The Nature of Acquired Dysgraphia:

Patterns of Impairment and Rehabilitation

Janna Geertruida (Trudy) Krajenbrink

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

ARC Centre of Excellence in Cognition and its Disorders (CCD)

Department of Cognitive Science

Macquarie University, Sydney, Australia

April, 2015

Declaration

The work in this thesis is my own original work. It has not been submitted for a higher degree in any other university or institution. All of the work reported in this thesis was undertaken during the time I was enrolled as a PhD student at Macquarie University, under the supervision of Prof Lyndsey Nickels and Dr Saskia Kohnen. Ethics approval for the studies reported in this thesis was obtained from Macquarie University's Human Research Ethics Committee, Reference No. 5201200905.

Signed:

Trudy Krajenbrink

10 April 2015

Table of Contents

Acknowledgements

During the time of my PhD I could not have wished for better guidance. I would like to express my deepest gratitude to Lyndsey Nickels, my primary supervisor, who convinced me to come back to Macquarie to start a PhD, and to my second supervisor, Saskia Kohnen. I have truly enjoyed learning from their wide range of expertise, and have been inspired by their energy and enthusiasm for research. Thank you for being so generous with your time and feedback, and for always reminding me to ask the right questions.

Thank you to the directors from IDEALAB for their time and energy in setting up a wonderful and valuable international collaboration: Prof Roelien Bastiaanse, Prof Ria de Bleser, Prof David Howard, Prof Barbara Höhle, Prof Gabriele Miceli, and Prof Lyndsey Nickels.

I would like to thank the participants in my research, in particular the people with aphasia. I admire the enthusiasm and energy with which they performed many tasks, that may have been challenging at times. In particular I would like to thank GEC and JOD and their partners. I am thankful for their hard work during the many sessions, and I have truly enjoyed the chats over tea and (many) biscuits. This research would not have been possible without their ongoing enthusiasm and commitment.

The friendly and supportive environment at CCD (MACCS) have contributed to my enthusiasm for research, and have made my time as a PhD student so wonderful. I would like to thank Lesley, Katie, Teri, Marcus and Craig for helping out along the way.
Over the years colleagues became friends, and have made it such a nice place to work.

In particular I would like to thank Anastasiya, Britta, Danielle, Huachen, Linda, Nathan, Rochelle, Serje, Thushara, and Yvette.
Thank you also to my dear friends back home, showing that friendship can cover great distances: Nienke, Laura, Martine, Vera, Rimke, Suzanne, Jeanine, Mariska and Rinie.

I have always felt greatly supported by my parents in my decision to start a PhD in Australia for which I am grateful. *Pap en mam – ik heb mij altijd zeer gesteund gevoeld in mijn keuzes. Bedankt voor jullie brede interesse en het voortdurende vertrouwen in mijn werk.* Also many thanks to my brother Henk, and Anne, for their valuable support and interest in my work. Finally, I am very grateful to Pete, for being so caring and supportive, and for always making me laugh.

This PhD has been carried out as part of the Erasmus Mundus joint

International Doctorate for Experimental

Approaches to Language and Brain

(IDEALAB)

Macquarie University (Australia)

University of Newcastle (UK)

University of Groningen (the Netherlands)

University of Potsdam (Germany)

University of Trento (Italy)

General Summary

This thesis focuses on the processes involved in spelling, and in particular explores the nature of impairment and rehabilitation of acquired dysgraphia, using a cognitive neuropsychological approach. The first study investigates the spelling impairment of GEC, a man with acquired dysgraphia. GEC showed characteristics of graphemic output buffer impairment, together with lexical influences (e.g., frequency) on performance and a large number of deletion errors of multiple letters (fragment errors). A detailed error analysis provided evidence to support the hypothesis that fragment errors can be the result of rapid decay of activation from the graphemic output buffer. The study also concludes that lexical influences could be observed in graphemic output buffer impairment due to cascading of activation.

The second study investigates the nature of acquired sub-lexical spelling impairments. Data from spelling sounds in isolation and spelling non-words are analysed to inform three issues: the relationship between performance on the two spelling tasks, the effects of phoneme-grapheme consistency and frequency on spelling, and the use of context when spelling a vowel in a non-word. Results indicated that people with aphasia show comparable difficulty spelling single sounds and sounds in initial position of the non-word. Furthermore, accuracy of individual PGCs was influenced by frequency and consistency of the non-word. Finally, no evidence was found for a loss context sensitive rules when spelling vowels.

The final two studies focus on rehabilitation of acquired dysgraphia. Study three comprises a literature review on generalisation effects after treatment of acquired dysgraphia. This study summarises 40 treatment studies, investigating the link between type of impairment, method of treatment, and generalisation. Some treatment studies in the literature have found an improvement in untreated items, however it is unclear what predicts such generalisation of treatment effects.

Study three highlights that a mechanism of interactive processing may play a role in treatment and generalisation. The final study therefore investigates this mechanism of interactivity within the spelling process in two treatment studies examining the role of orthographic neighbourhood size on the effects of treatment and generalisation in two individuals with acquired dysgraphia. Feedback between the orthographic lexicon and the graphemic output buffer predicts a target word will activate orthographically related words. However, while treatment improved spelling for treated items, there was no generalisation and no evidence for effects of orthographic neighbourhood size on treatment. It was hypothesised that severe impairment to the graphemic output buffer reduced the feedback within the spelling system.

This thesis contributes to our understanding of the nature of spelling impairment and rehabilitation. Furthermore, the thesis highlights the value of cognitive neuropsychological methods in research of acquired dysgraphia.

GENERAL INTRODUCTION

Written language is an invaluable means of communication in everyday life. Making a shopping list, filling in a form, taking notes – many important tasks require the translation from an idea into corresponding written words (Beeson, 2004). Recently this has become even more the case, mobile devices such as phones and tablets have become important tools for social communication for all generations. Writing e-mails and text messaging are a relatively new yet frequent form of communication which have resulted in increased importance of writing for many individuals.

The complex task of translating an idea into corresponding words relies on a number of cognitive components and processes. Moreover, skilled spellers can spell[1] a word in a variety of ways: spelling a word to dictation, spelling the name of a picture, spelling aloud by naming the individual letters in a word, and typing it on a keyboard (Tainturier & Rapp, 2001).

Acquired brain damage, such as a stroke, can result in aphasia: an impairment in the processing of language. This impairment can affect written language, resulting in difficulties in reading (dyslexia) and/or spelling (dysgraphia)[2]. Studying patterns of impairment has informed our understanding of the process of spelling (e.g., Rapp, 2002). This thesis will focus on impairment and rehabilitation of acquired dysgraphia within a cognitive neuropsychological approach. In this Introduction, we will discuss four types of methods commonly applied in cognitive neuropsychology that have informed our understanding of the spelling process to date: single case studies, case series, computational modelling, and treatment studies. The aims of the current thesis will be described in relation to these four approaches.

[1] In the literature different terms are used to describe the process of written word production – here the term 'spelling' is used.

[2] Following the tradition of cognitive neuropsychological research, the term dysgraphia refers to any difficulty in the process of spelling, where a distinction has been made between central and peripheral deficits. Thus, specific difficulties in producing a handwritten form (as opposed to typed) and poor accuracy in spelling are both a form of dysgraphia.

A Cognitive Neuropsychological Approach to Spelling Research

A main goal of cognitive neuropsychological research is to understand the characteristics of cognitive processes through the study of impaired performance (acquired or developmental) (Caramazza & Coltheart, 2006; Caramazza & Hillis, 1993; Kohnen & Nickels, 2015). For example, studying the pattern of impaired spelling in individuals with brain damage can reveal different components of this process, and hence help develop theories of the cognitive processes involved in spelling.

The cognitive neuropsychological approach to written language research came to the fore in the 1980s (Miceli & Capasso, 2006), where single case behavioural studies were used to develop and test theories of normal cognition (e.g., Kay & Ellis, 1987; Shallice & Coughlan, 1980; Warrington & Shallice, 1984). Initially this approach was mostly applied to the study of dyslexia, resulting in the development of detailed theories of reading (e.g., Marshall & Newcombe, 1973; Shallice & Warrington, 1980). Subsequently, detailed descriptions of impaired spelling performance were reported. Indeed, much of our understanding of spelling comes from studies of impaired performance (Tainturier & Rapp, 2001; Whitworth, Webster, & Howard, 2014).

One of the characteristics of cognitive neuropsychological research is the detailed study of symptoms within a single case study design, in order to answer questions about the underlying theory (Caramazza & Coltheart, 2006). For example, Beauvois and Dérousné (1981) reported the case of RG, a man who, following the operation on a tumour showed intact spelling of regular words (e.g., *cat*) and non-words (e.g., *mip)*, in the context of impaired spelling of irregular words (e.g., *choir*). This set of symptoms has since been labelled 'surface dysgraphia'. Shallice (1981) described the opposite pattern of PR, a man who, following a stroke, showed selective impairment of non-word spelling, in the context of relatively intact spelling of (regular and) irregular words, which has been labelled 'phonological dysgraphia'. These contrasting patterns - a double dissociation - have been interpreted as requiring a cognitive architecture of spelling with two distinct pathways (one for spelling irregular words and one for non-words). The majority of studies of spelling impairment have thus adopted this so-called dual route framework (e.g., Rapp, 2002; Tainturier & Rapp, 2001).

Early work in cognitive neuropsychology resulted in 'box and arrow' models: an understanding of the different components (boxes) of a certain process, and the connections between these components (arrows) (Caramazza & Coltheart, 2006). Caramazza and Coltheart (2006) argued that data from performance on a given task has informed this 'basic architecture', and, in addition, the analysis of error patterns has contributed to a more detailed understanding of cognitive processes. I will now explore in more detail the different methodologies used within this approach to inform our understanding of the spelling process, beginning with the single case study approach.

Single case studies to help define theory of impairment. One of the most influential single case studies in the domain of spelling is that of individual LB as first described by Caramazza, Miceli, Villa, and Romani (1987). This case study has provided evidence for the existence of a working memory component in spelling: the graphemic output buffer. Within this component, the abstract graphemic representation is held active while the output processes are being prepared. Caramazza et al. (1987) proposed a number of predictions about the nature of impairment after selective deficit to this component: a length effect in performance, similar performance across different spelling tasks, and errors that reflect a disruption of the graphemic structure (e.g., substitutions or transpositions of letters): a pattern that was shown by LB.

In addition to defining the nature of an impairment to the graphemic output buffer, further study of LB also helped specify the nature and structure of the graphemic representations processed by the graphemic output buffer (Caramazza & Miceli, 1990). Detailed error analyses showed that LB's performance was influenced by a number of variables. For example, the occurrence of deletion errors was constrained by consonant/vowel structure: deletions occurred frequently from a vowel or consonant cluster (e.g., *sfondo* – *sondo*), but a single vowel between two consonants was virtually never deleted (e.g., *tirare* did not result in *trare*). This pattern of errors was not compatible with a theory that assumed that graphemic representations were simple, linear representations of a set of graphemes, that only specify order and identity of letters (Caramazza & Miceli, 1990). Consequently, the results from the investigation of LB's errors provided evidence for representations of graphemes as complex, multi-layer representations.

In sum, a detailed single case study of an individual with graphemic output buffer impairment has informed our understanding regarding the nature of graphemic representations - detailed error analyses of single case studies of acquired dysgraphia allow further development of theories beyond the level of the basic architecture (Caramazza & Coltheart, 2006).

In addition, these case studies have informed our understanding of the impaired language system - LB has allowed us both to understand the workings of the graphemic output buffer and how brain damage may affect he working of this component - it allows us to specify a theory of the impairment. This is a vital step in any attempt to provide treatment for individuals with acquired dysgraphia - without a full understanding (theory) of the impairment treatment cannot be adequately targeted (e.g., Caramazza & Hillis, 1993).

Study One of this thesis takes this approach. It investigates a subtype of graphemic output buffer impairment with the aim of further defining the theory of graphemic output buffer impairment. GEC shows characteristics similar to the seminal case study of LB (Caramazza et al., 1987), and other cases of graphemic output buffer impairment that have since been reported (e.g., Cubelli, 1991; Miceli, Capasso, Benvegnù, & Caramazza, 2004, Tainturier & Rapp, 2004). GEC also showed errors ('fragment errors') whose origins have been much debated in the literature. Detailed error analyses in this case study aimed to determine whether or not fragment errors are another feature of graphemic output buffer impairment.

Case-series designs to define theory of impairment. The basic features of dual route architecture were based on dissociations that occurred across single cases. However, some have argued that relying on single case studies alone is problematic. Patterson and Plaut (2009) discuss these issues in detail, suggesting that, for example, individual differences in functional neural organisation limit the conclusions that can be drawn about cognition. In addition, they argued that even though two individuals may have similar underlying deficits, the use of strategies may lead to differences in performance, which again make it more difficult to draw general inferences from single case studies.

Schwartz and Dell (2010), however, noted that cognitive neuropsychology does not rely on single subject research alone. They showed that case series investigations are just as valuable a tool for informing the functional organisation of a cognitive process. Case series methodology involves a sample of participants with similar cognitive impairment. After systematic assessment on the same tasks and ideally using the same materials, inferences can be made about a certain cognitive ability, by examining the variability that is found across the case series. For some hypotheses only a case series approach is suitable (e.g. examining correlations between rate of errors and a particular psycholinguistic variable; Nickels, Howard, & Best, 2011). Nevertheless, researchers using case series must carefully consider the criteria used to select the individuals in the sample. While some heterogeneity in the study sample is required, heterogeneity of the 'wrong' type may limit the conclusions to be drawn (e.g., if individuals have other impairments that impact on the variable of interest; Nickels et al., 2010; Rapp, 2011).

An example of a case series that has informed our understanding of spelling is a study by Buchwald and Rapp (2009). This study of four individuals with acquired dysgraphia was used to explore the distinction between 'long term memory' (the orthographic lexicon) and 'working memory' (the graphemic output buffer) in spelling. The traditional distinction between these two types of impairments had been challenged (Buchwald & Rapp, 2009). For example, Sage and Ellis (2004) showed that BH's impaired performance included some classic features of graphemic output buffer impairment (e.g., letter errors), but lexical influences (e.g., frequency) on performance were also reported, which traditionally have been proposed not influence buffer processing (Caramazza et al., 1987). Sage and Ellis proposed that these effects are the result of lexical influences that can play out in the buffer.

Buchwald and Rapp (2009) compared four individuals to investigate the distinction between orthographic long term memory and working memory in spelling. The results showed two clear and distinct patterns of impairment, with regards to error types and the effect of lexical factors: damage to the orthographic working memory component resulted in strong effects of length on performance, and only mild effects of lexical variables, and damage to the long term memory component showed severe effects of lexical variables (e.g., frequency and imageability), and no effect of length. Similar to Sage and Ellis (2004), Buchwald and Rapp argued that mild lexical influences on performance in graphemic output buffer impairment could be explained as the result of interactive processing between the two components. This study showed how a detailed comparison of different profiles of impairment using a case series approach can further specify theory. It also helped specify a theory of impairment by delineating two impairments with symptoms that sometimes co-occur.

Study Two of this thesis uses a case series approach. It investigates sub-lexical impairment in dysgraphia, in order to further specify theories of impairment of the sub-lexical spelling route. People with aphasia often have sub-lexical spelling impairments. However, the exact nature of these impairments is underspecified. This is partly due to the fact that dual-route theories of spelling do not specify the exact nature of sub-lexical processes. By comparing a group of people with aphasia on tasks assessing sound spelling and non-word spelling we aim to inform theory of sub-lexical processing and sub-lexical impairment.

Computational modelling to help specify theory. A third method to refine theory is computational modelling, including modelling of cognitive impairment. An example is Nickels, Biedermann, Coltheart, Saunders, and Tree's (2008) use of computational modelling to investigate patterns of reading in phonological dyslexia, using the Dual-Route Cascaded (DRC) model of reading (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001).

Nickels et al. (2008) compared data from three individuals with acquired phonological dyslexia with data produced when different stages and processes within the DRC model were lesioned. Nickels et al. (2008) tested a specific theory of phonological dyslexia - that phonological impairment causes phonological dyslexia (e.g., Farah, Stowe, & Levinson, 1996). They simulated a phonological impairment by increasing phoneme decay and phoneme noise within the model. Lesions to these phonological components proved to successfully predict performance for one individual with phonological dyslexia, but different lesions were necessary for two other individuals. Consequently, this result informed theory of impairment in phonological dyslexia, indicating that a phonological impairment is not sufficient to explain all cases of phonological dyslexia. This study illustrates that using a computational model to investigate impairment requires a detailed and specific hypothesis about the underlying impairment in the human data.

