turn activated the incomplete or weak lexical representation in the orthographic lexicon and the letters at the grapheme level. During delayed copying, the letters making up a training word had to be held active at the grapheme level while the word was written. Due to bi-directional flow of activation between these levels of processing, activation cycles between the orthographic lexicon and the graphemic buffer (see Folk et al., 2002; McCloskey et al., 2006;

Miceli & Capasso, 2006). Thus, whilst letters are held active at the grapheme level, the corresponding lexical node for that training item would continue to receive activation from the grapheme level. This extended cycling of activation could potentially improve lexical representations by strengthening connections between the orthographic lexicon and the grapheme level.

In addition, letters that were not previously part of the incomplete or maybe incorrect lexical representation would have been activated via a direct link between the representation of a letter at the input level to the graphemic buffer (Brunsdon, Coltheart, & Nickels, 2006). Furthermore, without having been prompted to do so, KM sometimes rehearsed letter names of the training words. In this case, letter names would be activated via the phonological lexicon (Brunsdon et al., 2006) and these letter names would in turn activate letters at the level of the graphemic buffer. Hence, both of these processes could have led to the formation of connections between the orthographic lexicon and the grapheme level for incomplete lexical entries. This could also be the mechanism by which new lexical nodes were formed in the lexicon along with their corresponding connections to the grapheme level.

In summary, we suggest that copying and delayed copying could have improved spelling of training words by strengthening, completion and new formation of lexical entries.

Spelling to dictation. During spelling to dictation, the phonological representation of a training word is activated and in turn, activates its corresponding orthographic representation in the orthographic lexicon. Most models make the assumption that connections between phonological representations and orthographic representations are word specific, or ‘one-to-one’ (e.g. Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). When spelling to dictation, for example the word ‘spread’, the orthographic entry <SPREAD> will receive activation from the phonological lexicon. Repeated activation of the orthographic representation from the phonological representation could lead to strengthening of that specific connection for trained words.

Mechanisms underlying generalisation

In this section we will explore possible mechanisms underlying generalisation. We start by looking at previous suggestions regarding mechanisms mediating treatment generalisation.

Semantic spelling route. In line with previous research we found no improvement for the spelling of untrained homophones (Behrmann, 1987; Brunsdon et al., 2005). This is taken as evidence that processing along the semantic spelling route did not improve as a result of training. This was expected as the training did not focus on semantic processes. Hence, treatment generalisation in our case was not mediated by the semantic spelling route.

Improved access to orthography. Under the assumption of word specific one-to-one connections between the phonological lexicon to the orthographic lexicon, it seems unlikely that generalisation could be due to generally enhanced orthographic access from the phonological lexicon as only the links for trained words would be strengthened during training. This is in contrast to Weekes and Coltheart (1996) and Brunsdon et al. (2005) who hypothesized that treatment generalisation arises from improved access to orthography.

Orthographic similarity. Our data indicated that orthographic similarity is relevant in treatment generalisation. The similarity of untrained words to the training words played a role in predicting generalisation. This is explicable by considering the feedback of activation from the grapheme level to the orthographic lexicon. This feedback not only leads to activation of the training words but also to the partial activation of words that are orthographically similar to the training words: their neighbours (see also Sage & Ellis, 2006) and words that are orthographically related.

These findings lead to quite clear predictions. For example, the likelihood of treatment generalisation may be a function of orthographic similarity between untreated words and training words[5]. Whether this is indeed the case could be empirically tested in future research.

Neighbourhood size. In the previous section we argued that words benefit from their orthographic similarity to the training words because they are partially activated during training. This can result in long term strengthening of links between the grapheme level and the lexicon for these entries. Even if words are not direct neighbours of training words, having many neighbours means that the chance of partial activation for this lexical representation is higher than for a word with few or no neighbours. Words with a small neighbourhood size do not benefit in the same way from the feedback from the graphemic buffer. We would argue that the reason why neighbourhood size is such a good predictor of generalisation is because it is an estimate of the number of possible occasions (e.g. during training and regular school activities) on which the lexical representation receives activation, with each occasion leading to strengthening of that representation.

---

[5] Note that phonological neighbours of training words may receive some activation which then cascades to the orthographic lexicon, possibly leading to improved access for phonological neighbours of training words as well. The authors would like to thank M.J. Tainturier for this suggestion.

Frequency of occurrence. Another significant predictor of generalisation in this study was frequency of occurrence. Sage and Ellis (2004) mention the tradition in modelling of language processing to reflect frequency effects in varying strengths of connections between the semantic system and the (orthographic) lexicon. Under this view, what might make high frequency words more resistant to the pathological loss of information in the buffer is that they receive stronger top down support. Nevertheless, it is hard to see how the differences in connection strengths between the semantic system and the orthographic lexicon could explain why high frequency words were more likely to improve without specific training.

However, there are other ways of implementing frequency within cognitive models of language processing that may be more suitable for the present study. Within the DRC model (Coltheart et al., 2001), frequency is implemented as a constant associated with each lexical node. The value of the constant is derived from the frequency of the word and determines how quickly activation rises for a particular node, with the activation of higher frequency nodes rising quicker than low frequency nodes. As outlined above, lexical nodes corresponding to untreated items are activated via the feedback from the graphemic buffer to the orthographic lexicon. With frequency determining activation rise time, those words of higher frequency will have a quicker rise in activation. This might then lead to greater strengthening of the connections between the lexicon and the grapheme level. A possible effect of training therefore may be comparable to a priming effect, where training leads to a long lasting strengthening of connections from the orthographic lexicon to the grapheme level.

However, frequency of occurrence may just be a proxy for the quality of representation, where a word's frequency is positively correlated to its quality of representation. KM would have had more exposure to higher frequency words and thus more occasions on which she had the opportunity to acquire information about these words. Support for this hypothesis is the fact that written frequency and the number of times a word was spelled correctly prior to training were significantly correlated.

Quality of a lexical representation. Brunsdon et al. (2005) used the similarity of pre-treatment errors to target spellings (as measured by the five match values) as an indication of the quality of the underlying lexical representation. The present study showed clearly that high values on the five match schemes when comparing pre-treatment errors to target spellings were not of any further predictive value for generalisation over and above neighbourhood size and frequency. However, the quality of lexical representations is still very likely to play a role in generalisation. It seems unlikely that a word with an incomplete or non-existent lexical representation could be spelled correctly without specific training. Rather, the words that benefit from treatment generalisation may have complete but weak representations, which could then be strengthened as a result of the feedback from the graphemic buffer.

Thus, in order to maximise treatment effects it seems necessary to find a way of measuring the quality of lexical representations. Quantifying the quality of an orthographic representation is not an easy undertaking. Like Brunsdon et al. (2005), we used the similarity of spelling errors to their target spelling as a measure of the quality of the lexical representation. Given that we were dealing with words whose spelling is not perfectly predictable, it seems a reasonable assumption that if a word has been spelled correctly at least once before treatment, there was a complete representation in the orthographic lexicon for this word. Whether or not a word has been spelled correctly during baseline assessments could thus be another proxy of the quality of its lexical representation along with the similarity of the pre-treatment error to the target spelling. Indeed, we found significant correlations between the similarity of pre-treatment errors to target spellings and the number of times a word was spelled correctly prior to training (study 1 match 4, Pearson correlation .222 p = .003; match 5: .198 p = .008; study 2 match 1: .168 p = .049, match 4: .220 p = .009; match 5: .202 p = .018).

Future studies may determine whether combining the two (i.e. the number of times an item is spelled correctly at baseline and the similarity of the spelling error to the target spelling) provides a better measure.

Clinical implications

The results from this study have clear clinical implications regarding the selection of training words[6]. Overall, it seems that to maximise treatment outcomes, treatment should be focused on words of lower neighbourhood size and lower frequency that were never spelled correctly. Future research may determine what the limits of this claim are; for example, if none of the training words have neighbours and are orthographically not similar to any other words, they would be very unlikely to induce generalisation. Words with a high neighbourhood size and high frequency that were spelled correctly at least once during multiple baseline assessment are most likely to show generalisation and hence could be left untreated. (Of course. whether or not generalisation occurs to untreated words should be carefully monitored).

Summary and conclusions

Generalisation to untreated items in spelling is localised in the feedback of activation from the graphemic buffer to the orthographic lexicon. In the light of the results of this study we would argue that the mechanisms underlying generalisation are very similar

[6] Note that some of the clinical decisions and implications regarding both sublexical and lexical treatment are discussed elsewhere (Kohnen & Nickels, in preparation).

to priming. Consequently, generalisation is likely to only occur for words that have complete (but weak) lexical representations.

Essentially, feedback from the graphemic buffer to the orthographic lexicon leads to repeated activation of lexical nodes that are orthographically similar to the training words and/or that are orthographically similar to many words overall. Repeated activation results in a long term increase in connection strength of the links between the lexicon and the letters at the grapheme level for a particular representation. Stronger connections increase the chance of a correct response.

In summary, this study documents a successful treatment study with a child with lexical spelling deficits based on cognitive neuropsychological methods and models. The treatment was effective, as well as long lasting, and results could be explained within the model underlying this research. We hope that the findings may inspire future research as well as guide clinicians in their treatment of cases with similar difficulties.

References

Adams, C., Cooke, R., Crutchley, A., Hesketh, A., & Reeves, D. (2001). *Assessment of Comprehension and Expression 6-11 (ACE6-11).* Windsor: Nelson Publishing Company Ltd.

Aliminosa, D., McCloskey, M., Goodman-Schulman, R., & Sokol, S. M. (1993). Remediation of acquired dysgraphia as a technique for testing interpretations of deficits. *Aphasiology, 7*(1), 55-69.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX database (CD-ROM).* Philadelphia: Linguistic Data Consortium, University of California.

Badecker, W., Hillis, A., & Caramazza, A. (1990). Lexical morphology and its role in the writing process: Evidence from a case of acquired dysgraphia. *Cognition, 35,* 205-243.

Barry, C. (1994). Spelling routes (or roots or rutes). In G. D. A. Brown & N. C. Ellis (Eds.), *Handbook of Spelling: Theory, Process and Intervention* (pp. 27-50). Chichester: John Wiley & Sons Ltd.

Beauvois, M.-F., & Derouesné, J. (1981). Lexical or orthographic dysgraphia. *Brain, 104*, 21-50.

Beeson, P. M. (1999). Treating acquired writing impairment: Strengthening graphemic representations. *Aphasiology, 19*(9-11), 767-785.

Beeson, P. M., Hirsch, F. M., & Rewega, M. A. (2002). Successful single-word writing treatment: Experimental analyses of four cases. *Aphasiology, 16(*4/5/6), 473-491.

Behrmann, M. (1987). The rites of righting writing: Homophone remediation in acquired dysgraphia. *Cognitive Neuropsychology, 4*(3), 365-384.

Behrmann, M., & Bub, D. (1992). Surface dyslexia and dysgraphia: Dual routes, single lexicon. *Cognitive Neuropsychology, 9*(3), 209-251.

Brunsdon, R., Coltheart, M., & Nickels, L. (2005). Treatment of irregular word spelling in developmental surface dysgraphia. *Cognitive Neuropsychology, 22*(2), 313-351.

Brunsdon, R., Coltheart, M., & Nickels, L. (2006). Severe developmental letter-processing impairment: A treatment case study. *Cognitive Neuropsychology, 23*(6), 795-821.

Brunsdon, R. K., Hannan, T. J., Coltheart, M., & Nickels, L. (2002). Treatment in lexical processing in mixed dyslexia: A case study. *Neuropsychological Rehabilitation, 12*(5), 385-418.

Caramazza, A., & Miceli, G. (1990). The structure of graphemic representations. *Cognition, 37*, 243-297.

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

Coltheart, M., & Leahy, J. (1996). Assessment of nonlexical reading abilities in children: Some normative data. *Australian Journal of Psychology, 48*(3), 136-140.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia, 40*, 1196-1208.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist, 12*(4), 482-486.

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.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104,* 801-838.

De Partz, M.-P., Seron, X., & Van der Linden, M. (1992). Re-education of a surface dysgraphia with a visual imagery strategy. *Cognitive Neuropsychology, 9*(5), 369-401.

Dunn, L. M., & Dunn, L. M. (1997). *The Peabody Picture Vocabulary Test – 3rd edition (PPVT-III).* Circle Pines, MN: American Guidance Service.

Edwards, V. T., & Hogben, J. H. (1999). New norms comparing children's lexical and nonlexical reading: A further look at subtyping dyslexia. *Australian Journal of Psychology, 51*(1), 37-49.

Elliot, C. D. (1990). *Differential Ability Scales (DAS).* San Antonio: The Psychological Corporation.

Folk, J. R., Rapp, B., & Goldrick, M. (2002). The interaction of lexical and sublexical information in spelling: What's the point? *Cognitive Neuropsychology, 19*(7), 653-671.

Franklin, S. (1997). Designing single case treatment studies for aphasic patients. *Neuropsychological Rehabilitation, 7*(4), 401-418.

Gathercole, S., & Baddeley, A. (1996). *The Children's Test of Nonword Repetition.* London: The Psychological Corporation.

Goodman, R. A., & Caramazza, A. (1986). Aspects of the spelling process: Evidence from a case of acquired dysgraphia. *Language and Cognitive Processes, 1,* 263-296.

Hillis, A. E. (1991). Effects of separate treatments for distinct impairments within the naming process. In T. Prescott (Ed.), *Clinical Aphasiology* (Vol. 19, pp. 255-265). Austin, TX: Pro-Ed.

Hillis, A. E. (1993). The role of models of language processing in rehabilitation of language impairments. Aphasiology, 7(1), 5-26.

Hillis, A. E., & Caramazza, A. (1987). Model-driven remediation of dysgraphia. In R. H. Brookshire (Ed.), *Clinical Aphasiology* (Vol. 17, pp. 84-105). Minneapolis: BRK Publishers.

Howard, D. (2004). *Statistics Workbook: Unpublished Excel Workbook*.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic assessments of language processing in aphasia (PALPA).* Sussex, UK: Lawrence Erlbaum Ltd.

Kohnen, S. & Nickels, L. (in preparation). Clinical decisions in treating developmental mixed dysgraphia.

Kohnen, S., Nickels, L., Brunsdon, R., & Coltheart, M. (2008). Patterns of generalisation after treating sub-lexical spelling deficits in a child with mixed dysgraphia. *Journal of Research in Reading, 31*(1), 157-177.