However, it is important to note that to be a complete description of a cognitive process a model should be able to account for all features of performance found in human data (Kohnen & Nickels, 2015). Indeed, Nickels et al. (2008) argued that some individuals with phonological dyslexia have shown effects of concreteness or graphemic complexity on reading accuracy, which have not been implemented in the DRC model. Furthermore, while the model could simulate accuracy, it was not able to simulate the different error types that were found in the human data. Although further work in this area is still needed, computational modelling still provides an important means of constraining and specifying our theoretical thinking (Kohnen & Nickels, 2015).

Study Two of this thesis investigates the nature of sub-lexical impairment, as this is not always specified in descriptive models of spelling. In order to specify theory the computational model from Houghton and Zorzi (2003) is used to guide investigation and to specify hypotheses about impairment of sub-lexical spelling.

Using treatment studies to inform theory. Recently, Nickels, Kohnen, and Biedermann (2010) highlighted the importance of another methodology for cognitive neuropsychology: the use of treatment studies to inform theories of cognition. Nickels et al. suggest that within treatment studies, one methodology that can be used to inform theory is an examination of patterns of generalisation.

After any type of treatment of spelling impairment, the primary desired outcome is that the items targeted in treatment have improved. This may be the only improvement: an item-specific effect. Alternatively, there may also be improvement of untreated items or modalities: generalisation. To date, it remains unclear what the exact link is between the type of impairment, the type of treatment, and whether generalisation occurs. A better understanding of this process has both clinical advantages, as a better understanding of generalisation can improve treatment efficacy, as well as theoretical advantages, as it can inform the nature of the spelling process. Indeed, a number of treatment studies have investigated generalisation and have related the results to understanding the process of spelling.

For example, Rapp and Kane (2002) reported a treatment study with two individuals with acquired dysgraphia, suffering from an impairment to the graphemic output buffer (RSB) and to the orthographic lexicon (MMD). After treatment, both individuals showed an improvement of treated items, however RSB also showed an improvement of untreated items. The authors interpreted this result in relation to the underlying impairment, proposing that if treatment improves graphemic output buffer impairment, there is more likely to be generalisation as the graphemes activated are those used in the spelling of all words (Rapp & Kane, 2002). In contrast, improvement at a lexical level will only benefit the accessibility of items that have been treated.

Treatment studies have investigated the role the choice of stimuli might have, and how this relates to the nature of processing. For example, some treatment studies have reported generalisation to untreated items that were orthographically related to treated items (e.g., Harris, Olson, & Humphreys, 2012; Raymer, Cudworth, & Haley, 2003), including orthographic neighbours. (e.g., Sage and Ellis, 2006; see Kohnen, Nickels, Coltheart, & Brunsdon, 2008 for similar results after a treatment study of developmental dysgraphia). These results suggest that there is interactive processing within the spelling system, and hence, show how treatment studies can inform theory.

Study Three reviews 40 treatment studies to investigate the link between type of impairment, method of treatment, and generalisation. We outline general principles of generalisation and discuss how these can inform our understanding of the spelling process. In Study Four we describe a treatment exploring interactivity in the spelling system, by investigating the role of neighbourhood size in treatment.

Thesis Outline

This thesis is an investigation of the nature of spelling processes and spelling impairment using a cognitive neuropsychological approach.

Study One contains a detailed single case study investigating the underlying impairment in GEC, a man with acquired dysgraphia. GEC showed characteristics of graphemic output buffer impairment in combination of influences of lexical variables (e.g., frequency) on performance. In addition, GEC showed an intriguing pattern of errors, with many deletions of letters (e.g., *cheese: ch*). The aim of this paper was to investigate whether non-traditional error characteristics (lexical influences and fragment errors) could be part of the graphemic output buffer symptom complex.

Study Two reports data from a case series of people with aphasia and control participants on sound spelling to dictation task and a non-word spelling task, in order to investigate the nature of sub-lexical spelling and further specify sub-lexical processes. We consider the connectionist dual-route model of spelling from Houghton and Zorzi (2003) and use their description of the sub-lexical spelling process as a basis for our investigation.

The final two papers focus on rehabilitation of acquired dysgraphia. **Study Three** is a detailed literature review into treatment of acquired dysgraphia, with a focus on generalisation of treatment results.

Study Four uses treatment to investigate interactivity of processing in spelling. Interactivity between the lexicon and graphemic output buffer predicts that activated lexical representations co-activate orthographically similar lexical entries, such as orthographic neighbours (e.g., Kohnen et al., 2008, Sage & Ellis, 2006). We investigated this prediction by examining the role of neighbourhood in treatment of two individuals with acquired dysgraphia.

Finally in the General Discussion we review the contributions of the thesis to our understanding of the nature of spelling and spelling impairment. Furthermore, it discusses some important issues regarding the methodologies used in this thesis.

References

Beauvois, M., & Dérousné, J. (1981). Lexical or orthographic agraphia. *Brain, 104,* 21-50.

Beeson, P. M. (2004). The remediation of written language. *Topics in Stroke Rehabilitation, 11*(1), 37–48.

Buchwald, A., & Rapp, B. (2009). Distinctions between orthographic long-term memory and working memory. *Cognitive Neuropsychology, 26*(8), 724-751.

Caramazza, A., & Hillis, A. (1993). For a theory of remediation of cognitive deficits. *Neuropsychological Rehabilitation, 3*(3), 217-234.

Caramazza, A., & Coltheart, M. (2006). Cognitive neuropsychology twenty years on. *Cognitive Neuropsychology, 23*(1). 3-12.

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: Evidence from a case of acquired dysgraphia. *Cognition, 26,* 59-85.

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

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,* 204-256.

Cubelli, R. (1991). A selective deficit for writing vowels in acquired dysgraphia. *Nature, 353,* 258-60.

Farah, M. J., Stowe, R. M., & Levinson, K. L. (1996). Phonological dyslexia: Loss of a reading-specific component of the cognitive architecture? *Cognitive Neuropsychology, 13*(6), 849-868.

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

Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology, 20*(2), 115-162.

Kay, J., & Ellis, A. (1987). A cognitive neuropsychological case study of anomia. Implications for psychological models of word retrieval. *Brain, 110,* 613-629.

Kohnen, S., & Nickels, L. (2015). Methods of cognitive neuropsychology. In J.D. Wright (Ed.), *International encyclopedia of social and behavioral sciences* (pp. 90-94). Oxford: Elsevier.

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.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia. *Journal of Psycholinguistic Research, 2*, 175-199.

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

Miceli, G., Capasso, R., Benvegnù, B., & Caramazza, A. (2004). The categorical distinction of vowel and consonant representations: Evidence from dysgraphia. *Neurocase, 10*(2)*,* 109–121.

Nickels, L., Biedermann, B., Coltheart, M., Saunders, S., & Tree, J. J. (2008). Computational modeling of phonological dyslexia: How does the DRC model fare? *Cognitive Neuropsychology, 25*(2), 165-193.

Nickels, L., Howard, D., & Best, W. (2011). On the use of different methodologies in cognitive neuropsychology: Drink deep and from several sources. *Cognitive Neuropsychology, 28*(7), 475-485.

Nickels, L. Kohnen, S., & Biedermann, B. (2010). An untapped resource: Treatment as a tool for revealing the nature of cognitive processes. *Cognitive Neuropsychology, 27*(7), 539-562.

Patterson, K., & Plaut, D. (2009). “Shallow draughts intoxicate the brain”: Lessons from cognitive science for cognitive neuropsychology. *Topics in Cognitive Science, 1,* 39-58.

Rapp, B. (2002). Uncovering the cognitive architecture of spelling. In A. E. Hillis (Ed.), *The handbook of adult language disorders: Integrating cognitive neuropsychology, neurology and rehabilitation* (pp. 47-69). New York: Psychology Press.

Rapp, B. (2011). Case series in cognitive neuropsychology: Promise, perils, and proper perspective. *Cognitive Neuropsychology, 28*(7), 435-444.

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

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

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

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

Schwartz, M. F., & Dell, G. S. (2010). Case series investigations in cognitive neuropsychology. *Cognitive Neuropsychology, 27*(6), 477-494.

Shallice, T. (1981). Phonological agraphia and the lexical route in writing. *Brain, 104,* 413-429.

Shallice, T., & Coughlan, A. K. (1980). Modality specific word comprehension deficits in deep dyslexia. *Journal of Neurology, Neurosurgery and Psychiatry, 43,* 866-872.

Shallice, T., & Warrington, E. K. (1980). Single and multiple component central dyslexic syndromes. In M. Coltheart, K. E. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 119-145). London: Routledge & Kegan Paul.

Tainturier, M. J., & Rapp, B. (2001). The spelling process. In B. Rapp (Ed.), *What deficits reveal about the human mind/brain: A handbook of cognitive neuropsychology* (pp. 263-289). Philadelphia, PA: Psychology Press.

Tainturier, M. J., & Rapp, B. C. (2004). Complex graphemes as functional spelling units: Evidence from acquired dysgraphia. *Neurocase: The neural basis of cognition, 10*(2), 122-131.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107*(3), 829-854.

Whitworth, A., Webster, J., & Howard, D. (2014)*. A cognitive neuropsychological approach to assessment and intervention in aphasia* (2nd edition). Hove, E. Sussex: Psychology Press.

STUDY ONE

Graphemic Output Buffer Impairment with Fragment Errors: a Case of Rapid Decay of Activation from the Buffer

Abstract

This paper provides a detailed investigation of the pattern of spelling impairment shown by GEC, a man with acquired dysgraphia. GEC showed clear characteristics of graphemic output buffer impairment as detailed by Caramazza et al. (1987): Performance was equally impaired across different modalities of spelling, showed an influence of length, and errors were characterised by letter errors. However, in addition, GEC showed influences of lexical factors (e.g., frequency and imageability) on spelling, which although traditionally not associated with buffer impairment, have been reported previously, and can be explained as the result of different strength of lexical activation cascading down to the level of the buffer (Sage & Ellis, 2004).

GEC's errors showed a linear serial position effect, including many deletions of letters towards the end of words, and this linear position has been reported in other studies (Katz, 1991; Bormann et al., 2008). We argue that these 'fragment errors' can also be explained within buffer impairment (a rapid decay of information, Katz, 1991, Schiller et al., 2001) and that there was no evidence for a more central impairment causing fragment errors (e.g., Bormann et al., 2008, Cipolotti, et al., 2004).

In sum, we argue that GEC's pattern of impairment including fragment errors is the result of activation decaying in the buffer too quickly (reduced temporal stability, Costa et al., 2011), with lexical influences on processing within the impaired buffer.

Introduction

Studies of individuals with acquired dysgraphia have greatly informed our understanding of the process of written word production and have revealed the components underlying the process of writing a word. An example of such a component is the graphemic output buffer (Caramazza, Miceli, Villa, & Romani, 1987): a working memory component that holds an orthographic representation active while the act of spelling takes place. This paper describes an intriguing spelling impairment, where spelling was mainly characterised by the production of only a fragment – the first few letters – of the target word (e.g., 'ch' for *cheese*). We report different experimental tasks which investigate the characteristics of this error pattern in relation to the role and functioning of the graphemic output buffer in the spelling process.

We will first provide an overview of the process of written word production. Most cognitive theories of spelling to dictation involve a dual-route architecture (see Figure 1), including a lexical route and a sub-lexical route (e.g., Tainturier & Rapp, 2001; Beeson & Rapcsak, 2002).

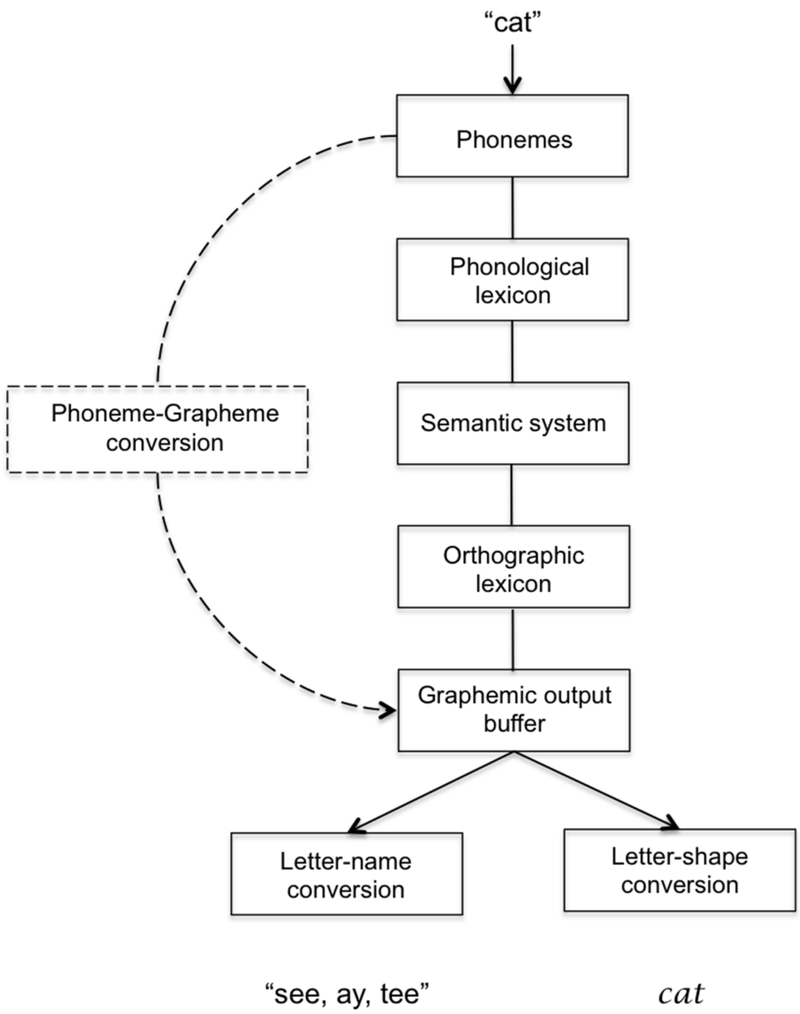


Figure 1. Schematic overview of the architecture for spelling to dictation. Solid lines depict the lexical route, the dotted lines represent the sub-lexical route.

Central processes. The lexical route for spelling to dictation can be used to spell words that have a representation in the lexicon. This is the only route that results in correct spelling of all familiar words (both regularly and irregularly spelled and homophones (e.g., bear/bare)). In writing to dictation, the orally presented word activates its representation in the phonological lexicon, which subsequently activates the corresponding concept in the semantic system. This concept can in turn activate the representation in the orthographic output lexicon where spellings of familiar words are stored. Written naming also occurs via the lexical route, after visual analysis of the object or picture the corresponding concept is activated in the semantic system, and subsequently the orthographic word form is retrieved from the orthographic lexicon.

Unfamiliar words and non-words, that do not have a representation in the lexicon, must be spelled via the sub-lexical route. In order to produce a plausible spelling, phonemes will be converted to graphemes. This route will produce phonologically plausible errors for irregular words (e.g., *yacht*: *yot*).

Graphemic output buffer. After the graphemic representation is retrieved through the lexical or sub-lexical route, it is stored in a working memory component: the graphemic output buffer. In the buffer, orthographic information regarding, for example, letter identity and order is stored temporarily while the physical process of writing is completed. In the case of oral spelling this process consists of grapheme-letter name conversion, while written spelling requires the conversion of graphemes into letter shapes.

Considering the position of the buffer in the spelling process, and the nature of this component, a number of features that characterise a specific impairment to the buffer have been proposed (e.g., Caramazza et al., 1987, Miceli, Silveri, & Caramazza, 1985). We will now discuss these characteristics of buffer impairment and how they are addressed in previous case studies.

Characteristics of graphemic output buffer impairment.

1. Effects of task and modality. A graphemic buffer impairment produces symptoms which are input and output modality independent. This means that all tasks such as written naming, delayed copying, and spelling to dictation should show comparable spelling difficulties. For example, both writing the name of a picture of a cat, and writing the word *cat* to dictation, require the orthographic representation *cat* to be held active in the working memory component while the letter shapes are retrieved. Similarly, because the buffer is a shared component in both oral and written spelling, no difference is expected between the tasks of orally spelling the word *cat* as *see-ay-tee*, and writing *cat* as cat. Furthermore, as noted above, the buffer processes the output of both the lexical and the sub-lexical route and therefore an impairment should manifest in both words and non-words (Caramazza et al., 1987).

2. Effects of stimulus properties. One of the main characteristics of buffer impairment is an effect of length in spelling. Because the buffer functions as an orthographic working memory system, spelling performance will be influenced by the length of words, such that errors are more likely to occur in long compared to short words (Caramazza et al., 1987).