McCloskey, M., Badecker, W., Goodman-Schulman, R. A., & Aliminosa, D. (1994). The structure of graphemic representations in spelling: Evidence from a case of acquired dysgraphia. *Cognitive Neuropsychology, 11*(3), 341-392.

McCloskey, M., Macaruso, P., & Rapp, B. (2006). Grapheme-to-lexeme feedback in the spelling system: Evidence from a dysgraphic patient. *Cognitive Neuropsychology, 23*(2), 278-307.

Miceli, G., & Capasso, R. (2006). Spelling and dysgraphia. *Cognitive Neuropsychology, 23*(1), 110-134.

Miceli, G., Silveri, M. C., & Caramazza, A. (1987). The role of phoneme-to-grapheme conversion system and of the graphemic output buffer in writing. In M. Coltheart, G. Sartori & R. Job (Eds.), *The Cognitive Neuropsychology of Language* (pp. 235-252). London: Lawrence Erlbaum Associates Ltd.

Patterson, K., & Shewell, C. (1987). Speak and spell: Dissociations and word-class effects. In M. Coltheart, G. Sartori & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 273-294). London: LEA.

Perry, C., Ziegler, J. C., & Coltheart, M. (2002). How predictable is spelling? Developing and testing metrics of phoneme-grapheme contingency. *The Quarterly Journal of Experimental Psychology, 55A*(3), 897-915.

Rapp, B. (2005). The relationship between treatment outcomes and the underlying cognitive deficit: Evidence from the remediation of acquired dysgraphia. *Aphasiology, 19*(10/11), 994-1008.

Rapp, B., & Kane, A. (2002). Remediation of deficits affecting different components of the spelling process. *Aphasiology, 16*(4/5/6), 439-454.

Rastle, K., & Coltheart, M. (1999). Lexical and nonlexical phonological priming in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 461-481.

Raymer, A. M., Cudworth, C., & Haley, M. A. (2003). Spelling treatment for an individual with dysgraphia: Analysis of generalisation to untrained words. *Aphasiology, 17*(6/7), 607-624.

Robinson, G., & Weekes, B. (1995). Subtypes in developmental dysgraphia. In J. Fourez & N. Page (Eds.), *Treatment Issues and Long Term Outcomes* (pp. 43-51). Brisbane: Academic Press.

Romani, C., Ward, J., & Olson, A. (1999). Developmental surface dysgraphia: What is the underlying cognitive impairment? *The Quarterly Journal of Experimental Psychology, 52A*(1), 97-128.

Sage, K., & Ellis, A. W. (2004). Lexical influences in graphemic buffer disorder. *Cognitive Neuropsychology, 21*(2/3/4), 381-400.

Sage, K., & Ellis, A. W. (2006). Using orthographic neighbours to treat a case of graphemic buffer disorder. *Aphasiology, 20*(9/10/11), 851-870.

Schmalzl, L., & Nickels, L. (2006). Treatment of irregular word spelling in acquired dysgraphia: Selective benefit from visual mnemonics. *Neuropsychological Rehabilitation, 16*(1), 1-37.

Scott, C. & Byng, S. (1989). Computer assisted remediation of a homophone comprehension disorder in surface dyslexia. *Aphasiology, 3*(3), 301-320.

Stadie, N., & van de Vijver, R. (2003). A linguistic and neuropsychological approach to remediation in a German case of developmental dysgraphia. *Annals of Dyslexia, 53*, 280-299.

Tainturier, M.-J., & Rapp, B. C. (2004). Complex graphemes as functional spelling units: Evidence from acquired dysgraphia. *Neurocase, 1*(2), 122-131.

Temple, C. (1985). Developmental surface dysgraphia: A case report. *Applied Psycholinguistics, 6*, 391-406.

Temple, C. (1986). Developmental dysgraphias. *The Quarterly Journal of Experimental Psychology, 38*A, 77-110.

Temple, C. M. (1997). Spelling Disorders. In C. M. Temple (Ed.), *Developmental Cognitive Neuropsychology* (pp. 225-252). Hove: Psychology Press.

Venezky, R. L. (1970). *The Structure of English Orthography.* The Hague: Mouton.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence – 3rd edition (WPPSI-III).* New York City: Psychological Corporation.

Weekes, B., & Coltheart, M. (1996). Surface dyslexia and surface dysgraphia: Treatment studies and their theoretical implications. *Cognitive Neuropsychology, 13*(2), 277-315.

Appendix A:

Example of different ways of aligning target word and error response for Match schemes 1-4 comparing WEALTH to WELTH

Match scheme 1: Alignment from leftmost letter

Position	1	2	3	4	5	6	
	W	E	A	L	T	H	
	W	E	L	T	H		
overlap	*	*					2/6

Match scheme 2: Vowel centred and left alignment

Position	O1	O2	O3	V1	V2	V3	C1	C2	C3	
	W			E	A		L	T	H	
	W			E			L	T	H	
overlap	*			*			*	*	*	5/6

Match Scheme 3: Vowel centred and right alignment

Position	O1	O2	O3	V1	V2	V3	C1	C2	C3	
			W	E	A		L	T	H	
			W	E			L	T	H	
Overlap:			*	*			*	*	*	5/6

Match Scheme 4: External letters alignment

Position	L	L+1	L+2	L+3	F-3	F-2	F-1	F	
	W	E	A			L	T	H	
	W	E	L				T	H	
Overlap	*	*					*	*	4/6

Different Match values for selected spelling errors

Target Word	Error Response	Match Value 1	2	3	4	5
swap	swop	.75	.75	.75	.75	.64
wealth	welth	.33	.83	.83	.67	.62
thief	thef	.4	.6	.6	.8	.58
comb	come	.75	.75	.5	.75	.49
took	tock	.75	.5	.75	.75	.64

Table 1

Assessment

Test name	n =	KM score	Control Sample*/ normative score
Initial assessment			
Phonological Processing			
Auditory Rhyme Judgement (PALPA 15)	60	60	
Initial Sounds (PALPA 16)	45	43	
Final Sounds (PALPA 17)	45	44	
Before lexical treatment			
Test name	n =	KM score	Control Sample*/ normative score
Phonological Processing			
Children's Test of Nonword Repetition (CNRep)			Controls aged 8-8;11:
2 syllables	10	8	9
3 syllables	10	7	8
4 syllables	10	8	7
5 syllables	10	7	8
Naming and Semantics			
Picture Naming (ACE6-11)	20	15	16^{th} percentile
Spoken Picture Naming Test (PALPA 53)	40	38	
Written Picture Naming Test (PALPA 53)	40	19	
Peabody Picture Vocabulary Test (PPVT)		127	Standard Score 97
Spoken Word Picture Matching (PALPA 47)	40	39	

Written Word Picture Matching (PALPA 48)	40	38	
Single Letter Processing (PALPA 22)			
Naming Letters (upper case)	26	26	
Naming Letters (lower case)	26	26	
Sounding out Letters (upper case)	26	26	
Sounding out Letters (lower case)	26	26	
Spelling Letter Sounds to Dictation	25	25	
Visual Lexical Decision (PALPA 27)			
Regular Words	15	14	Mean = 14 SD = 1.09
Irregular Words	15	10	Mean = 11.5 SD = .84
Pseudohomophones	15	3	Mean = 9.33 SD = 2.07
Nonwords	15	9	Mean = 12.3 SD = 2.86
Single Word Reading (Castles & Coltheart Lists)			
Regulars	30	25	**Mean = 27.8 SD = 2.9
Irregulars	30	16	**Mean = 21.8 SD = 3.9
Nonwords	30	18	**Mean = 24.2 SD = 5.9
Single Word Spelling (Robinson & Weekes Lists)			
Regulars	30	20	Mean = 23.7 SD = 2.9
Irregulars	30	9	Mean = 16.8 SD = 3.1
Nonwords	30	21	Mean = 21.8 SD = 3.8