Some authors have argued that lexical factors such as frequency or imageability should not affect spelling performance in graphemic output buffer impairment (e.g. Caramazza et al., 1987), because these processes are assumed to take place at a higher level in the spelling process. Indeed, an effect of frequency of spelling in the context of characteristics of buffer impairment has led some to consider a second locus of impairment: damage to the orthographic output lexicon (e.g., Aliminosa, McCloskey, Goodman-Schulman, & Sokol, 1993) that could have an influence on buffer functioning (Schiller, Greenhall, Shelton, & Caramazza, 2001). However, others have argued that there can be lexical influences on spelling performance in buffer impairment (e.g., Sage & Ellis, 2004, Buchwald & Rapp, 2009). It is proposed that, for example, higher frequency words have stronger connections between semantic representations and word forms. The strength of these connections will cascade down to the level of the graphemic output buffer, and as a result higher frequency words will be more resistant to buffer damage (Sage & Ellis, 2004). Similarly, some lexical effects should also impact on the buffer in interactive models of spelling (Buchwald & Rapp, 2009).

3. Error types. A specific impairment to the buffer is characterised by a large proportion of letter errors (Caramazza, et al., 1987; Caramazza & Miceli, 1990). When writing the word 'tongue' for example, letters may be substituted (tangue), added (trongue), transposed (tognue) or deleted (togue). Other errors that are generally agreed to result from a more central impairment such as phonologically plausible, semantic or morphological errors are infrequent, if lexical processes are intact. As mentioned above, an impairment to the buffer is considered to be spelling modality independent. Therefore, these specific error types are expected to be found across different input and output modalities (e.g., in written naming as well as in writing to dictation).

4. Serial position of errors. Caramazza et al. (1987) showed that in graphemic output buffer impairment the distribution of errors across position in the word resulted in a bow shaped pattern, with more errors occurring for medial letters. This error pattern seems to be an accentuation of a normal error pattern found in unimpaired adult spellers: Wing and Baddeley (1980) investigated 'slips of the pen' made by unimpaired spellers and found that most errors occur in the middle of words. The serial position of errors in spelling has been related to theoretical dimensions of serial behaviour. A bow shaped error pattern has been attributed to *interference*, meaning adjacent letters impose a cost on each other (Olson, Romani, & Caramazza, 2010). As exterior letters do not experience as much interference from neighbouring letters compared to middle letters, these positions are less error prone (Olson et al., 2010).

However, other cases of buffer impairment have not shown a bow shaped error distribution (e.g., Katz, 1991; Miceli, Capasso, Benvegnù, & Caramazza, 2004; Schiller et al., 2001). Rather than a bow shaped curve, for these cases a linear decline in accuracy has been reported, with more errors at final positions compared to middle and initial letters. Katz (1991) attributed this pattern to rapid decay of information at the level of the buffer which makes those graphemes that are held in the buffer the longest the most vulnerable. This account was also adopted by Schiller et al. (2001) to explain a linear increase in errors in individuals TH and PB.

Indeed, Cipolotti, Bird, Glasspool, and Shallice (2004) proposed a subdivision of graphemic output buffer impairment, suggesting that there may be two types: Type A and Type B[1]. Characteristics of Type A are the 'traditional' buffer impairments, similar to the first descriptions by Miceli et al. (1985) and Caramazza et al. (1987): length effects on spelling, many letter errors, no influence of lexical factors such as frequency, and a bow shaped serial position effect in errors (e.g., Caramazza et al., 1987, Jónsdóttir, Shallice, & Wise, 1996). In contrast, individuals like BA (Ward & Romani, 1998) fit a Type B buffer impairment, that also shows features of deep dysgraphia: some effects of lexical factors on spelling, an inability to write non-words, and a more linear increase of errors across the word.

This account has also been implemented computationally by Glasspool, Shallice and Cipolotti (2006), based on a competitive cueing (CQ) account (Houghton, Glasspool, & Shallice, 1994). The dissociating serial position curves are explained by impairments to different components of the model. The bow shaped pattern of errors concentrated in the middle of words (Type A) is the result of 'noise' at the grapheme level. In contrast, the linear increase of errors that characterises Type B is caused by reduced semantic activation of the final positions of representations, which consequently leaves the graphemes at the end of the words with an insufficient activation level to be produced. In sum, Type A is caused by an impaired graphemic output buffer, whereas Type B is the result of impaired input to an otherwise relatively intact graphemic output buffer (Glasspool et al., 2006).

[1] Glasspool et al. (2006) acknowledged there may be other types that show characteristics of both Type A and B. Therefore, Type A and Type B can be seen as two patterns on a continuum.

Costa, Fischer-Baum, Capasso, Miceli, and Rapp (2011) also proposed that the difference in serial position effects was due to different underlying impairments but to different components *within* the graphemic output buffer. Costa et al. described two distinct cases of graphemic output buffer impairment. One individual, GSI, showed a linear increase in errors, whereas the other participant, CRI, showed a bow shaped error pattern. The authors argued that the different serial position curves reflect two key functions of orthographic working memory that are necessary for the selection and ordering of the elements of a representation.

The first function of the buffer is maintaining 'temporal stability', a process to ensure the representation is held active for the duration of writing. This function is related to the concept of 'decay' in the working memory literature (Costa et al., 2011). An impairment to this component will result in more errors towards the end of words: a linear increase (e.g., individual GSI, Costa et al., 2011). The second function of orthographic working memory is 'representational distinctiveness'. For writing this means the correct letter has to be produced in the correct order. This is related to the process of 'interference' in working memory. Impairment of this function will result in a bow shaped error pattern (e.g., individual CRI, Costa et al., 2011). These two dimension can also capture previously described patients with buffer impairment (temporal stability: e.g., Katz, 1991; impaired representational distinctiveness: e.g., Caramazza & Miceli, 1990).

Fragment errors. In its extreme form a linear increase in errors results in omissions of single or multiple letters at the end of words which have been referred to as fragment errors. Katz (1991) reported individual HR, who showed this pattern, omitting a large number of letters in spelling, with a linear increase in the number of errors from the initial to the final letters of the word. As noted above, Katz proposed these errors to be the result of rapid decay of information at the level of the buffer. By the time the last letters of the word have to be written, activation has decayed away completely resulting in omissions.

In contrast, other authors have argued that deletions of letters towards the end of the word are not the result of an impairment to the graphemic output buffer, but are a consequence of lexical impairment. Ward and Romani (1998) described individual BA, who showed a large proportion of deletions of single or multiple letters at the end of words. The errors clearly included (incomplete) portions of the target spelling (e.g., *sulphur*: *sulp*; *book*: *b*). Ward and Romani (1998) argued that these fragment errors were the result of an incomplete activation of the (lexical) orthographic representation, where the final positions of the representation receive less support at the letter level. Indeed, BA's other errors (lexical substitutions, morphological and semantic errors) do suggest a more central impairment, potentially impaired access to the orthographic lexicon and possibly the semantic system. BA (Ward & Romani, 1998) showed symptoms of deep dysgraphia: semantic errors in writing and an inability to write sub-lexically. Other authors have suggested that fragment errors can co-occur with symptoms of deep dysgraphia (e.g., Cipolotti et al., 2004; Bormann, Wallesch, & Blanken, 2008). Cipolotti et al. argued that both semantic errors and fragment errors could stem from the same impaired underlying mechanism of reduced activation of graphemes from the lexical and/or semantic system.

More specifically, Ward and Romani (1998) argued that BA had difficulty retrieving the corresponding letter units for lexical representations, either because of a reduced level of activation of letters, or because the lexical representation provides insufficient support to letter units. This would explain for example the lexical substitutions BA produced that were orthographically similar to the target word (*broom: book*).[2]

In sum, fragment errors have been explained as the result of two different patterns of impairment: rapid decay at the level of an impaired graphemic output buffer, or reduced activation to the lexical representation, in the context of a relatively intact graphemic output buffer. In this case study we report an individual, GEC, who showed characteristics of graphemic output buffer impairment in the context of lexical influences on performance. Errors contained a large proportion of fragments, in the absence of semantic errors. We will describe GEC's spelling[3] performance in detail, and use the characteristics of his impairment to inform the debate regarding fragment errors. First, we describe his general language profile, followed by a more detailed description of his spelling performance in the context of the features of graphemic output buffer dysgraphia. We go on to discuss fragment errors, both outlining the previous literature in more detail and examining the characteristics of GEC's fragment errors.

Case Report

GEC was a 69 year old, right-handed man who suffered a left middle cerebral artery stroke in March 2008. He had a medical history of heart bypass surgery, hypertension, elevated cholesterol, and a previous transient ischaemic attack. After his stroke GEC was diagnosed with global aphasia and a moderate right hemiparesis. At the time of testing, between four and five years post-onset, GEC presented with a moderate aphasia with clear word finding difficulties resulting in non-fluent speech in conversation.

---

[2] However, it is important to note when considering a lexical source for fragment errors that Ward and Romani (1998, p. 193) differ from most authors in considering both word nodes and letter nodes to be part of the orthographic lexicon.
[3] In this paper we will use the term 'spelling' to refer to the process of written word production.

GEC held a university degree in business, and prior to his stroke he worked as a financial planner. When GEC was five years old he moved from Poland to the United Kingdom, and then to Australia when he was eight years old. He continued speaking Polish with a number of friends and relatives for the first 20 years he lived in Australia but indicated that at the time of his stroke he did not speak Polish regularly and had forgotten most of the language. GEC has also been reported in Fieder, Nickels, and Biedermann (2015), and Krajenbrink, Nickels, and Kohnen (2015).

Background Assessment

An assessment of language functions took place over multiple sessions to determine GEC's underlying language impairment (summarised in Table 1), and provide an overview of his written language skills (summarised in Table 2). This assessment used the Comprehensive Aphasia Test (CAT, Swinburn, Porter, & Howard, 2004), selected subtests from Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1992) and other tests as appropriate to give a reasonably comprehensive overview of language processing. Furthermore, visual memory was assessed using the Wechsler Memory Scale (4^{th} edition: WMS-IV; Wechsler, 2009).

Visual memory. A number of subtests from the Wechsler Memory Scale were administered to assess visual (working) memory. GEC's score on the visual memory index (VMI) and the visual working memory index (VWMI) were within normal range for his age group (58^{th} and 27^{th} percentile, respectively).

Semantic processing. GEC's semantic processing of spoken and written input was relatively intact for pictureable stimuli, shown by unimpaired performance on word to picture matching tasks (PALPA and CAT), but he showed a mild impairment of conceptual semantic knowledge (Pyramids & Palm trees test, Howard & Patterson, 1992). His scores on auditory synonym judgements were also below average (especially for the low imageability items), indicating a mild semantic impairment.

Speech production. GEC's spoken picture naming was unimpaired on the relatively easy naming subsection of the CAT. However, his impairment in word retrieval was evident when tested on the word fluency task (CAT) where he was not able to produce the name of a single animal. GEC also scored below the cut off score on a picture description task, which further indicates impairment in production of connected speech (see Appendix A for results on the picture description task).

Auditory processing. GEC was almost flawless in discriminating non-word minimal pairs, indicating an intact phonological input buffer. Repetition was impaired for non-words compared to flawless repetition of words (CAT), suggesting an impairment to the sub-lexical repetition route (from phonological input buffer to phonological output buffer).[4]

Input lexical access. We administered a visual lexical decision task that consisted of regular (e.g., mist) and irregular words (e.g., dove), non-words (e.g., fute), and pseudohomophones (e.g., gote) (PALPA subtest 27, Kay et al., 1992). An adapted version of this test was used in the auditory modality, where the set of pseudo-homophones was changed to non-homophonic non-words. GEC's lexical decision scores were similar across modalities, and compared to available norms for the visual modality GEC's score was below the controls, indicating some difficulties accessing representations in the orthographic input lexicon.

Reading. The results of preliminary assessments of reading and writing are summarised in Table 2.

[4] For visual simplicity, in Figure 1 non-word repetition is subsumed within the link labelled phoneme-grapheme conversion: the 'phonemes' component can be considered the phonological input buffer.

Individual letters and sounds. GEC was impaired on reading aloud letter names, and sounding out individual letters, indicating impaired letter identification and an impairment in grapheme-phoneme conversion.

Word reading. GEC was mildly impaired in reading words. His errors mainly consisted of visual word errors (*choice* for *choir*), letter position errors (*casual* for *causal*, *trail* for *trial*), and morphological errors (*castle* for *castles*, *moves* for *move*). There were no semantic errors. Reading was not significantly influenced by frequency or imageability (CAT reading task: Fisher exact, p = .109 one-tailed, for both variables).

We also administered two reading tasks which aim to detect impairments in assigning positions to letters within the word (letter position dyslexia; Letter Position Test, Kohnen, Marinus, Friedmann, Anandakumar, Nickels, McArthur, & Castles, 2012) and across words (attentional dyslexia; FriCasKo word pairs, Friedmann, Castles, & Kohnen, 2011, unpublished; see notes Table 2 for more details).

On the Letter Position Test GEC made 11 errors, of which five were migration errors (*bread* read as *beard*). According to some adult data (Marinus, Kezilas, Kohnen, & Castles, *in preparation*) normal younger adults rarely make these errors. While we have no data on older adults, five migration errors may indicate some difficulties with letter position coding. On the FriCasKo word pairs test GEC made seven attentional errors (e.g. *even* read as *ever* when listed next to *over*), which could indicate some difficulty with assigning letter positions across words.

We combined stimuli from four of the reading lists (LetPos, FriCasKo word reading and word pairs, PALPA reading). The 214 words for which all variables were available were used in a regression analysis to ascertain the effects of frequency (log written word frequency), imageability, neighbourhood size (number of words that can be formed by substituting a letter in any one position in a letter string) and number of letters on accuracy (values obtained from N-Watch, Davis, 2005). In the regression, the model was significant (χ^2 (3) = 10.604, p = .031), and neighbourhood size was the only significant predictor of accuracy (see Appendix B for regression data).

Non-word reading. GEC's reading of monosyllabic non-words was poor. A third of errors consisted of 'no responses'. Of the remaining errors, 38% were lexicalisation errors (e.g., *spidge* read as *spy*).

Spelling. The results of the initial spelling assessment are summarised in Table 2.

Single letters and sounds. GEC was flawless in matching and same and cross case copying upper and lower case letters, suggesting intact visual feature analysis and abstract letter identification in reading, and of accessing and implementing motor plans for letters in spelling. Spelling single letters to dictation from letter names was also unimpaired. In contrast, spelling letter sounds to dictation was impaired, indicating a difficulty with phoneme-grapheme rules.

Spelling words. Writing was an effortful process for GEC. He wrote slowly, in lower case, with his dominant right hand. When writing to dictation, GEC was able to repeat the target word accurately and he indicated knowledge of the meaning of the word (e.g., spelling the word *physics*, GEC commented: "physics was a subject of mine, but I don't know", and wrote *ph*). However, only three out of 10 responses were correct (CAT written naming and writing to dictation combined). On a homophone spelling task errors consisted of single or multiple letter deletions (*bear: b*, *colonel: col*) and one homophone error (*gait: gate*).

Written picture description was very effortful and GEC only managed to write eight appropriate single words (see Appendix A).

Spelling non-words. GEC's non-word spelling was very poor even for monosyllabic non-words. However, for 73% of GEC's errors the first phoneme was spelled correctly. 60% of errors were orthographically related non-words (e.g., *zie: zict*).

17% of the errors were word responses, of which 83% were orthographically related (e.g., *leet: let*).

A subset of 20 non-words was matched for length to a set of 20 monosyllabic irregular words (from the Krajenbrink, Nickels, & Kohnen list, see Table 5) to compare spelling for words and non-words. Writing non-words (0 correct) was more impaired than spelling words (4 correct) which approached significance (Fisher exact, $p = .053$, one-tailed).

Oral spelling and typing. We asked GEC to orally spell words to dictation, using a set of 10 imageable and high frequency words of four and five letters long (e.g., *mouse, fire*). He was able to orally spell only one word correctly (*crab*). GEC was able to name a number of letters from other items correctly, but either only spelled a fragment (*snail*: "S… A… I don't know") or sounded out some of the letters (*fire*: /f/, R…E..). He often indicated he knew the correct order of letters (*dove*: "D... O… (don't know)... E"). Due to his performance being at floor we did not test this output modality in more detail.

GEC was also presented with a letter board, and was asked to spell by pointing to the letters the same 10 and an extra 10 words that were dictated to him (three to seven letters long). Four words were spelled correctly, and errors resulted in partially correct or orthographically related responses (*market: ma, nerve: nen, drop: do, chicken: chickow, horse: hosme*), a pattern that seemed similar to his written responses.

Table 1

Background assessment

Task		N (items)	*% correct Cut-off* [3]	*GEC (% correct)*
Comprehension				
Word comprehension – spoken	Spoken word picture matching (PALPA 47)	40	95	98
	Spoken word-picture matching (CAT)	30 [1]	83	80*
	Auditory synonym judgement (PALPA 49)	60	n/a	80
	- High imageability	30		90
	- Low imageability	30		70
Word comprehension – written	Written word picture matching (PALPA 48)	40	95	98
	Written word-picture matching (CAT)	30 [1]	90	93
	Written synonym judgement (PALPA 50)	60	87	82 *
	- High imageability	30	91	90 *
	- Low imageability	30	82	73 *
Sentence comprehension – spoken	Spoken sentence-picture matching (CAT)	32 [1]	90	93
	Comprehension of spoken paragraph (CAT)	4	50	100
Sentence comprehension – written	Written sentence-picture matching (CAT)	32 [1]	72	87
Conceptual semantics	PPT 3 pictures	52	94	85 *
Production				
Word fluency	Semantic fluency task (CAT)	n/a	13 items	0 items*
Picture naming – spoken	Spoken picture naming – objects (CAT)	48 [1]	90	90*
	Spoken picture naming – actions (CAT)	10 [1]	80	100
Picture naming – written	Written picture naming (CAT)	21 [1]	71	33*
Connected speech and writing	Spoken picture description (CAT)	n/a	score of 33	score of 19 *
	Written picture description (CAT)	n/a	score of 19	0 *
Repetition	Word repetition (CAT)	32 [1]	91	97
	Sentence repetition (CAT)	12	83	100
	Non-word repetition (CAT)	10 [1]	50	10 *
	Digit string repetition (CAT)	14	57	57
Auditory discrimination	Non-word minimal pairs (PALPA 1 – subset)	36	n/a	97
Auditory lexical decision	Auditory lexical decision (PALPA 27 – subset)	60	n/a	95
	- Regular	15		100
	- Exception	15		87
	- Non-homophonic non-words	30		97

Visual lexical decision	Visual lexical decision (PALPA 27)	60		93
	- Regular	15	94	87 *
	- Exception	15	94	87 *
	- Pseudo-homophones	15	87	100
	- Non-homophonic non-words	15	95	100

Note. **CAT** = Comprehensive Aphasia Test (Swinburn, et al., 2004); **PALPA** = Psycholinguistic Assessment of Language Processing in Aphasia (Kay, et al., 1992); **PPT** = Pyramid and Palm trees Test (Howard & Patterson, 1992)

[1] This CAT subtest gives a score of 2 (immediate correct response) or 1 (>5 seconds delayed correct response or a self-correction). Therefore '*n*' reflects maximum score here, which is two times the actual number of items.

[2] This CAT subtest gives a letter accuracy score. Therefore '*n*' reflects the maximum score here, which is the total number of letters for 5 items

[3] Cut-off scores are generally scores more than 2 standard deviations below the mean score of healthy controls. Cut-off scores from the CAT represent the score that at least 95% of normal subjects exceed. Cut-off scores are taken from the tests manuals or from Nickels and Cole-Virtue (2004) norms.

* score indicates an impairment (at or below cut-off).

Table 2

Initial assessment of reading and writing

Task	Test	N (items)	% correct Cut-off [2]	GEC (% correct)
Reading				
Letter processing	Letter naming (PALPA 22)			
	- Upper case	26	100	53 *
	- Lower case	26		53 *
	Letter sounding (LeST)	51	n/a	27
	Cross case letter matching	29	n/a	100
Pre-lexical processes	Word pairs (FriCasKo)	70	n/a	75
	Migratable words (LetPos)	60	n/a	82
Reading – lexical	Words (CAT)	48[1]	94	90 *
	- High frequency	12		100
	- Low frequency	12		83
	- High imageability	12		100
	- Low imageability	12		83
	- Regular	8		88
	- Irregular	8		100
	- 1 syllable	16		94
	- 3 syllables	8		88
	Words: Regularity (PALPA 35)	60	n/a	95
	- Regular	30		93
	- Exception	30		97
Reading – sub-lexical	Reading non-words (DiRT)	105	n/a	28
	Reading non-words (FriCasKo)	30	n/a	33
Writing				
Letter processing	Cross-case letter copying	28	n/a	100
	Writing letter names to dictation	26	n/a	100
	Writing letter sounds to dictation (DiSTs)	32	n/a	66
Writing – lexical	Writing to dictation words (CAT)	28[1]	86	72 *
	Writing homophones (PALPA 46)	20	n/a	15
	- Regular	10		15
	- Irregular	10		0
Writing: sub-lexical	Non-word spelling (DiSTn)	74	n/a	4
Buffer processing	In sight copying letters and words (CAT)	27	25	100

Note. **CAT** = Comprehensive Aphasia Test (Swinburn, et al., 2004); **DiRT** = Diagnostic Reading Test for nonwords (Colenbrander, Kohnen, & Nickels, 2011); **DiSTn** = Diagnostic Spelling Test – Non-words (Kohnen, Nickels, & Castles, 2009); **DiSTs** = Diagnostic Spelling Test for Sounds (Kohnen, et al., 2009). **FriCasKo** = A screening test for dyslexia subtypes (Friedmann, Castles, & Kohnen, 2011, *unpublished test*) - The word pair reading task assesses attentional difficulties in reading and consists of 35 word pairs that can create new words when between-word migrations of letters occur (e.g., clown – frown could lead to errors such as 'crown – frown'); **LeST** = Letter Sound Test (Larsen, Kohnen, Nickels, & McArthur, In Press); **LetPos** = Letter Position Test (Kohnen, et al., 2012) - The Letter Position Test assesses the ability to assign position to letters, and consists of 60 'migratable' words', in which letters can be rearranged into a new word (beard – bread).); **PALPA** = Psycholinguistic Assessment of Language Processing in Aphasia (Kay, et al., 1992);

[1] This CAT subtest gives a letter accuracy score. Therefore '*n*' reflects the maximum score here, which is the total number of letters for 5 items.
[2] Cut-off scores are generally scores more than 2 standard deviations below the mean score of healthy controls. Cut-off scores from the CAT represent the score that at least 95% of normal subjects exceed. Cut-off scores are taken from the tests manuals or from Nickels and Cole-Virtue (2004) norms.
* score indicates an impairment (at or below cut-off).

Cross modality testing. In order to localise the spelling impairment in more detail than was possible from the initial background screening, we compared performance on the same items across different modalities. To ensure that all reading and spelling tasks were using lexical processes we used irregular words.

Stimuli. 120 irregularly spelled nouns were selected by first choosing words that were irregular for reading according to the N-Watch database (Davis, 2005). Although most words that are irregular for reading are also irregular for spelling (e.g., yacht), we confirmed this was the case by checking the predictability of spellings using a list of English phoneme-grapheme correspondences (Perry, Ziegler, & Coltheart, 2002). This list is based on the frequency of occurrence of the correspondences in different positions in the word. For example, in word-initial position, the sound /s/ is most frequently written as S (*set*), followed by C (*cell*), followed by SC (*scene*). We considered a word to be irregularly spelled if one or more of the PGCs in the word was not the most frequently occurring spelling according to this list. For example, the word *cigar* was considered irregular because use of the most frequent PGCs would result in the spelling *sigar*. We excluded words that were heterographic homophones (e.g., *bear/bare*) and words that had no imageability rating in N-Watch (Davis, 2005).

The final list consisted of four sets of words controlled for frequency and imageability. Half of the irregular words were 'high frequency' with a written frequency of >20 per million (CELEX, Baayen, Piepenbrock, & van Rijn, 1995; a frequency count derived from the COBUILD corpus of 16.6 million words from written sources). The other 60 'low frequency' words had a CELEX written frequency <15. Half of all words had a CELEX imageability rating of >500 ('high imageability'), and the other 60 words had imageability ratings of <450 ('low imageability'). This resulted in four subsets of words (see Appendix C).

All 120 irregular words were tested in three different modalities: writing to dictation, repetition and reading. The 60 high imageability items were also tested in oral and written naming. The five tasks were spread across testing sessions with each item only presented once in a session. Table 3 summarises GEC's performance on the list measured both in terms of accuracy of the whole word (word accuracy), and in terms of numbers of letters correct (letter accuracy). For letter accuracy, we used a scoring procedure based on Caramazza and Miceli (1990) and adapted by, for example, Buchwald and Rapp (2009). In this analysis the target and response are maximally aligned. Furthermore, when a response could be scored in different ways, the one resulting in the least number of error points was chosen (e.g., Schiller et al., 2001). Any letter that is present in the correct (relative) position[5] is given 1 point. A letter that is deleted or substituted receives 0 points. Any letter present but in the incorrect position receives 0.5 point. When two adjacent letters are transposed, both will be scored as 0.75. For example, the target 'algae' written as 'agli' receives a score of 1+0.75+0.75+0+0 = 2.5, and therefore a letter accuracy of 2.5/5 = 0.5.

GEC was relatively unimpaired when reading, repeating and orally naming the items (see Table 3). In contrast, written naming was significantly more impaired than oral naming (McNemar's test exact, $p < .001$). When the two written modalities are compared, written naming of the high imageability words resulted in 13% correct (8/60), and writing to dictation of the same items, 5% correct (3/60). This difference

[5] Relative position indicates the position of a letter in relation to the other letters in the word. For example: the response *huse* for *house* results in a score of 1 for the *u*, as it is still in the correct relative position in relation to its adjacent letters (even though after deletion of *o* it is no longer in the correct absolute (3rd) position).

did not reach significance for either word accuracy (McNemar's test exact, p = .125, two-tailed) or letter accuracy (writing to dictation 0.49, written naming 0.47; $t(59)=0.70$, p = .485, two-tailed).

***Effects of frequency and imageability*.** There was no significant effect of frequency on any of the tasks (see Table 3). There was a significant effect of imageability on reading, as all high imageability words were read correctly compared to 88% low imageability items. When measured as letter accuracy, GEC showed a significant effect of imageability on writing to dictation, with higher imageability items resulting in a higher letter accuracy score (0.49) compared to low imageability items (0.41) (see Table 3). We will return to the influence of lexical factors on GEC's performance in later analyses.

Table 3

Cross modality testing of 120 irregular items

	Repetition	Reading	Oral naming	Written naming		Writing to dictation	
Subsets (n=30)	Word accuracy %	Word accuracy %	Word accuracy %	Word accuracy %	Letter accuracy	Word accuracy %	Letter accuracy
Hi Im. Hi Freq.	100	100	83	17	0.51	10	0.53
Hi Im. Lo Freq.	93	100	93	10	0.43	0	0.45
Lo Im. Hi Freq.	93	93	n/a	n/a	n/a	10	0.40
Lo Im. Lo Freq.	87	83	n/a	n/a	n/a	3	0.43
Effect of Frequency [a]	p = .137	p = .219	p = .212	p = .358	$t(58) = 1.08$, p = .141	p = .122	$t(118) = 0.63$, p = .265
Effect of Imageability [a]	p = .137	p = .010 *	n/a	n/a	n/a	p = .306	$t(118) = 1.67$, p = .049 *
Written naming versus writing to dictation				word accuracy: McNemar, p =.125 two-tailed			
				letter accuracy: Related t-test, $t(59) = 0.70$, p = .485 two-tailed			

Note. Freq. = Frequency; Im. = Imageability; Hi = High; Lo = Low.

[a] word accuracy: Fisher exact, one-tailed; letter accuracy: two sample t-test, one-tailed.

* $p<.05$.

Summary of level of impairment in spelling. Cross modality testing showed that GEC was impaired in the written production with spoken production and reading being relatively unimpaired (as measured by repetition and reading). His good oral picture naming skills compared to his written naming rules out a semantic impairment as the source of his written naming impairment and indicated a problem specific to written word production. The impairment in writing was independent of input modality, as no significant difference was found between written naming and writing to dictation. Difficulties in writing to dictation were unlikely to reflect deficits in phonological or auditory processing since GEC was unimpaired at phoneme discrimination and showed good repetition.

When writing, GEC seemed aware of his errors. He often said he did not know how to spell part of a word, or circled the part of the word that was spelled incorrectly, commenting it “does not look right”, but he was not able to correct himself. When a word was spelled correctly he usually indicated that he knew it was correct. His better reading and monitoring abilities were also evident in his relatively good scores on a lexical decision task (see Table 1).

GEC’s writing errors showed an intriguing pattern - his responses often consisted of fragments of the target item. For example, for the item *cheese* he wrote *ch*. He did not rely (solely) on phoneme-grapheme conversion skills to produce fragment errors, as he was able to produce irregular word onsets. For example, his response for the item *ghost* was *gh*. The *h* is the irregular part of the digraph *gh* representing the sound /g/, and spelling the digraph correctly, reflects (at least partial) intact orthographic knowledge. GEC often repeated the item before, after and even during his spelling attempt, which did not lead to the retrieval of more letters.

As GEC’s errors resembled cases in the literature characterised as graphemic output buffer dysgraphia (e.g., Sage & Ellis, 2004; Schiller et al., 2001), we carried out further investigations using tasks that have been suggested to be diagnostic of buffer impairments (e.g., delayed copying). In addition, we used eight controlled word lists to determine which item characteristics (frequency, imageability, age of acquisition, neighbourhood size, length) influenced GEC's performance. Finally, we investigated whether GEC's errors showed characteristics of spelling errors typically associated with graphemic output buffer impairment (i.e., letter errors).

Copying: effects of delay and item length. Tasks that require the graphemic output buffer to keep letters active in memory have been argued to be diagnostic of impairment to the graphemic output buffer (Caramazza et al., 1987). In the case of a specific impairment to this component, errors are expected to increase when words have to be held active in the buffer for a longer period of time.

GEC was asked to copy, in the same case as well as across case (lower to upper case), 40 irregular words that differed in length (20 four-letter words, 20 eight-letter words). All words contained at least two letters that differ in shape and size between upper and lower case (e.g., g – G). The task consisted of three copying conditions: (1) with the word in sight, (2) covered and then immediately recalled, and (3) covered and recalled after a five second delay. Same-case copying was carried out in all three conditions. Cross case copying was not administered in the five second delay condition as performance was so close to floor. As Table 3 shows, GEC was almost flawless at copying words in sight which rules out a peripheral impairment in copy and transcoding skills. However, when a delay was introduced, accuracy decreased. The longer the delay, the worse his performance, particularly for long words (Jonckheere trend test exact, same case, short words: $z = 2.94$, $p = .001$; long words: $z = 6.41$, $p < .001$).

GEC also copied and recalled 20 non-words in the three conditions in the same case. As for word copying, he was almost flawless at copying the non-words in sight, but again accuracy decreased with increased delay (short and long words combined: Jonckheere trend test, $z = 4.57$, $p < .001$).

Table 4

Copying (% correct)

	Same case copying				Cross case copying	
	words		*non-words*		*words*	
	Short n=20	Long n=20	Short n=10	Long n=10	Short n=20	Long n=20
In sight	100	100	100	95	95	95
Immediate recall	80	5	20	20	35	0
Delayed recall	60	0	0	0	n/a	n/a

Writing to Dictation: Comprehensive Analysis

In order to investigate GEC's writing skills in more detail, we administered a number of controlled word lists to test which factors influenced performance. Table 5 provides an overview of the word lists that were administered.

Table 5

Overview of writing to dictation lists administered

List	Abbreviation	Factors of interest
Krajenbrink, Nickels, & Kohnen (unpublished)	KNK	Frequency, Imageability, Regularity
Buchwald & Rapp (2009, revised version from Lavidor & Ellis, 2002)	B&R	Length
Coltheart, Laxon & Keating (1988)	CLK	Age of Acquisition, Imageability
Laxon, Coltheart & Keating (1988)	LCK	Neighbourhood size
PALPA 44 Regularity and spelling (Kay, et al., 1992)	PALPA Reg	Regularity
Johns Hopkins Dysgraphia battery (Goodman & Caramazza, 1985):		
- Length	JHU Length	Length, Frequency
- Part of Speech	JHU PoS	Word class, Frequency
- PGC	JHU PGC	Phoneme-Grapheme probability, Frequency

Writing to dictation: effects of lexical variables. GEC showed effects of a number of lexical variables, such as frequency, imageability and age of acquisition. The overall pattern indicated that GEC performed better on items that were high in frequency, imageability and that were early acquired, however, results were variable in their strength across lists (see Table 6). For example, on one of the lists that examines effects of frequency (JHU – PGC), GEC was significantly better on the high frequency items compared to the low frequency items, but on a different list (JHU – length) no significant effect of frequency was found. In addition, because GEC's performance was at floor, some effects were only significant when measured in letter accuracy. Because of the variability in results, and floor performance across lists we also examined the effects of these variables using regression analysis for both word and letter accuracy, which we report below.