*control score from N = 5 age-matched control children **norms from Edwards & Hogben (1999)

Table 2

Spelling errors on irregular words at baseline 1.1 and 1.2

Error type	Definition	Example	Baseline 1.1	Baseline 1.2
Phonologically plausible spelling	Every phoneme spelled by using the most or second most frequent phoneme-to-grapheme correspondences according to the Perry List (Perry et al., 2002).	<BREATH> as <BRETH>; <MONK> as <MUNK>	43.6%	38.9%
Close phonological approximation	At least 67% of the phonemes represented by the most or second most frequent grapheme	<PLEASE> as <PLEES>; <CRINGE> as <CRING>	31.9%	35%
Errors indicating lexical knowledge	Presence of unpredictable letters in the error response	<SCISSORS> as <SCISERS>	7.5%	10.5%
Whole word substitution	Substituting one word by another	<PULL> as <POOL>; <WERE> as <WHERE>	4.8%	7.8%
Phonologically implausible nonword	Less than 67% of the phonemes represented by the most or second most frequent grapheme	<WASP> as <YOST>; <PALM> as <NAM>	12.2%	7.8%

Table 3

Reading errors on irregular words at baseline 1.1 and 1.2

Error type	Definition	Example	Baseline 1.1	Baseline 1.2
Phonologically plausible error	Every grapheme read as the phoneme it represents most frequently	SHOVE> as /shə□v/	37.6%	29.7%
Close phonological approximation	At least 67% of the graphemes read as the most or second most frequent phoneme	<GROUP> as /gr□p/	20.8%	17.8%
Whole word substitution	Substituting one word for another	<PALM> as ‘asylum’	34.7%	43.6%
Phonologically implausible nonword	Less than 67% of the graphemes are read as the most or second most frequent phoneme	<GASP> as /kæst/	6.9%	8.9%

Table 4

Time line of the study

Assessment or Training	Abbreviation	Number of Sessions	Weeks no.
Baseline 1	BL1.1	3	1 – 2
Baseline 2	BL1.2	2	3
Training 1	T1	9	5 – 7
Post Training 1	PT1	2	8
Follow Up Training 1	FU1 & BL2.1	2	18 –19
Baseline for Training 2	BL2.2	2	22
Training 2	T2	6	25 – 26
Post Training 2	PT2	2	27
Follow Up Training 2	FU2	2	38

Table 5

Correlations between words that generalised at PT1 and FU1 and predictor variables for 153 words misspelled at BL1

Variable	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Generalised at Post Test 1	**-.167***	**.210***	.153	**.205***	**.166***	.080	**.193***	**.160***	.151	**.180***	.130	-.120	**.180***
2 Generalised at Follow Up 1	**.292***	**.272***	**.219***	**.315***	.101	.100	**.214***	**.240***	**.231***	**.215***	**.165***	.017	**.198***
3 Length in Letters	-	-.139	-.076	**-.716***	-.148	-.074	-.015	**-.166***	-.105	-.128	-.038	-.056	**-.238***
4 Written Frequency		-	**.930***	.074	-.014	.094	.100	.084	.124	.096	.090	-.087	.109
5 Spoken Frequency			-	.028	-.035	.059	.047	.027	.057	.038	.032	-.075	.037
6 Neighbourhood Size				-	**.221***	**.176***	.108	.098	.070	.158	-.013	-.030	**.189***
7 Match 1					-	**.533***	**.570***	**.473***	**.518***	**.733***	.021	.155	**.499***
8 Match 2						-	**.781***	**.572***	**.697***	**.687***	**.840***	.099	**.683***
9 Match 3							-	**.633***	**.777***	**.884***	**.208***	.145	**.696***
10 Match 4								-	**.732***	**.814***	**.600***	**.261***	**.721***
11 Match 5									-	**.912***	**.468***	**.162***	**.789***
12 Composite Match Score										-	**.375***	**.197***	**.813***

13 External Letters Correct	-	**.235***	**.327***
14 Phonolog. Plausible Error		-	.113
15 Similarity 0-2			-

*significant at p<.05

Table 6

Means for item groups in study 2

	Low value training set	Low value control set	High value N, Fq set	High value N, Fq and matches set
Written log frequency	2.21	2.16	2.93	2.98
Spoken log frequency	.72	.66	1.56	1.58
N	3.00	2.97	5.15	5.06
Match 3	.64	.64	.63	.74
Match 4	.59	.56	.58	.76
Match 5	.70	.67	.68	.76

Table 7:

Subregularities

Proportion of generalised words that contained subregularities present in trained words vs proportion of generalised words that did not contain subregularities present in trained words for studies 1 and 2

		Untrained words including subregularities	Untrained words without subregularities
Study 1	Proportion generalised*	33%	32%
	n=	86	68
Study 2	Proportion generalised*	49%	32%
	n=	43	46

*generalised=words incorrect at first baseline and correct at post test

Figure Caption

Figure 1. Schematic representation of the functional architecture of spelling

Figure 2. Bi-directional activation between the grapheme level and the orthographic lexicon (e.g. <SWAMP>, <SWAP>, <SWARM>)

Figure 3. Spelling Results Training 1.

Figure 4. Reading Results Training 1.

Figure 5. Item Structure for Training 2.

Figure 6. Spelling Results Training 2.

Figure 7. Reading Results Training 2.

Figure 8. Results Homophone Spelling.