Regularity. A number of lists allow to test for an effect of regularity (see Table 5), however, regularity was defined in different ways across lists. The definition for the KNK list is based on Perry et al. (2002) frequencies and N-Watch (Davis, 2005), while for the JHU test, rather than regular and irregular, the lists are defined as high or low in phoneme-grapheme probability (PGC probability: the likelihood of generating a correct spelling using phoneme-grapheme conversion rules[6]). In addition to examining effects in each list separately, we also combined all three lists that investigated 'regularity' to provide more power (KNK (regular and irregular subsets), Johns Hopkins Dysgraphia battery: PGC probability, PALPA 44 spelling by regularity) giving a total of 172 items (excluding items that were the same across lists). Irregular items were as accurate as regular items (word accuracy: Fisher exact z = 1.39, p = .082, letter accuracy: $t(170)$ = .391, p = .103, one-tailed). Only one phonological plausible error was made on the irregular items, but this error could also be the result of a single letter substitution (*voice: voise*). Therefore, regularity does not seem to have a significant effect on GEC's spelling.

[6] A low PGC probability means a less than 10% chance of being spelled correctly using phoneme-grapheme mappings from Hanna et al. (1966; as cited in Sanders & Caramazza, 1990). A high probability word has a greater than 50% chance of being spelled correctly using the Hanna et al. PGC frequency norms.

Table 6

Influence of item characteristics on writing to dictation

Factor	List	Items	N	Word accuracy %	Statistical test [a]	Letter accuracy	Statistical test [b]
Length	B&R	5-letters	38	8	*p* = .115	0.54	*t*(55.15) = 3.92, *p* < .001 ***
		8-letters	39	0		0.36	
	JHU Length	4-letters	14	43	Jonckheere trend test, *z* = 2.76 *p* = .003 **	0.73	One way trend ANOVA, F(1,65) = 19.37, *p* < .001 ***
		5-letters	14	29		0.71	
		6-letters	14	21		0.57	
		7-letters	14	14		0.52	
		8-letters	14	0		0.38	
Age of Acquisition (AoA)	CLK	Early AoA	40	28	*p* = .018 *	0.66	*t*(78) = 1.61, *p* = .056
		Late AoA	40	8		0.57	
Imageability (Im.)	CLK	High Im.	40	20	*p* = .014 *	0.65	*t*(78) = 1.86, *p* = .033 *
		Low Im.	40	3		0.59	
	KNK	High Im.	60	5	*p* = .500	0.49	*t*(118) = 1.67, *p* = .049 *
		Low Im.	60	7		0.41	
Neighbourhood size (N)	LCK	High N	39	49	*p* = .082	0.76	*t*(76) = 1.21, *p* = .115
		Low N	39	31		0.69	
Frequency	JHU Length	High Frequency	35	29	*p* = .122	0.63	*t*(68) = 1.54, *p* = .064
		Low Frequency	35	14		0.53	
	JHU PGC	High Frequency	55	38	*p* = .002 **	0.69	*t*(108) = 2.89, *p* = .002 **
		Low Frequency	55	13		0.53	
	KNK	High Frequency	60	10	*p* = .060	0.47	*t*(118) = 0.63, *p* = .265
		Low Frequency	60	1		0.44	

[a]: All comparisons are Fisher exact tests (unless specified otherwise), all reported *p*-values are one-tailed; [b] all reported *t*-values are two-sample *t*-tests.

* *p* < .05. ** *p* < .01. *** *p* < .001.

Table 7

Comparison measures of regularity

List	Items	No of items	Word accuracy %	Statistical test [a]	Letter accuracy	Statistical test [b]
JHU PGC probability	High PGC	29	31	*p* = .290	0.60	$t(108) = 0.16$, $p = .438$
	Low PGC	81	23		0.61	
PALPA 44	Regular	20	25	*p* = .642	0.57	$t(38) = 0.10$, $p = .462$
	Exception	20	25		0.58	
KNK	Regular	20	35	*p* = .064	0.68	$t(38) = 0.84$, $p = .204$
	Irregular	20	10		0.61	
Combined analysis across sets [c]	'regular'	65	32	*p* = .082	0.63	$t(170) = 0.391$, $p = .103$
	'irregular'	107	21		0.62	

[a] Comparisons are Fisher exact tests (unless specified otherwise), all reported *p*-values are one-tailed; [b] All reported *t*-values are two-sample *t*-tests, one-tailed; [c] Double items were excluded.

Regression analyses. Regression was performed to ascertain the effects of stimulus properties on the likelihood that GEC wrote a word correctly. The variables included were: written logarithmic frequency, bigram frequency (average bigram token frequency across the entire letter string), imageability, neighbourhood size, number of letters (all obtained from N-Watch (Davis, 2005), using CELEX database (Baayen et al., 1995) and age of acquisition (Kuperman, Stadthagen-Gonzales, & Brysbaert, 2012). For the regression analysis we only included the response from the first time GEC was asked to spell each word, and excluded words that did not have an imageability value. Of the remaining 609 words that GEC wrote to dictation 112 (18%) were correct, and the average proportion of correct letters was .56.

All the variables showed significant effects on word accuracy (see Appendix D) and significant correlations with letter accuracy (see Appendix E). As expected, the correlations between many of these variables was significant, and hence regression is required in order to determine which variables uniquely predict performance. To examine word accuracy we used logistic regression and to examine letter accuracy we used simultaneous multiple regression.

Results. The regression model was statistically significant, both for word accuracy (χ^2 (6) = 111.974, $p < .001$) and letter accuracy ($F(6, 602) = 20.948$, $p < .001$, adj. $R^2 = .416$.). Of the six predictor variables only three were statistically significant: frequency, length and imageability (see Table 8 and 9): Increasing frequency and imageability were associated with increased accuracy, but increasing length was associated with an increased likelihood of making an error. Bigram frequency was close to significant in the analysis on word accuracy: words with more frequent bigrams being more accurate.

Table 8

Summary of logistic regression predicting word accuracy (n=609 items)

	B	SE	Wald	*df*	*p*	Odds Ratio	95% C.I. for Odds ratio	
							Lower	Upper
Number of letters	-0.595	0.142	17.589	1	<.001***	.552	.418	.728
Log written frequency	0.815	0.232	12.309	1	<.001***	2.259	1.433	3.561
Imageability	0.003	0.001	5.073	1	.024 *	1.003	1.000	1.005
Bigram frequency	0.000	0.000	3.679	1	.055	1.000	1.000	1.000
Age of acquisition	-0.061	0.064	.898	1	.343	.941	.830	1.067
Number of Orthographic neighbours	0.027	0.032	.675	1	.411	1.027	.964	1.094
Constant	-0.982	1.259	.609	1	.435	.374		

* *p* < .05. *** *p* < .001.

Table 9

Summary of multiple regression analysis predicting letter accuracy (n=609 items)

	B	SE_B	β	*t*	*p*
Intercept	0.622	0.103			
Log written frequency	0.073	0.020	.164	3.550	<.001 ***
Length in letters	-0.053	0.009	-.273	-6.043	<.001 ***
Imageability	0.000	0.000	.126	2.847	.005 **
Bigram frequency	0.000	0.000	.054	1.209	.227
Number of orthographic neighbours	0.003	0.003	.036	0.765	.444
Age of acquisition	-0.002	0.005	-.021	-0.444	.657

** *p* < .01. *** *p* < .001.

Writing to dictation: error patterns.

Error types. Graphemic output buffer dysgraphia is characterised by a large proportion of 'letter errors' (e.g., *yacht: yahct,* or *yicht,* or *yaicht,* or *yaht*), relative to a low number of other errors such as phonologically plausible errors (*yacht: yot*) and semantic errors (*yacht: boat*).

In the next section we analyse GEC's errors. First we report a general classification of errors in writing to dictation and written naming of the types of errors that are expected to result from a lexical impairment: semantic errors, orthographic errors and morphological errors and phonologically plausible errors. An error was classified as orthographically related when either at least 50% of target letters were in the response (*task: trash*), or at least 50% of response letters were target letters (*hatred: hit*; based on Nickels' (1995) analysis of phonological errors in spoken production).

Finally, we will investigate whether GEC's error pattern is similar to the characteristic pattern of graphemic output buffer dysgraphia by examining the occurrence of letter errors.

General classification. There was no evidence that GEC's errors were the result of lexical impairment: only 1.9% of all errors could be categorised as semantic, morphological or phonologically plausible errors (see Table 14). Furthermore, the errors in these categories could all be the result of letter substitutions and (multiple) deletions (e.g., *caravan: car, noise: noisy, squad: squod*) and therefore did not clearly indicate a lexical impairment. The majority of errors (87%) were orthographically related to the target. Only 10% of errors were orthographically unrelated to the target (e.g., *brother: odlam*).

Table 10

General error classification in writing to dictation (n=586 errors)

Error type	Example	N	%
No response		5	0.9
Semantic error	anybody: anyone	4	0.7
Morphological error	ski: skis	3	0.5
Phonologically plausible error	squad: squod	4	0.7
Orthographically related [1]		511	87.2
- words	task: trash	93	
- non-words	priest: prient	418	
Orthographically unrelated		59	10.1
- words	though: value	16	
- non-words	bottom: brude	43	
Total		586	100

[1] Either ≥50% of target letters are in response (task: trash) or ≥50% of response letters are target letters (hatred: hit)

We also classified the errors GEC made in written naming of 60 irregular items (Krajenbrink et al., unpublished list, see Table 3), and compared this to writing to dictation of the same 60 items (see Table 11). The distribution across tasks is similar: For both modalities over 90% of errors are orthographically related to the target word, and few error types were found that could indicate a lexical impairment.

Table 11

General error classification in written naming (n=52 errors) compared to writing to dictation (n=57 errors) of same 60 items

Error type	Example	Errors written naming N	Errors written naming %	Errors writing to dictation N	Errors writing to dictation %
No response		1	1.9	1	1.9
Semantic error	car: bus	0	0	0	0
Morphological error	ski: skis	0	0	1	1.9
Phonologically plausible error	sugar: suger	2	3.8	1	1.9
Orthographically related [1]		48	92.3	54	94.7
- words	chalk: cork	3		11	
- non-words	hockey: holked	45		43	
Orthographically unrelated		1	1.9	0	0
- words	school: value	0		0	
- non-words	syringe: ch	1		0	
Total		52	100	57	100

[1] Either ≥50% of target letters are in response (task: trash) or ≥50% of response letters are target letters (hatred: hit)

Analysis of letter errors. The previous analyses showed that GEC made a large proportion of orthographically related errors, where target and response share a large proportion of their letters. 'Letter errors' have been reported as a characteristic of graphemic output buffer dysgraphia. These errors can be further subdivided into: letter additions (e.g., tongue as *tonguer*), deletions (e.g., tongue as *togue*), substitutions (e.g., tongue as *tongul*), and transpositions (e.g., tongue as *tognue*). We made a distinction between error categories involving a single letter, or multiple letters. Errors involving more than one subtype of error or ambiguous errors were categorised as a mixed error.

We excluded the five ‘no response’ errors and the 11 phonologically plausible, semantic and morphological errors, and categorised the remaining 570 overt errors (see Table 12).

Table 12

Number of letter errors (with % in parentheses) as a function of stimulus length (n=570 errors)

Length of stimulus	Addition		Deletion		Substitution		Transposition		Mixed	Total
	single	multiple	single	multiple	single	multiple	single	multiple		
3									3 (100)	3 (100)
4	1 (1)		4 (4.2)	10 (10.4)	22 (22.9)	13 (13.5)	1 (1)		45 (46.9)	96 (100)
5	1 (0.5)		7 (3.7)	19 (10.1)	18 (9.6)	6 (3.2)	2 (1.1)		135 (71.8)	188 (100)
6			4 (3.4)	27 (22.7)	5 (4.2)	2 (1.7)	1 (0.8)	1 (0.8)	79 (66.4)	119 (100)
7			2 (3.1)	23 (35.9)					38 (59.4)	63 (100)
8				37 (45.7)	1 (1.2)				43 (53.1)	81 (100)
9			1 (5)	8 (40)					11 (55)	20 (100)
Total	2 (0.4)		18 (3.2)	124 (21.8)	46 (8.1)	21 (3.7)	4 (0.7)	1 (0.2)	354 (62.1)	570 (100)

Over 60% of GEC's errors were mixed errors involving more than one error type (see Table 12). Furthermore, about a quarter of all errors involved a deletion of single or multiple letters. Only 12% of errors could be classified as a single letter error, and the number of single letter errors decreased with stimulus length (Jonckeere trend test comparing items of 4 to 9 letters, $z = 6.22$, $p < .001$). Table 13 displays the analysis of these single letter errors, and shows that the majority were substitution errors. In 80% of the substitution errors the consonant/vowel (CV) status of the substituted letter remained constant. This is consistent with reports showing that substituted letters in dysgraphia overwhelmingly preserve CV status of the target (e.g., Caramazza & Miceli, 1990; Miceli et al., 2004; Schiller et al., 2001: see Miceli and Capasso, 2006, for an overview), argued to support the distinct representation of consonants and vowels in the spelling system.

Table 13

Classification of single letter errors

Types	N (%)
Additions	2 (2.8)
Deletions	18 (25.7)
Substitutions	46 (65.7)
- Consonant-Consonant	27
- Vowel-Vowel	10
- Consonant-Vowel	6
- Vowel-Consonant	3
Transpositions	4 (5.7)
Total	70 (100)

Serial position of errors. As already noted above, traditionally, letter errors show a bow shape distribution with more errors occurring in the middle of words (e.g., Caramazza et al., 1987). However, other cases in the literature have shown a linear increase of errors across the word (e.g., Katz, 1991; Schiller et al., 2001).

We investigated whether there was an effect of serial position on GEC's spelling by dividing the total letter accuracy score per position by the total number of target letters in that position (see explanation of scoring method in Assessment section 'Cross modality testing').

Single letter errors. Following the literature, we first calculated GEC's accuracy by position with a subset of his responses that contained single letter errors only (n=70; e.g., *priest*: *prient*) (e.g., Jónsdóttir et al., 1996; but see Sage and Ellis (2004) and Buchwald and Rapp (2009) for criticism of this approach). The target words were 4-9 letters long, and words of different lengths were normalised to five positions according to the method described by Wing and Baddeley (1980).

The accuracy per normalised position shows a generally linear decrease of accuracy, with a slight increase for the last position (see Figure 2). A chi square analysis on accuracy per position shows a significant effect of position on accuracy: $\chi^2(4) = 37.79$, $p = <.001$. The letters in first position were always written correctly.

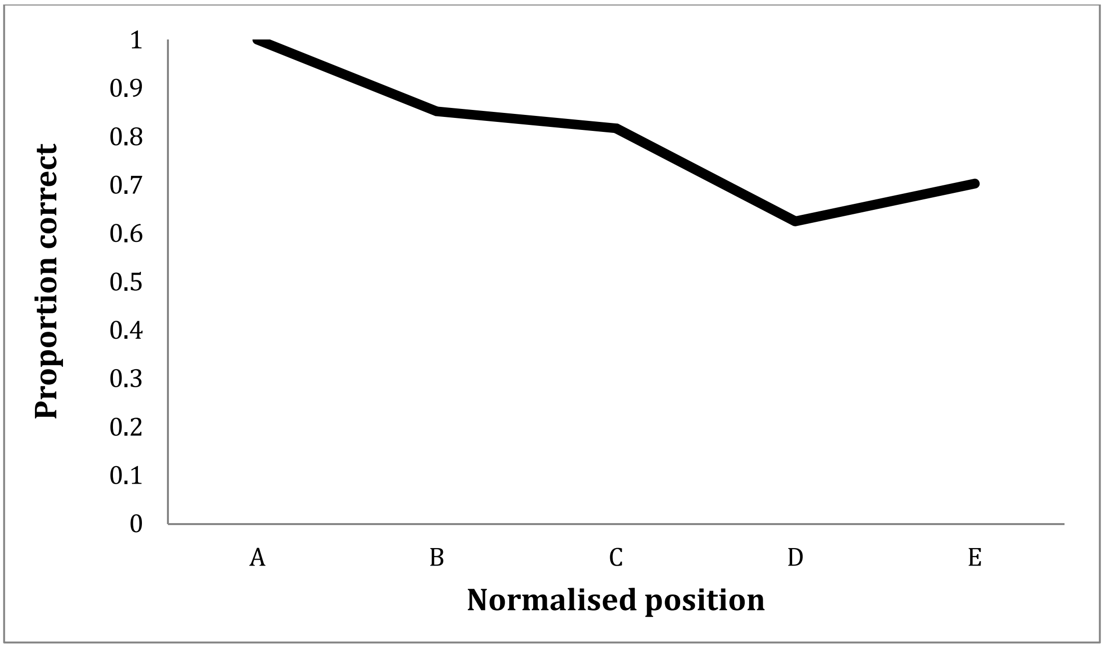
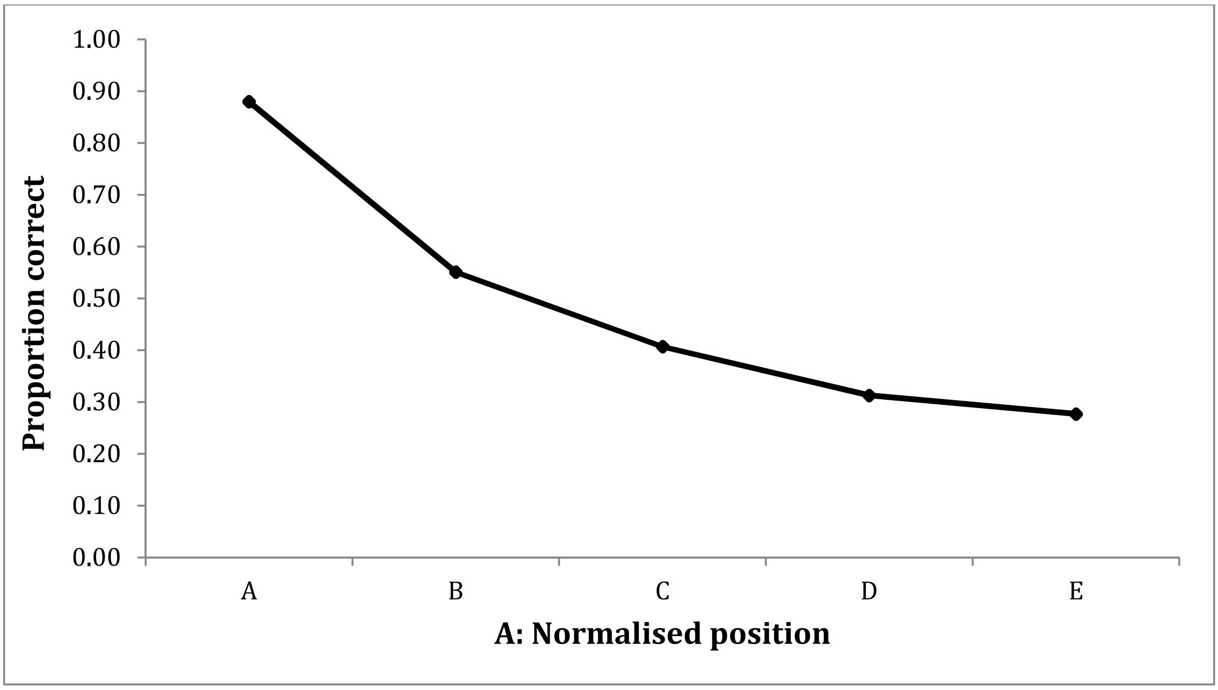