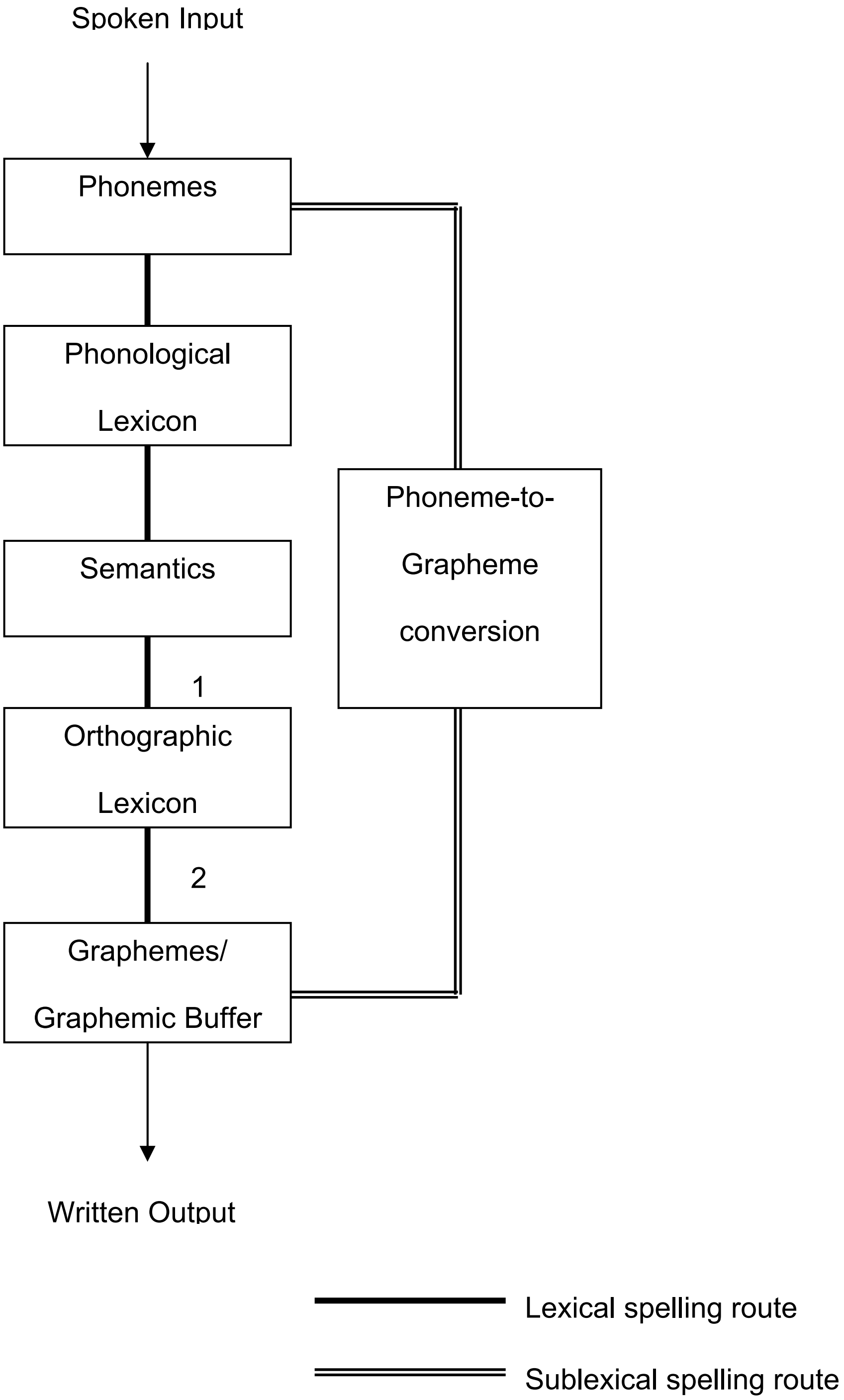
Spoken Input
Phonemes
Phonological
Lexicon
Semantics
1
Orthographic
Lexicon
2
Graphemes/
Graphemic Buffer
Written Output
Phoneme-to-
Grapheme
conversion
Lexical spelling route
Sublexical spelling route

Figure 1

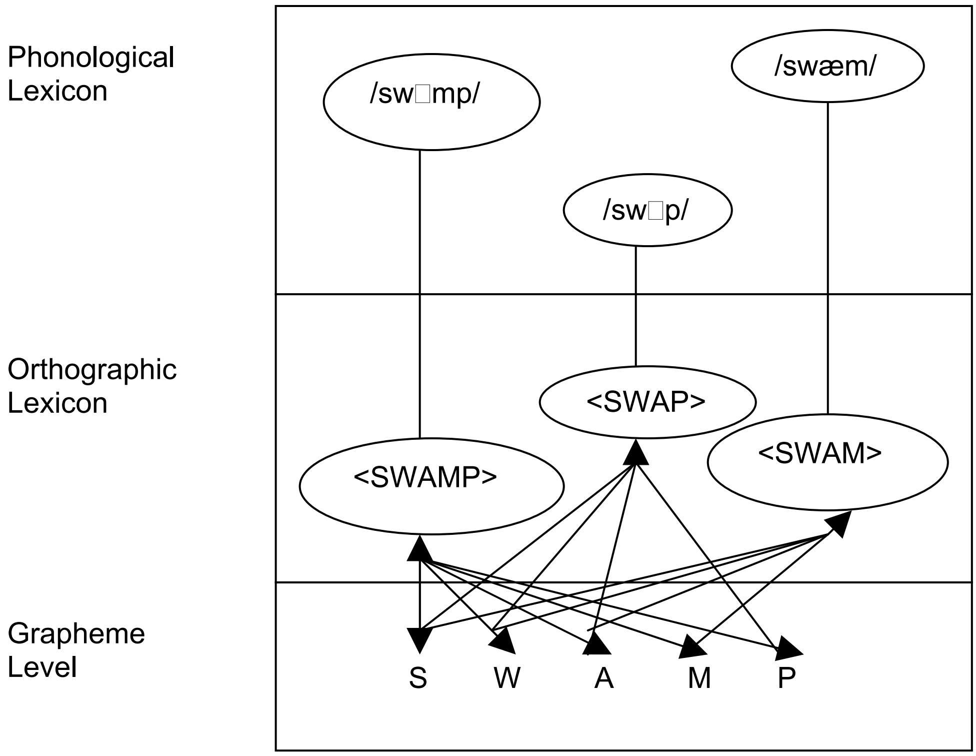
Phonological
Lexicon
/sw□mp/
/swæm/
/sw□p/
Orthographic
Lexicon
<SWAP>
<SWAMP>
<SWAM>
Grapheme
Level
S
W
A
M
P