Figure 2. Serial position for stimuli with single letter errors (n=70) across five normalised positions.

All errors. In order to increase the power of the analysis we included a larger error sample, involving all the orthographically related errors in the corpus, excluding orthographically unrelated responses, semantic errors, morphological errors, no responses, and phonological plausible errors. To report a representative error analysis we only included item lengths that were present at least 15 times in the corpus, which led to the exclusion of items with 3, 9, 10 and 11 letters. The analysis is therefore based on of a sample of 492 errors ranging from four to eight letters in length, across five normalised positions (see Figure 3A) and shows the same pattern as for the single letter errors: a linear decrease in accuracy.

Machtynger and Shallice (2009) pointed out that combining words of different lengths in a normalised position analysis can misrepresent the shape of a serial position curve because words of different lengths contribute differently to the five regions. If, for example, the sample of words consists of many six-letter words, this can possibly give too much weight to the medial position, because for a six-letter word, letters three and four are combined to the third region (C). Consequently, we also analysed the same data using absolute positions and again only included positions that occurred at least 15 times in the sample (n=490 errors, ranging from four to eight letters in length).

Figure 3B reports the proportion correct by absolute position in the word. This graph displays a similar pattern to the normalised position curve, with a clear serial position effect: GEC's accuracy decreased for positions later in the word. However, there was no effect of length on the proportion correct per position: the proportion correct for the fourth position in a four-letter word was similar to the same position in an eight-letter word (comparison accuracy fourth position in 4, 5, 6, 7, and 8 letter words: Jonckheere trend test, z = -0.07, p = .473, one-tailed).



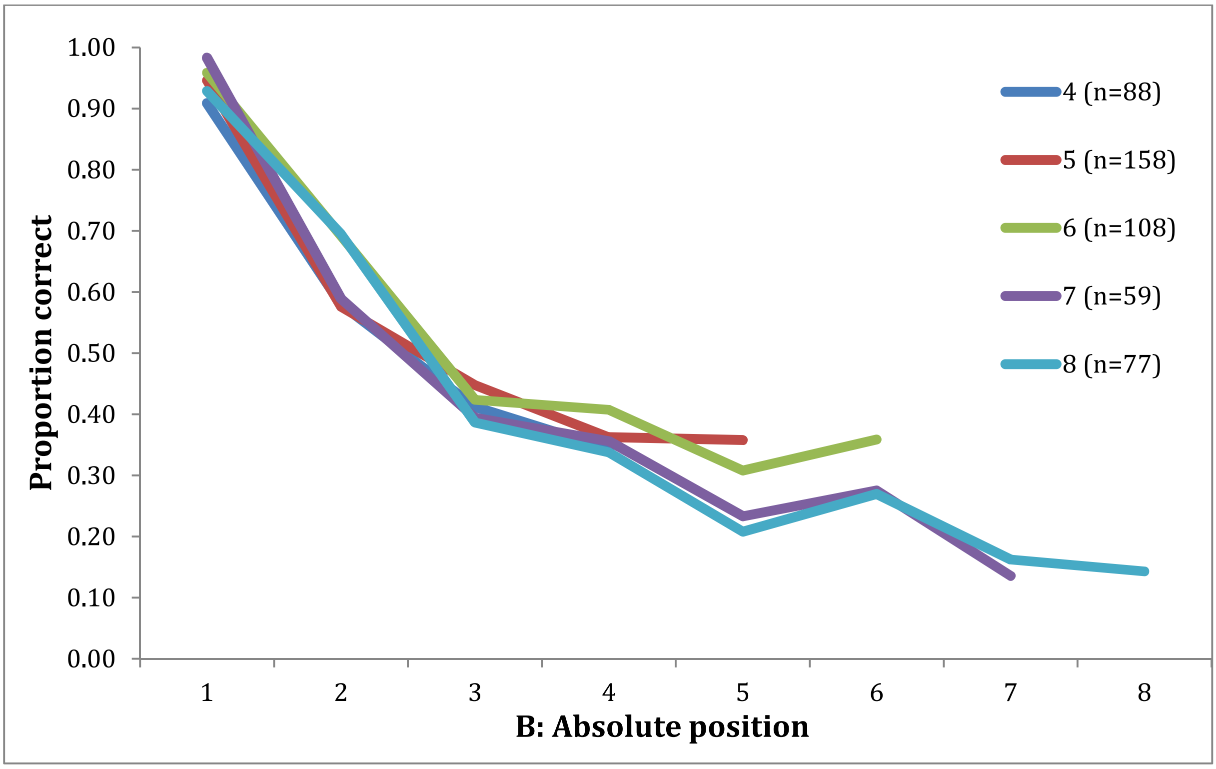


Figure 3. Serial position of orthographically related errors in writing to dictation (n=490) across normalised positions (Figure A) and absolute positions (Figure B).

Summary: nature of GEC's impairment. GEC's performance showed many characteristics of a deficit to the graphemic output buffer (GOB).

- GEC's performance was poor, independent of input modality: his spelling was equally poor for written naming and writing to dictation.
- GEC showed an effect of length: longer words were more error prone.
- In copying an increased delay resulted in an increase in errors.
- Errors were characterised by mainly deletions, but also substitutions, transpositions and insertions of letters.

However, GEC also showed characteristics on which there is less agreement. First he showed an effect of lexical variables (frequency, imageability) on performance: some authors consider that accuracy should not be influenced by any lexical variables in graphemic output buffer impairment (Caramazza et al., 1987). However, GEC's results are consistent with other reports attributed to graphemic output buffer impairment. For example, Sage and Ellis (2004) described BH, who showed characteristics of buffer impairment (i.e., a length effect, letter errors) together with effects of frequency, imageability, age of acquisition, and number of orthographic neighbours (however the effect of length was reduced when other factors (especially neighbourhood size) were controlled, see Sage and Ellis, 2004). Sage and Ellis argued that effects of frequency are compatible with buffer impairment, because a stronger lexical activation of words can cascade down to the buffer (see Introduction). Stronger lexical activation would be predicted for words of higher frequency, higher imageability and that are earlier acquired.

Second, GEC's position of error within the word did not show the classical bow shaped pattern, but rather a linear decrease in accuracy, which, as discussed above, has nevertheless been associated with buffer impairments by some authors (Katz, 1991; Schiller et al., 2001).

In sum, GEC shows characteristics that have all been attributed to graphemic output buffer deficits, albeit some that are more contentious. We now turn to the fact that GEC's errors often involved deletion of single and multiple letters.

Fragment Errors

A large proportion of GEC's errors (61%) were one or more letters shorter than the target. As noted in the Introduction, some cases with buffer impairment have shown deletions of many letters towards the end of the word, and these errors have been termed 'fragment errors' (e.g., Bormann et al., 2008; Ward & Romani, 1998). In the literature there has been debate whether these errors occur in the context of graphemic output buffer impairment, or as part of a different underlying impairment, affecting the lexicon.

We begin by examining the factors affecting the likelihood of a fragment error occurring. Then we examine predictions of two different accounts on the origin of these errors, by presenting tasks that have been used in previous studies as supporting evidence for the two accounts.

Definition of fragment errors. In their description of participant BA, Ward and Romani classified fragment errors as a non-word response that was at least two letters shorter than the target (Ward & Romani, 1998).

Using this definition 218 of GEC's errors (37%) were fragments: 28% of fragments were correct (correct order of letters in correct position: *diamond: dia*), 62% were related (50% or more response letters shared with target: *havoc: huv*) and 9% were unrelated fragments (<50% response letters shared with target: *venom: wa*).

Factors affecting the production of fragment errors. To investigate the origin of fragment errors we analysed which target characteristics predicted the production of a fragment error. We first compared the set of fragment errors to correct responses, and subsequently to other types of errors, to see whether the same factors predict a fragment error compared to other responses.

Fragment errors vs. correct responses. We used a logistic regression to ascertain the effects of written logarithmic frequency, imageability, age of acquisition, neighbourhood size, bigram frequency (mean average bigram token frequency across the entire letter string) and number of letters, on the likelihood that GEC made a fragment error compared to a correct response (variables obtained from N-Watch (Davis 2005) and Kuperman et al., 2012).

As reported above, 9% of fragment errors were unrelated to the target (e.g., *braq* for the word *choice*). Such errors may be an indication of complete failure of lexical access (Bormann et al., 2008) and therefore seem different to a correct or related fragment (e.g., *cho* for *choice*). In contrast to Ward and Romani (1998), we therefore only included fragment errors that were *orthographically related* non-words at least two letters shorter than the target (i.e., 50% or more letters of the fragment were shared with the target). We excluded the semantic, morphological and phonologically plausible errors and no responses. This led to 198 correct or related fragment errors and 119 correct responses. 38 items with missing variables were excluded, leaving 279 items in the analysis (167 related fragment errors, 112 correct responses). The model was significant (χ^2 (6) = 151.135, $p < .001$). Of the six predictors, only length was statistically significant (as shown in Table 14) (although frequency approached significance). Increased length was associated with an increase in the likelihood of a fragment error[7].

[7] We also performed the same analysis including 39 orthographically related *word* fragments (asthma – ash). The results were the same: only length was a significant predictor for fragments ($p < .001$).

Table 14

Binary logistic regression predicting likelihood of fragment error vs. correct response

	B	S.E.	Wald	*df*	*p*	Odds ratio	95% CI for Odds Ratio	
							Lower	Upper
Number of letters	-1.141	0.204	31.323	1	<.001 ***	0.320	0.214	0.477
Log written frequency	0.624	0.343	3.310	1	.069	1.866	0.953	3.653
Imageability	0.000	0.002	0.067	1	.796	1.000	0.996	1.003
Age of acquisition	-0.153	0.095	2.575	1	.109	0.859	0.713	1.034
Number of orthographic neighbours	0.006	0.062	0.010	1	.921	1.006	0.891	1.136
Bigram frequency	0.000	0.000	2.503	1	.114	1.000	1.000	1.001
Constant	5.654	1.874	9.103	1	.003	285.288		

*** $p < .001$

Bormann et al. (2008) investigated the influence of word frequency on fragments, and compared the frequency for a set of 79 correct targets with 79 targets resulting in a fragment error. The two sets were matched for length. They reported that correctly spelled targets were significantly higher in frequency than those that resulted in fragments. Bormann et al. argued that this suggested a lexical locus for the occurrence of fragment errors. However, It may be the case that the effect of frequency on fragment errors that was reported, was in fact an effect of frequency on overall accuracy. In their initial assessment Bormann et al. (2008) tested for an effect of frequency on overall accuracy only on a small subset of words (20 high and 20 low frequency words), which resulted in a non-significant trend for higher frequency words to be more accurate (10/20) compared to low frequency words (6/20). Similarly, when we compared a subset of five-letter targets that resulted in correct responses (n=50) with five-letter targets that resulted in fragments (n=45), we also found a significant difference in frequency with correct responses having a higher frequency (mean log written frequency of 1.81) compared to fragment errors (1.30; $t(93) = 4.02$, $p < .001$), which also seems compatible with the effect of frequency we reported on GEC's overall word accuracy (see Table 8).

In order to further explore the source of fragment errors, we investigated whether different variables (e.g., frequency) play a role in the occurrence of fragments (e.g., *cheese: ch*), by comparing these errors to other erroneous responses (e.g., *choice: chaila*) in a regression analysis. If fragments are errors that stem from a different source (i.e., lexical) we may find different factors (i.e., frequency) influencing performance compared to other errors.

Fragment errors vs. other errors. For this analysis we included the same set of 198 orthographically related fragment errors and compared these with all other orthographically related non-words (e.g., *choice: chaila)*. Semantic, morphological and phonologically plausible errors and no responses were again removed from the analysis. 62 items without complete variable data were excluded, leaving 356 items in the analysis (167 fragments, 189 other non-words).

The model was significant (χ^2 (6) = 77.757, $p < .001$), and only length was a significant predictor ($p <.001$, see Table 15). Increased length was associated with an increase in the likelihood of a fragment error[8].

[8] When we included the 39 orthographically related word fragments in the analysis the same results were found: only length was a significant predictor ($p < .001$).

Table 15

Logistic regression predicting likelihood of fragment error vs. other related non-word error

	B	SE	Wald	*df*	*p*	Odds Ratio	95% CI for Odds Ratio	
							Lower	Upper
Number of letters	0.748	0.113	44.178	1	<.001 ***	2.113	1.695	2.635
Log written frequency	0.179	0.242	0.550	1	.459	1.196	0.745	1.922
Imageability	0.002	0.001	2.137	1	.144	1.002	0.999	1.004
Age of acquisition	0.084	0.062	1.838	1	.175	1.088	0.963	1.229
Number of orthographic neighbours	0.052	0.048	1.198	1	.274	1.054	0.959	1.157
Bigram frequency	0.000	0.000	0.390	1	.532	1.000	1.000	1.000
Constant	-6.224	1.347	21.352	1	.000	0.002		

*** $p < .001$.

In sum, the analyses show a robust effect of length on the occurrence of a fragment error compared to a correct response or another error: for longer words, the probability of a fragment error increases, compared to either a correct response or another orthographically related error. We will now discuss different accounts on the possible underlying impairment resulting in fragment errors that have been proposed.

Investigating theoretical accounts of fragment errors. We will discuss different theoretical accounts of fragment errors and the tasks that have been used to test these accounts.

1. Backward spelling. Katz (1991) argued that fragment errors occur when the information in the graphemic output buffer decays too rapidly, resulting in less support for word final letters compared to initial letters. Under this account, spelling performance should be affected by the order of writing, and Katz used results from a backward spelling task to support this account. In this task, writing starts with the last letter, which means this letter has to be held in the buffer a shorter time and consequently word final letters should be less impaired than in forward writing. This was the pattern shown by HR (Katz, 1991).

In contrast, opposing results on this task have also been argued as evidence against buffer impairment. Individual BA (Ward & Romani, 1998) was also asked to spell words backward, but her error pattern in backward spelling was similar to spelling forwards. When spelling the word *bone* backwards, BA responded ***i**nob*, with an error on the letter that is written first but that is the final letter of the word. The authors argued that BA's result was not compatible with the hypothesis of rapid decay from the buffer: BA's letter errors in backward spelling were related to the ordinal position of the letter in the word, and not to the order in which the word was spelled (Ward & Romani, 1998). Consequently, BA's similar serial position effect on the forward and backward spelling task were argued to be support for a lexical impairment, with the final positions of the lexical entry *bone* having lower activation (Ward & Romani, 1998), an explanation also adopted by Bormann et al. (2008).