Figure 2

Generalisation irregular word spelling

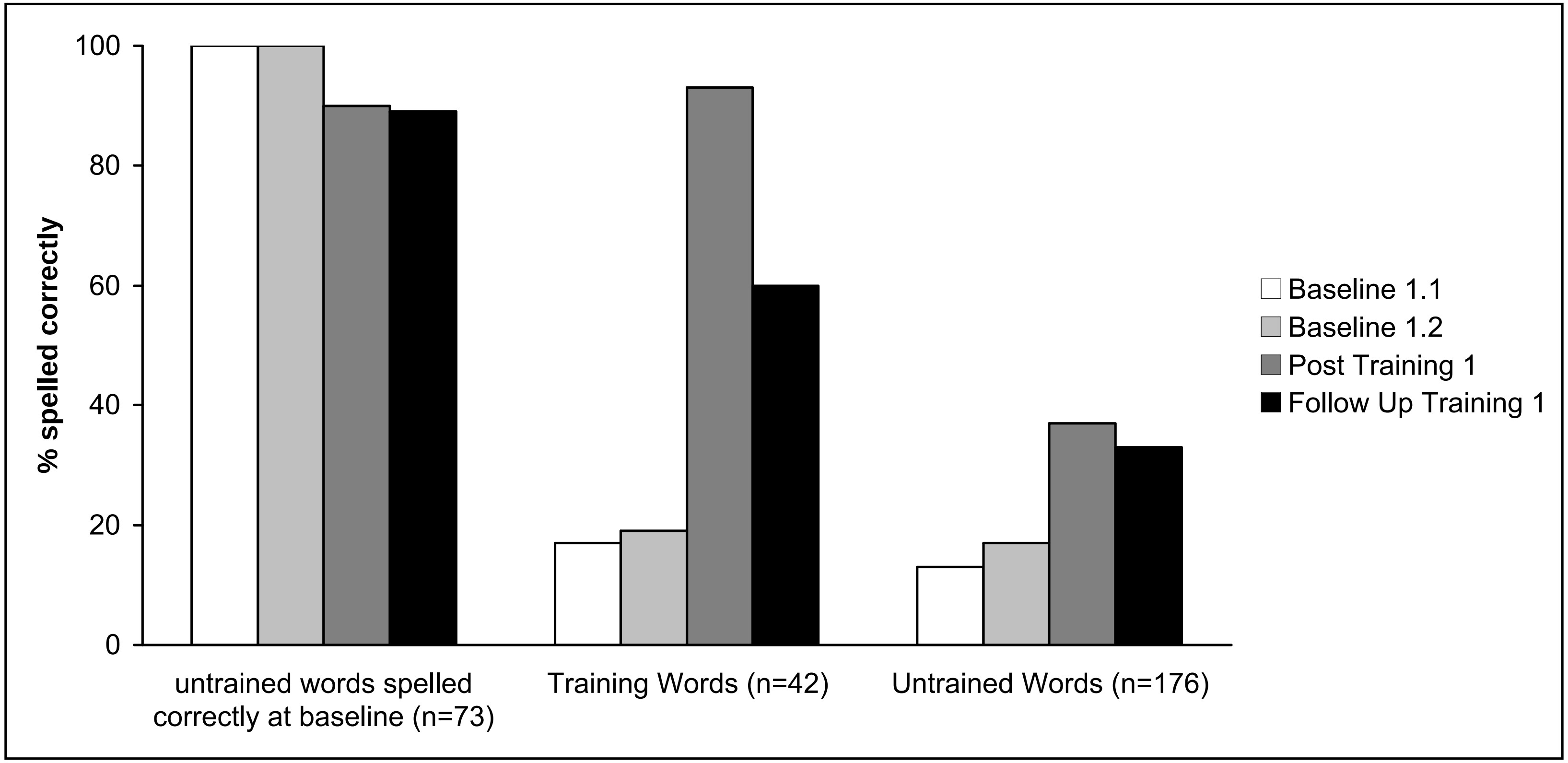

Figure 3

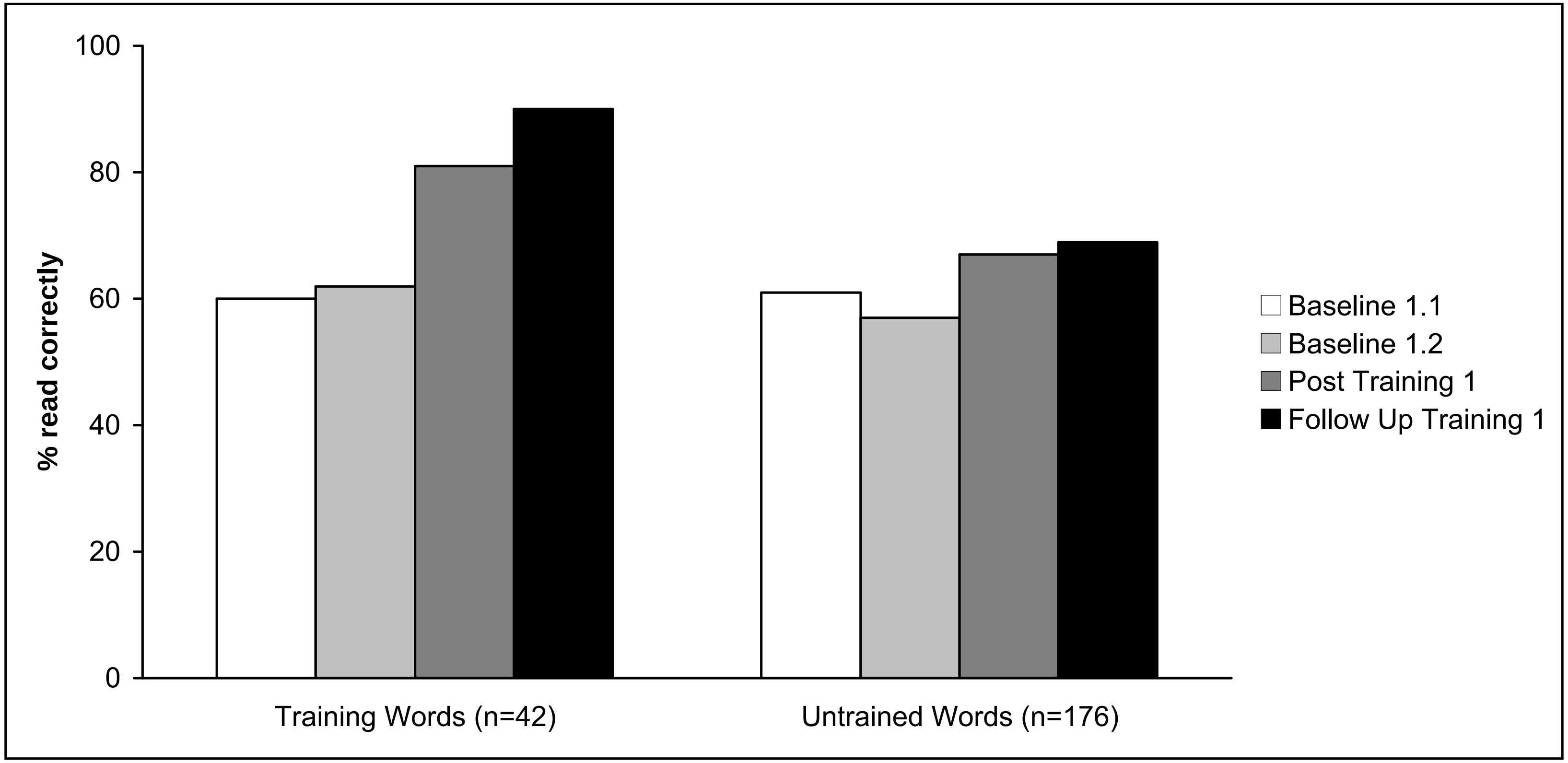
100
80
60
40
20
0
% read correctly
Training Words (n=42)
Untrained Words (n=176)
Baseline 1.1
Baseline 1.2
Post Training 1
Follow Up Training 1

Figure 4

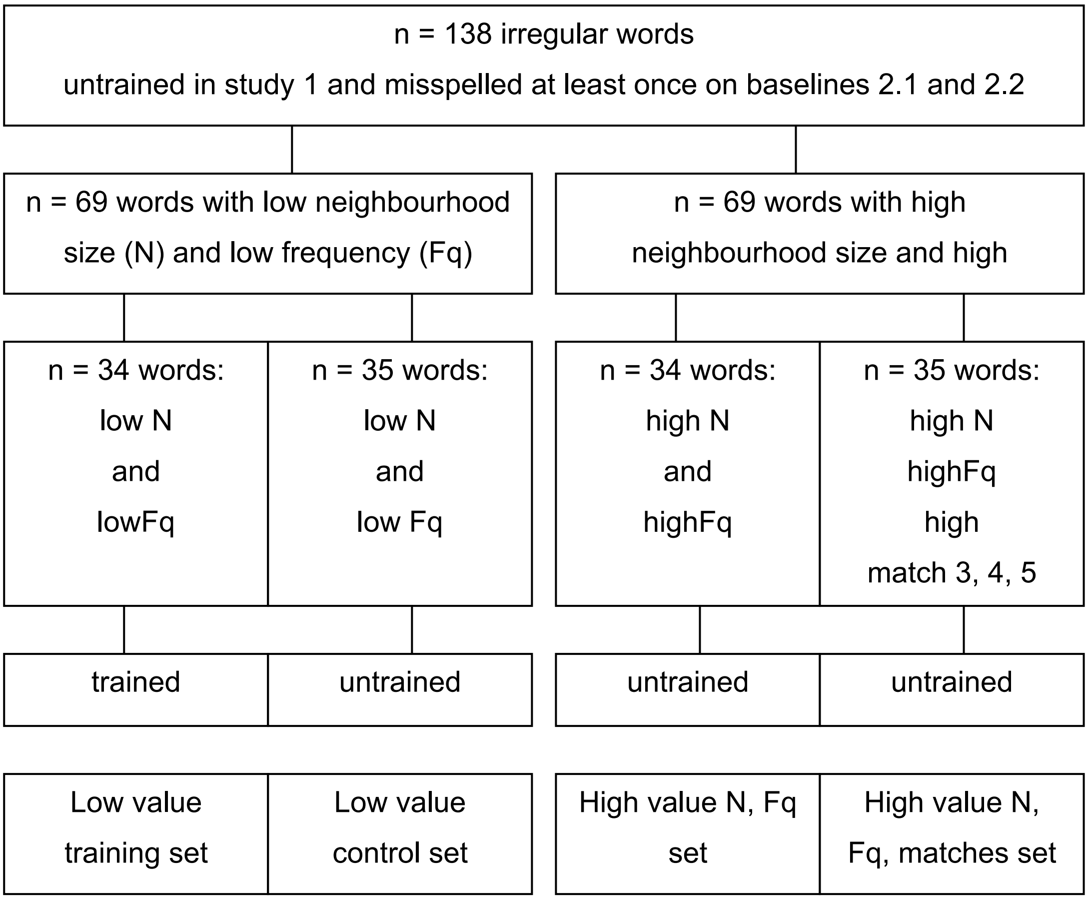
n = 138 irregular words
untrained in study 1 and misspelled at least once on baselines 2.1 and 2.2
n = 69 words with low neighbourhood size (N) and low frequency (Fq)
n = 69 words with high neighbourhood size and high
n = 34 words: low N and lowFq
n = 35 words: low N and low Fq
n = 34 words: high N and highFq
n = 35 words: high N highFq high match 3, 4, 5
trained
untrained
untrained
untrained
Low value training set
Low value control set
High value N, Fq set
High value N, Fq, matches set