Method. GEC was asked to perform a written spelling task, spelling backwards 30 words from the Krajenbrink et al. word list (20 regular and 10 irregular words, ranging from four to eight letters).

Results. In the forward spelling condition GEC wrote seven out of these 30 words correctly, compared to three out of 28 words in the backward spelling condition (we excluded two words GEC wrote in a forward direction). GEC frequently had to be reminded to start with the last letter and continue in serial order, as he often wrote the letters in a different order (e.g. for *bone*, writing *e*, *b*, *n*...).

The serial position curve of GEC's errors was calculated by normalising the different word lengths into five positions (Wing & Baddeley, 1980). Both correct responses and errors were included in the analysis, excluding two responses that were correct but written in forward direction which resulted in 28 words in both conditions.

In the backward condition, GEC showed a clear bow shaped curve, compared to a more linear decrease in accuracy in forward spelling (though with a slight increase for position E; Figure 4): when GEC spelled words starting with the last letter, the last letters are preserved to the same degree as the first letters. The most striking example was GEC's response to the word *monkey*, which resulted in a fragment error in both conditions, however when spelled in forward direction GEC wrote 'mo', and in the backward condition he wrote 'ye'.

The initial position was significantly less accurate in the backward condition compared to the forward condition (Wilcoxon matched pairs test, z = 2.87, p = .004, two-tailed). For the final position, the difference between the two conditions was not significant (Wilcoxon matched pairs test, z = 1.34, p = .091, two-tailed). Furthermore, the first position differed significantly from the last position in forward but not in backward spelling (Wilcoxon matched pairs test, forward: z = 3.80, p <.001; backward: z = 0.32, p = .746 two-tailed).

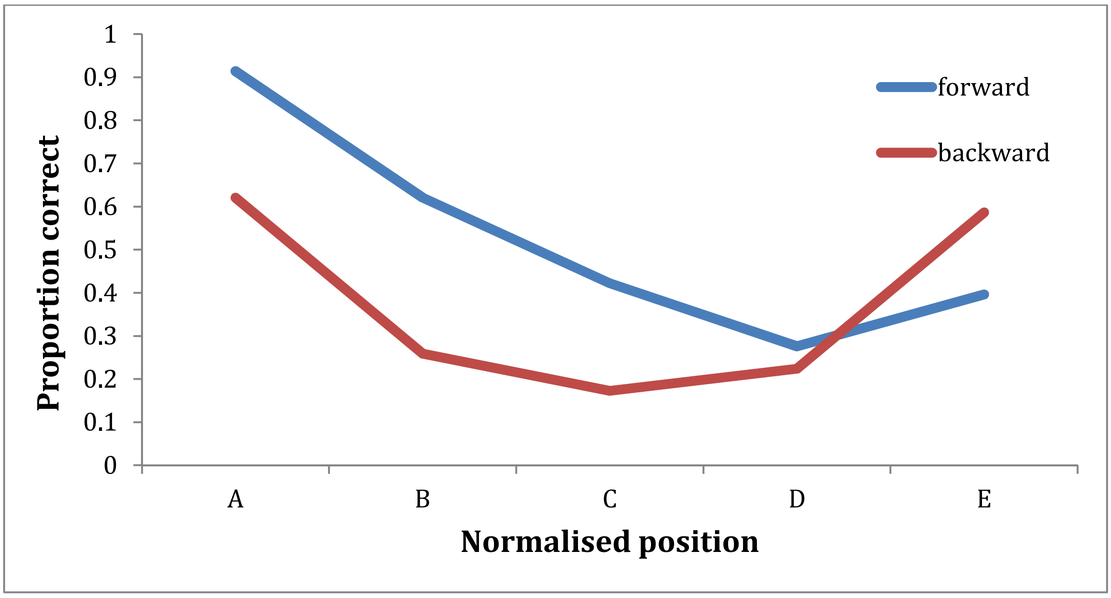


Figure 4. Serial position curves for forward and backward spelling (n=28).

This pattern seems consistent with HR (Katz, 1991), who also showed a difference in serial position curves when forward and backward spelling were compared. GEC's results are different to BA's pattern (Ward & Romani, 1998) and MD (Bormann et al., 2008). For BA and MD there was no effect of the order of writing: errors were related to the ordinal position in the word, and serial position curves were similar across the two spelling task. GEC's result is consistent with a rapid decay of information at the level of the buffer.

However, both HR (Katz, 1991) and GEC show a linear decrease in accuracy in the forward condition, but in the backward condition the linear decrease did not fully reverse into a pattern with highest accuracy for final letters and most errors on the initial letters. Katz (1991) argued this may have been a result of HR's tendency to write the beginning of the word before the middle when writing in the backward condition. Katz (1991) suggested that an explanation for this nonlinear writing order may be related to how attention is guided when 'reading out' the information from the graphemic output buffer. Attention may be automatically guided to the beginning of a word held in the graphemic output buffer, therefore in the forward condition this is consistent with the direction of writing and as a result the beginning letters are spelled most accurately (Katz, 1991).

In the backward condition HR may have only been partially successful in shifting attention to the final letters, and therefore after writing the final letters, attention immediately moved back to the beginning of the word, therefore writing the middle letters last and as a result of information decaying over time, thus least accurately (Katz, 1991). This account would be possible for GEC who also seemed to make most errors in medial position and for a number of items also wrote the medial letters last (see also TH (Schiller et al., 2001)).

However, Schiller et al. (2001) pointed out it is not possible to infer what an individual's performance on this task can indicate about the underlying impairment without understanding the exact way this task is performed, arguing that it remains unclear what mechanisms underpin backward spelling. One could for example, even in backward spelling, access the final letters in a forward manner by repeatedly working towards this position: for example when spelling 'chair', in order to retrieve the final letter, one could access the information in a forward manner by scanning the word: C-H-A-I-**R**, and then again in a similar manner to retrieve the next letter (C-H-A-**I**), and so on. Executing the task in this manner would be similar to spelling in the forward direction, and no difference in error pattern would be expected (although overall performance might be poorer due to an overall slower rate of writing necessitated by the scanning). However, Bormann et al. (2008) questioned this proposal and argued that such a strategy of scanning the word in a forward direction to perform a backward spelling tasks assumes the participant is able to effectively manipulate the orthographic representation, which they argue unlikely in the case of poor performance in dysgraphia patients. Therefore, Bormann et al. (2008) agree with Ward and Romani (1998) that a similar position curve for forward and backward spelling (as was also shown by MD, Bormann et al., 2008) is evidence against the buffer account of serial position effects.

In conclusion, GEC showed an effect of order of writing on letter accuracy which is consistent with the pattern of HR in Katz (1991), and provides further support for a buffer impairment.

2. Word completion task. The rapid decay account states that fragment errors are a result of information decaying too rapidly at the level of the graphemic output buffer. A prediction that follows from this hypothesis is that the error pattern should change when the workload for the buffer is reduced.

Bormann et al. (2008) and Schiller et al. (2001) described a word completion task which aims to reduce the memory load on the buffer. In this task, participants are presented with a list of words, where either the first, or the last, letter(s) are missing. Individual TH (Schiller et al., 2001) was almost flawless when performing this task, compared to impaired writing to dictation of words of similar length (20-40% correct).

MD (Bormann et al., 2008) was also better at retrieving initial and final letters (88% and 73% correct, respectively), compared to writing to dictation (45%). MD showed better performance in completing missing first letters than final letters. This was argued to be inconsistent with a buffer impairment. Bormann et al. argued that as the workload for the buffer in this task is reduced (though probably not completely eliminated) retrieving initial and final letters should be equally difficult. However, when trying to retrieve the missing letters of a target in this task, one has to 'scan' the spelling of the word to find the relevant letter. During this process, the target spelling has to be kept active. Scanning is presumably quicker and easier when retrieving initial letters compared to final letters because the word has to be kept active. Therefore, even though the working memory demands might be reduced, an effect of position is still likely to be found in the case of buffer impairment. Hence, in contrast to Bormann et al., we could suggest that a difference between accuracy for initial and final letters on this fragment completion task would be compatible with buffer impairment.

Method. We provided GEC with the spellings of 40 irregular words from the Krajenbrink et al. (unpublished) list: 20 words of 4-6 letters and 20 words of 8-9 letters. Each word (e.g., potato) had to be completed by filling in two missing letters either at the beginning (_ _ tato), middle (po_ _to) or end (pota_ _) of the word. Each word was presented in all three conditions, in separate testing sessions.

An item was counted as correct when both letters were present in the correct position. For letter accuracy we counted the number of letters correct. If two letters were transposed (e.g., pro**pe**rty: pro**ep**erty) half a point was given for each letter. When a correct letter was written but an incorrect position, only half a point was given (e.g., if UL was required and LE was written, 0.5 was given for L).

Table 16

Performance on the word completion task

Position	Words correct (n=40 per position)	Letters correct (n=80 per position)
Initial	23	62
Medial	17	48
Final	27	61

Results. Table 16 summarises GEC's performance across positions. When spelling these same words to dictation (see earlier sections), GEC only spelled eight words correct. Hence, the results on this task show that when the workload for the buffer is reduced, GEC performs better compared to writing to dictation of these items. Yet, just like TH (Schiller et al., 2001) and MD (Bormann et al., 2008) GEC still made errors. For GEC, the difference in word accuracy across positions only approached significance (word accuracy: Cochran's Q test: Q(2) = 5.43, p = .066; letter accuracy: Friedman's test chi square(2) = 5.65 p = .059). Pairwise comparisons show that initial and medial position differed significantly for letter accuracy (Wilcoxon matched pairs test, z = 2.25 p = .024, two-tailed), medial and final positions for both word and letter accuracy (word: McNemar's test exact, p = .031 two-tailed; letter: Wilcoxon matched pairs test, z = 2.07 p = .039, two-tailed). The initial and final position, however, did not differ significantly in either word or letter or accuracy (word: McNemar's test exact, p = .481, two-tailed; letter: Wilcoxon matched pairs test, z = 0.22 p = .824, two-tailed).

In summary, unlike MD (Bormann et al., 2008), GEC shows similar performance for initial and final positions in the fragment completion task: It seems that when the working memory load is reduced, GEC shows a more bow shaped pattern, perhaps an accentuation of a 'normal' pattern, with more errors likely in the middle of words, where there is more interference from neighbouring letters. The results on this task therefore again provide support for a buffer impairment because they show that reducing working memory load increases accuracy.

3. Letter Probe Task. Schiller et al. (2001) described another task that aimed to reduce the workload for the buffer: They asked the participant whether a letter was present in the spelling of an orally presented word. For example: 'Is there a 'b' in 'debt''? Schiller et al. found that individual PB was 93% correct on this task, which was far better than spelling of the same words. This task assesses whether an individual is able to activate a word's constituent letters even if not able to produce them in writing, and if this is the case, this argues against a lexical impairment and in favour of a localisation of impairment within the graphemic output buffer.

Method. GEC was presented with a spoken word and was asked whether a particular letter (named and shown on letter board) was in the spelling of that word (e.g.: is there a 'b' in 'book'?)[9]. There were 120 words. Half of the stimuli were presented with correct letters and the other half incorrect letters. We manipulated two variables: position of the letter within the word, and sound-letter regularity.

<u>Position</u>: For half of the words the letter in question was present in the first three letters of the word (book: is there a b?) and for the other half the letter in question was located after the first three letters (book: is there a 'k'?). This criterion was chosen as GEC's mean length of fragment was 2.9 letters.

<u>Regularity</u>: As GEC performed moderately accurately on letter sounding, it was possible that he could use knowledge of grapheme to phoneme correspondences to answer a letter probe (knowing that B is pronounced /b/ and hence when given B,

[9] We would like to thank Adam Buchwald for suggesting this task.

converting it to the sounds and comparing the sound /b/ to the sounds in the word *book* rather than knowing the letter B is part of the spelling of *book*). Therefore, we manipulated whether or not the letter in question was a phonologically plausible spelling that was in the actual spelling. For example, the letter J is pronounced /dʒ/. Both the word *jingle* and *ginger* contain the sound /j/, however it is only spelled as a J in *jingle*. Hence, to answer the question about whether *ginger* has a 'j', one has to know the actual spelling. Each letter was administered in one regular item (jingle: is there a J?) and one irregular item (ginger: is there a J?). Words were administered in a random order.

Results. GEC was still relatively poor[10] on this task, although he was significantly above chance, answering 80 out of 120 questions correctly (66.7%; Binomial Test, p < .001). There was no effect of regularity (both irregular and regular sounds were 66.7% correct), nor was there an effect of position (70% initial versus 63% final, Fisher exact, p = .563, two-tailed).

The fact that GEC was as accurate when retrieving information about irregular as about regular words indicates that he did not solve this task purely by relying on non-lexical processes. Rather, he must have accessed lexical information. The lack of an effect of position argues against an impairment to certain letter positions of the lexical representation (e.g., Ward & Romani, 1998). Once again it seems that by reducing the working memory load of the task, the effect of letter position on GEC's accuracy is reduced, and compared to his serial position analysis of writing to dictation, final letters seem to benefit in particular. Nevertheless, in order to complete this task, the spelling of a word still has to be kept in memory and therefore errors still occur due to an impairment to the mechanism that keeps graphemic representations active. Therefore,

[10] There is however no writing to dictation data on these words available for direct comparison.

the results on this task are compatible with a localisation of impairment within the buffer.

4. Copying versus writing to dictation. If difficulties in spelling are the result of information decaying too quickly in the graphemic output buffer, spelling should be impaired when a delay is introduced in a copying task. We reported above that GEC's spelling performance got worse with increasing delay (comparing copying in sight, immediately after the word has been removed, and after a five second delay). However, Bormann et al. (2008) reported that introducing a five second delay in a copying task did not affect their participant MD's performance. Bormann et al. argued that in the case of rapid decay, a delay in the copying task should negatively influence performance and therefore, results were in favour of a more central impairment for their participant (Bormann et al., 2008).

However, in this task Bormann et al. compared copying immediately after the word had been removed from sight, with copying after a five second delay. That is, both conditions were taxing orthographic working memory to some degree because the target was not in sight. For GEC we compared performance in three conditions: with the word in sight (no working memory load), immediately after the word has been removed (increasing working memory load), and after another five seconds delay (highest working memory load). The comparison between the first two conditions is particularly important when considering an effect of delay, and therefore it would be interesting to know how MD's (Bormann et al. 2008) performance compared across these conditions. Indeed for GEC, as his performance is already so poor in the 'immediate' condition when we compare this with 'delay' we also don't find a significant difference.

Ward and Romani (1998) argued that the different nature of copying and writing to dictation can result in different error patterns. They found that in writing to dictation, BA made many fragment errors but when she copied words immediately after they were removed from sight her writing seemed more fluent, more accurate and with fewer fragment errors, compared to writing to dictation (BA copied seven letter words with 70% accuracy, compared to writing to dictation of seven letter words with only 28% accuracy). Furthermore, BA's errors in copying were concentrated in the middle of words. Somewhat counterintuitively, Ward and Romani argued that a copying task places a greater demand on the graphemic output buffer, compared to writing to dictation which requires retrieving a lexical representation. In addition, they suggest that the copying task has less lexical involvement. Consequently, the reduction in fragment errors in the copying task led the authors to conclude that the fragment errors in writing to dictation were the result of an impaired lexical representation that was weaker at the final letter nodes.

In order to investigate these hypotheses, we compared GEC's serial position curve in writing to dictation and copying. When we compared a set of words used both in the same case copying task (direct copying without the word in sight, no extra delay) and in writing to dictation (n=20), we found a more linear decrease in accuracy in writing to dictation (see Figure 5): the reverse pattern to that observed by Ward and Romani (1998). This pattern is consistent with the pattern of results in the letter probe and backward spelling task; when memory load is reduced, performance on final letters improves. This suggests that writing to dictation has a greater memory load than delayed copying.