Figure 5

Generalisation irregular word spelling

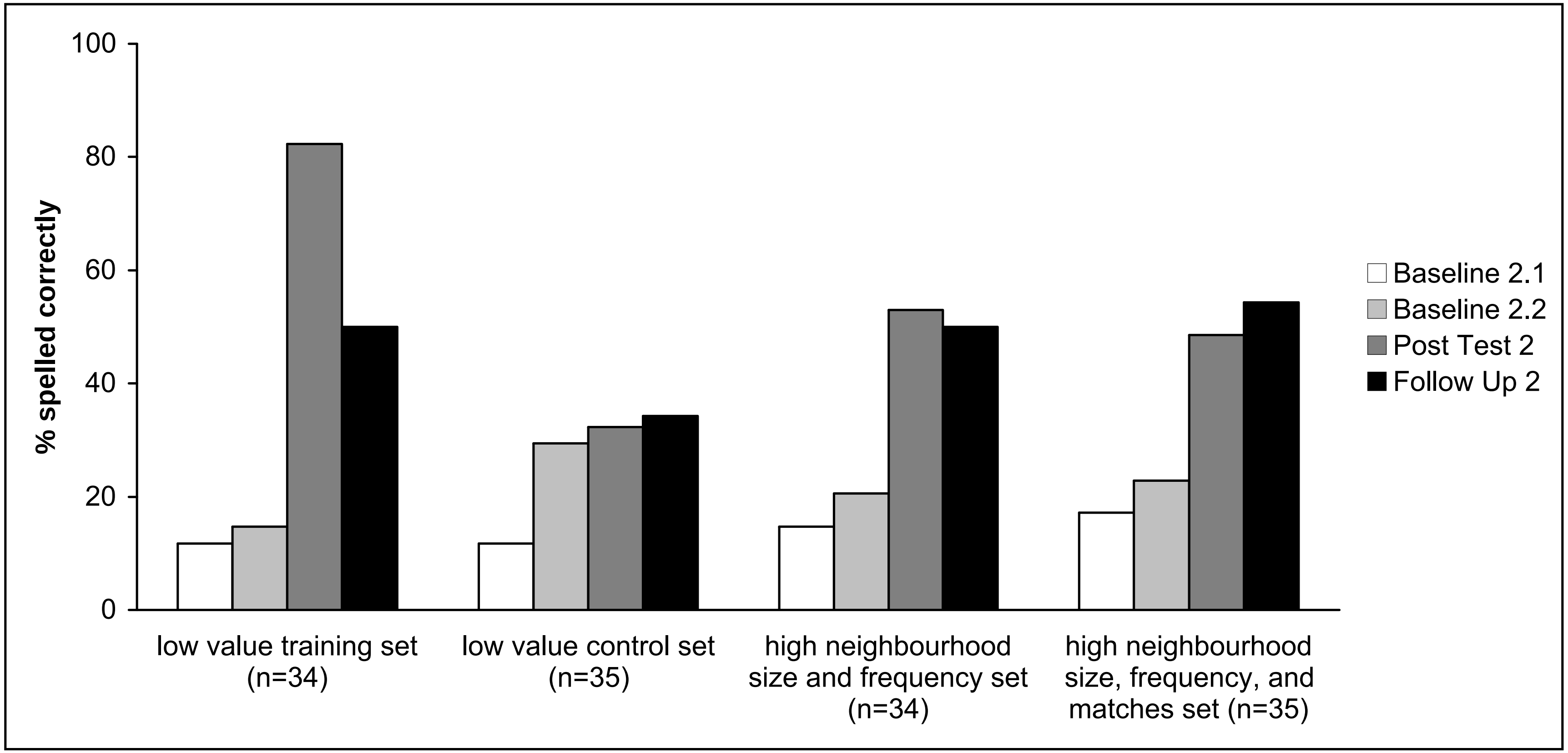

Figure 6

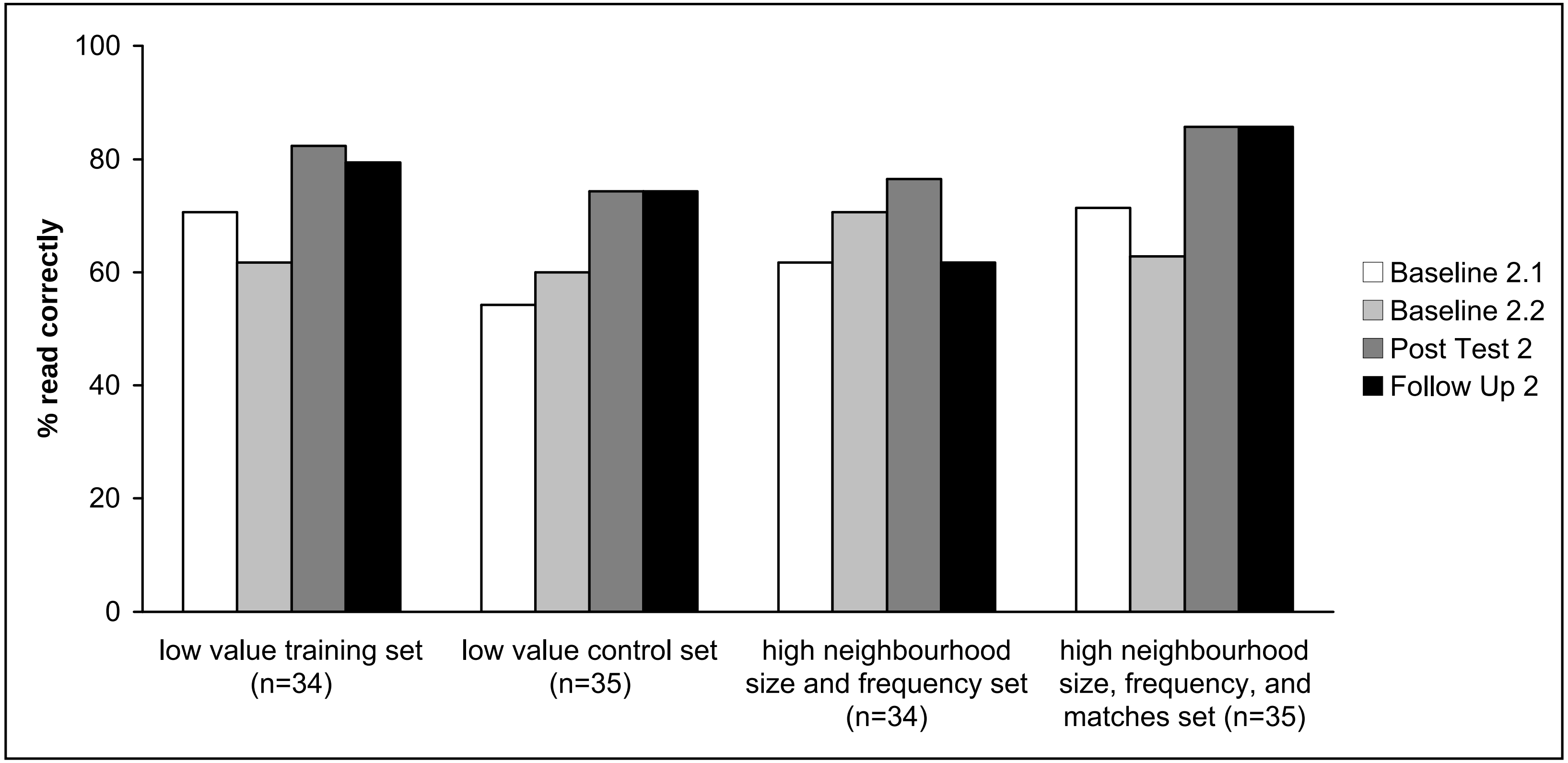
100
80
60
40
20
0
% read correctly
low value training set (n=34)
low value control set (n=35)
high neighbourhood size and frequency set (n=34)
high neighbourhood size, frequency, and matches set (n=35)
Baseline 2.1
Baseline 2.2
Post Test 2
Follow Up 2

Figure 7

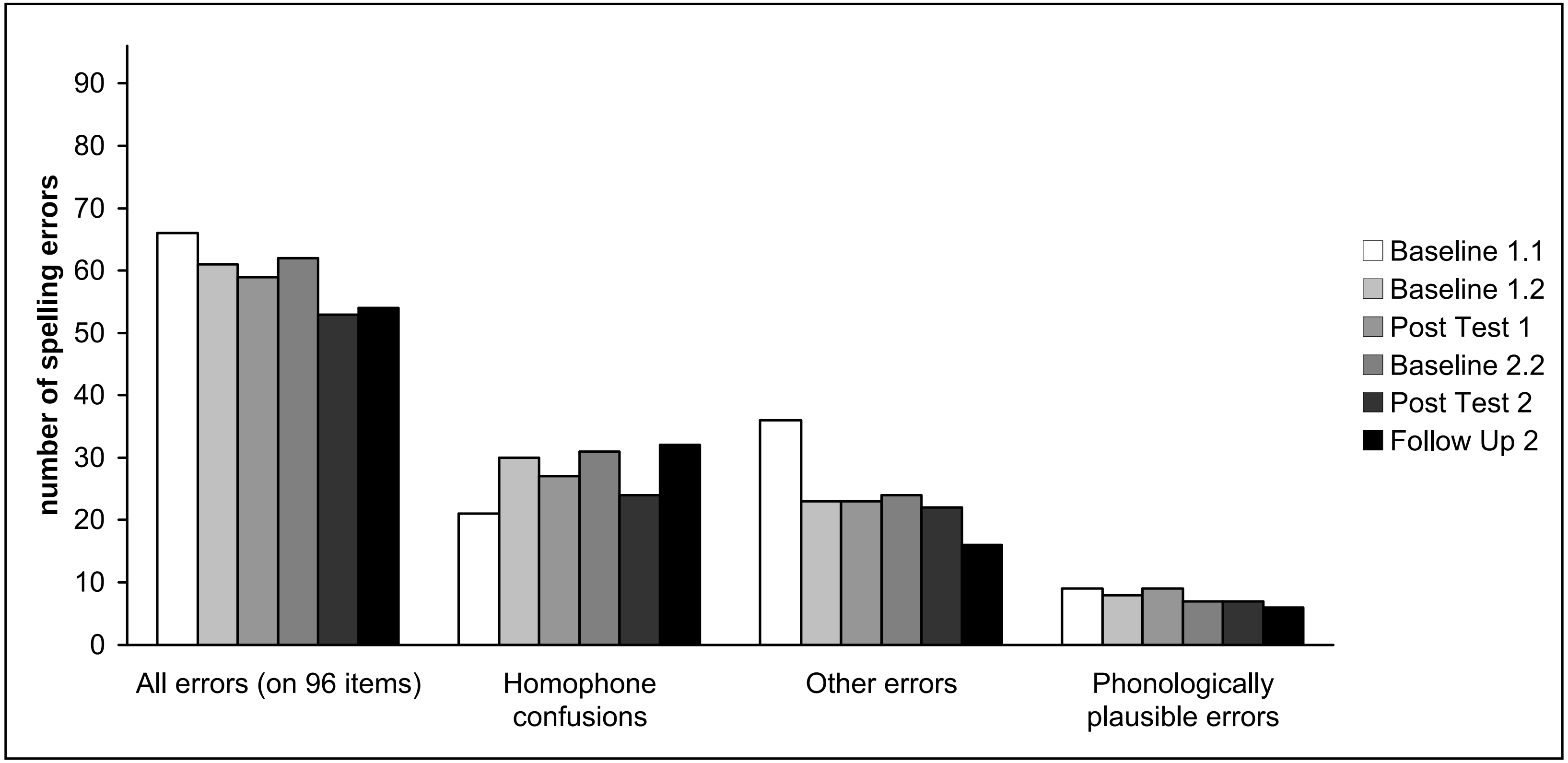
number of spelling errors
90
80
70
60
50
40
30
20
10
0
All errors (on 96 items)
Homophone confusions
Other errors
Phonologically plausible errors
Baseline 1.1
Baseline 1.2
Post Test 1
Baseline 2.2
Post Test 2
Follow Up 2

Figure 8