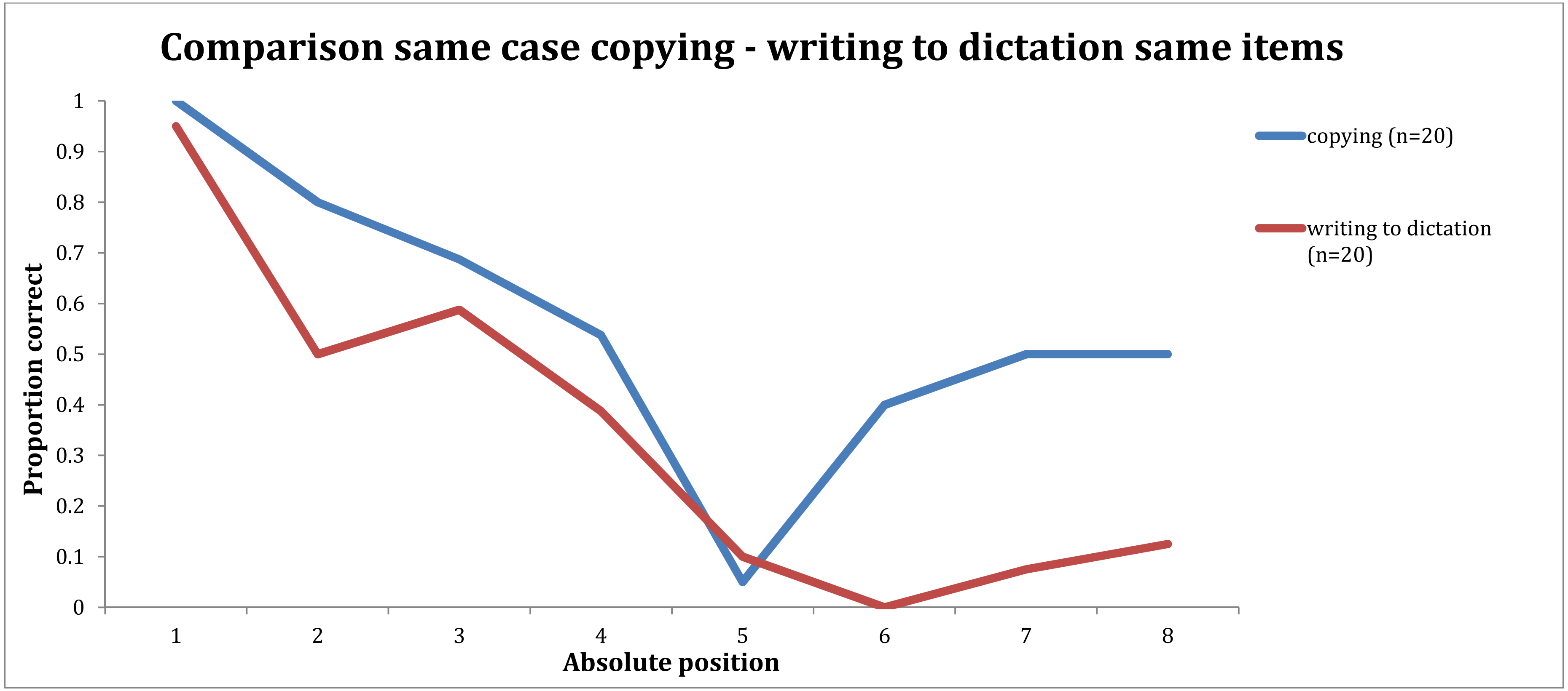


Figure 5. Letter accuracy across absolute position in the word, for direct copying (no extra delay) and writing to dictation of the same 20 items (10 words of four letters, 10 words of eight letters).

Summary for localisation of fragment errors. Around a third of GEC's errors in writing to dictation could be categorised as fragment errors: non-words that were two or more letters shorter than the target. Of these errors 90% consisted of correct and orthographically related fragments. Regression analyses showed that length of the target was a significant predictor for the production of a fragment error.

We discussed a number of tasks that have previously been used to inform whether fragment errors result from rapid decay of information in the graphemic output buffer, or from an impairment to final positions of lexical representations. GEC's results seemed to be most readily interpreted as a rapid decay of information from the graphemic output buffer, as error patterns changed on tasks that reduced the working memory load for the buffer. We will now discuss GEC's results more broadly and the contribution of his data to the literature and models of spelling.

Discussion

We have reported a detailed investigation into the spelling impairment of GEC, a man with extremely poor spelling and characteristics of impairment to the graphemic output buffer as described by Caramazza et al. (1987). GEC also made a large proportion of what have been called 'fragment errors': deletions of letters towards the end of a word. We compared GEC's error pattern to similar cases in the literature aiming to inform the debate on different types of graphemic output buffer impairment, and the origin of fragment errors. We will now discuss the contribution of GEC's result to this topic to investigate the nature of graphemic output buffer impairment and the underlying impairment resulting in fragment errors.

Types of Graphemic Output Buffer Impairment

Caramazza et al. (1987) assigned the graphemic output buffer a rather specific role in spelling, and therefore damage to this component results in a characteristic pattern of impairment. GEC showed a number of these features: an effect of length on spelling, similar performance across input modalities (writing to dictation compared to written naming), greater impairment in delayed compared to direct copying, and a large proportion of orthographically related errors in the absence of many semantic, morphological and phonologically plausible errors.

GEC's impairment also showed other characteristics that traditionally have not been considered to be part of buffer impairment, such as an effect of frequency and imageability on spelling accuracy. Caramazza et al. (1987) argued that impairment of the graphemic output buffer as a post-lexical component is not expected to show an influence of factors that play a role in processes higher up in the spelling system. Indeed, a number of case studies have reported characteristic of buffer impairment without (consistent) influences of frequency on performance (Caramazza et al., 1987, Jónsdóttir et al., 1996, Miceli et al., 1985, Posteraro, Zinelli, & Mazzucchi, 1988). In other cases of characteristics of buffer impairment the co-occurring effect of frequency has led authors to conclude that some (additional) lexical impairment cannot be ruled out (Schiller et al., 2001). However, others have argued that these effects are compatible with buffer impairment: words that are low in frequency will result in lower levels of activation in the buffer resulting in less support for production (Sage & Ellis, 2004). Furthermore, Sage and Ellis argued that lexical influences on the graphemic output buffer are not exceptional, in their review of the literature, 12 of 17 case studies with graphemic output buffer impairment showed signs of lexical influences (e.g., frequency, imageability or concreteness) on spelling (Table 7, Sage & Ellis, 2004). In addition, Buchwald and Rapp argued that the (mild) effects of frequency reported for individuals BWN and RSB can be explained as the result of bidirectional activation between the lexicon and the buffer in an interactive spelling system (Buchwald & Rapp, 2009). GEC's error pattern fits the pattern of graphemic output buffer impairment, with lexical variables (frequency and imageability) also impacting on performance.

However, Cipolotti et al. (2004) argued that the influence of lexical factors is an indicator of a different subtype of buffer impairment, proposing two functional syndromes, graphemic output buffer impairment Type A and Type B, corresponding to two locations of damage. We will consider the suggested subtypes in the context of evidence from GEC's data.

Type A impairment corresponds to the 'traditional' descriptions of graphemic output buffer impairment: no influences of lexical factors on performance, letter errors, and a bow-shaped error curve. In contrast, Type B refers to buffer impairment with features of deep dysgraphia (semantic errors in spelling): a linear increase in the number of errors towards the end of words, lexical and semantic influences on performance and semantic errors, but also letter errors. This pattern fits with that shown by individuals HR (Katz, 1991: effect of word class and impaired non-word spelling however no semantic errors reported), BA (Ward & Romani, 1998: impaired non-word spelling and semantic errors) and DA (Cipolotti et al., 2004: lexical effects and semantic errors).

Within a computational model of spelling, using a competitive queuing account from Houghton et al. (1994), it is proposed that Type B impairment is explained as an impairment in the mappings from semantic to orthographic lexical representations, resulting in reduced activation to these representations. In contrast, Type A results from an impairment to output letter nodes.

GEC's error pattern does not map easily onto the Cipolotti's types A and B. GEC showed lexical influences, so Type A does not apply. However, GEC did not show the main characteristic of Type B either: Less than 1% of his errors could be classified as semantic errors. This is similar to BH, in Sage & Ellis (2004), HR (Katz, 1991) and the individuals reported in Schiller et al. (2001). GEC shows characteristics of graphemic output buffer impairment, but no semantic errors, which is one of the key symptoms of deep dysgraphia. However, GEC does show a linear increase in errors towards the end of the word as definitive of Type B, in contrast to a bow shaped curve defining Type A. Indeed, Cipolotti et al. (2004) categorised other individuals showing this linear error pattern (DA, Cipolotti et al., 2004, HR, Katz, 1991; BA, Ward & Romani, 1998) as Type B. As GEC did not fit other characteristics of Type B impairment, we suggest that he does not fit within Cipolotti et al.'s subtypes and will instead provide an explanation for the linear increase of errors within accounts of buffer impairment.

However, a possible overlap between GEC's impairment and symptoms of deep dysgraphia could be considered in relation to a possible common underlying impairment. Even though GEC did not make semantic errors in spelling, he did show a lexicality effect in some tasks (poor non-word compared to word performance), which may be interpreted as impaired phonological skills. Within a 'primary systems' view of language, both phonological and deep dysgraphia are seen as two impairments on a continuum of severity (Crisp & Lambon Ralph, 2006; Jefferies, Sage, & Lambon Ralph, 2007), with both impairments resulting from a deficit to a common phonological system. This is evident in phonological deficits across different tasks such as repetition, reading, and spelling (Jefferies et al., 2007). The two disorders may reflect different degrees of impairment, where semantic errors may be expected in a severe impairment, whereas milder impairments may not result in these errors, whereas in these cases a lexicality effect may be present across different tasks.

GEC's initial assessment did show a lexicality effect in some tasks. For example, GEC seemed more impaired in reading and spelling of non-words compared to words. However, when spelling accuracy on matched subsets of non-words (0 correct) and words (4 correct) were compared, the lexicality effect only approached significance.

Although we can not exclude an additional impairment of sub-lexical processing affecting spelling of non-words, we believe that the nature of GEC's pattern of spelling impairment is the result of an impairment to the buffer, and is not defined by phonological impairment.

Schiller et al. (2001) argued that two types of impairment may affect the buffer: rapid decay and general noise. Rapid decay results in errors increasing towards the end of the word, which they suggested fits the patterns shown by BA (Ward & Romani, 1998) and HR (Katz, 1991). General noise in the system, on the other hand, was argued to amplify the normal pattern found by unimpaired spellers (Wing & Baddeley, 1980), with more errors in the middle of words (e.g., LB, Caramazza et al., 1987). As explained in the Introduction, Costa et al. (2011) also explained distinct error patterns as the result of impairment to different functions of the buffer: temporal stability and representational distinctiveness.

Costa et al. (2011) proposed that impairment to either of these functions results in a length effect, but with a different underlying cause. In the case of reduced distinctiveness, the interference cost is highest for letters in the middle of words (see also Jones, Folk, & Rapp, 2009), and the longer a word, the more interference there will be. If on the other hand impairment affects the process of temporal stability, information decays too rapidly from the buffer, which mostly affects letters further away from the beginning of the word, resulting in a linear increase in errors, and again longer words will be more error prone compared to short words. This is the pattern found by Katz (1991), Schiller et al. (2001).

We suggest that GEC's data shows evidence for impaired temporal stability resulting in increased decay of activation from the graphemic output buffer. Not only did GEC show a strong linear decrease of accuracy in the word, he also showed a different error pattern when the workload for the buffer was reduced, when performance is less dependent on temporal stability. In the word completion task and the letter probe task, accuracy for retrieving the final letters did not differ from initial letters. Hence, when the representation is required to be active over a shorter time, this is particularly beneficial for final letters. As a result, the error pattern becomes more similar to the 'normal' bow-shaped error pattern (Wing & Baddeley, 1980).

Fragment Errors in Spelling

A large number of GEC's errors were fragment errors. We have provided an analysis of these errors in order to inform the debate regarding the origin of these errors: an underlying impairment at the level of the buffer (Katz, 1991) or a lexical impairment (e.g., Ward & Romani, 1998).

We reported GEC's performance on a number of tasks that have been used in the literature to distinguish between the two accounts. We reported a letter probe task and a word completion task: both tasks are assumed to place a smaller demand on orthographic working memory. Although GEC still made errors on these tasks, retrieval of the word final letters was facilitated in these task conditions compared to writing to dictation. Similarly, accuracy of letters that were in the final ordinal position of a word was higher in a backward spelling task when they were the first letter to be written. For example, when writing the word *monkey* in the (usual) forward direction GEC wrote *mo*, whereas in the backward condition he wrote *ye*. Hence, we conclude that GEC's fragment errors are the result of an impaired component in the buffer to keep the orthographic representation active during spelling.

Methodological Considerations

When investigating the underlying impairment that cause a certain error pattern, findings can differ depending on which errors are included in this error analysis. Sage and Ellis (2004) reported that BH almost always wrote initial letters correctly, and that she made more (multiple) letter errors towards the end of words. However, in the serial position analysis, as is standard, they only included 59 single letter errors out of the total of 701 errors. This resulted in a bow shaped error pattern. Sage and Ellis noted that the serial position function would have shifted to the right if multiple letter errors had been included. It is therefore essential to take into account a representative sample of errors.

In addition, it seems important for the discussion of fragment errors to agree to a clear definition of fragment errors and analysis of these errors. Ward and Romani (1998) categorised non-word responses that are two letters shorter than the target as fragment errors. However this definition also includes unrelated fragments (e.g., *house* as *gib*), which could be the result of a different impairment compared to correct fragments. Furthermore, as mentioned previously, it is important to include multiple letter errors, as some of these may in fact be fragment errors: Sage and Ellis reported that 20% of BH's error involved omission of more than one letter, which often concerned the last part of a word (*sledge: sle*) (Sage & Ellis, 2004). Hence, to be able to investigate the nature of fragment errors, it is important to consider a definition of these errors that will include a representative set of clearly defined fragment errors.

In the analysis of GEC's errors, we reported error analyses on a subset of single letter errors as well as on a larger set including all orthographically related errors (using both absolute and normalised positions) which resulted in a detailed overview of the error pattern. Furthermore, we decided on a definition of fragments errors resulting in the exclusion of unrelated errors, in order to get a better understanding of this error type.

Conclusion

We propose that GEC's fragment errors are the result of insufficient working memory capacity to keep the representation active. The information is decaying too rapidly from the graphemic output buffer, and letters that have to be kept active the longest (i.e., final letters) are most vulnerable. GEC's fragment errors can be explained as an extreme form of reduced temporal stability within the graphemic output buffer. We did not find evidence for a lexical impairment and argued that the lexical influences on GEC's performance are the result of differences in strength of lexical activation cascading down to the level of the buffer.

References

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,* 55-69.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX Lexical Database. Release 2* [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

Beeson, P. M., & Rapcsak, S. Z. (2002). Clinical diagnosis and treatment of spelling disorders. In A. E. Hillis (Ed.), *The handbook of adult language disorders: Integrating cognitive neuropsychology, neurology and rehabilitation.* New York: Psychology Press.

Bormann, T., Wallesch, C-W., & Blanken, G. (2008). "Fragment errors" in deep dysgraphia: Further support for a lexical hypothesis*. Cognitive Neuropsychology, 25*(5), 745-764.

Buchwald, A., & Rapp, B. (2009). Distinctions between orthographic long-term memory and working memory. *Cognitive Neuropsychology, 26*(8), 724-751.

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: Evidence from a case of acquired dysgraphia. *Cognition, 26,* 59-85.

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

Cipolotti, L., Bird, C. M., Glasspool, D., & Shallice, T. (2004). The impact of deep dysgraphia on graphemic buffer disorders. *Neurocase, 10*(6), 405-419.

Colenbrander, D., Kohnen, S., & Nickels, L. (2011). *Diagnostic Reading Test for Nonwords (DiRT).* Retrieved from www.motif.org.au.

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology, 79*, 1-12.

Costa, V., Fischer-Baum, S., Capasso, R., Miceli, G., & Rapp, B. (2011). Temporal stability and representational distinctiveness: Key functions of orthographic working memory. *Cognitive Neuropsychology, 28*(5), 338-362, DOI: 10.1080/02643294.2011.648921.

Crisp, J., & Lambon Ralph, M. A. (2006). Unlocking the nature of the phonological-deep dyslexia continuum: The keys to reading aloud are in phonology and semantics. *Journal of Cognitive Neuroscience, 18*(3), 348-362.

Davis, C. (2005). N-Watch: A program for deriving neighbourhood size and other psycholinguistic statistics. *Behavior Research Methods, 37*(1), 65-70.

Fieder, N., Nickels, L., & Biedermann, B. (2015). *How 'some garlic' turns into 'a garlic' or 'capsicum': A follow-up study on mass and count processing in aphasia.* Manuscript submitted for publication.

Friedmann, N., Castles, A., & Kohnen, S. (2011). *FriCasKo – a screening test for dyslexia subtypes.* Unpublished Test.

Glasspool, D., Shallice, T., & Cipolotti, L. (2006). Toward a unified process model for graphemic buffer disorder and deep dysgraphia. *Cognitive Neuropsychology, 23*(3), 479-512.

Houghton, G., Glasspool, D., & Shallice, T. (1994). Spelling and serial recall: Insights from a competi- tive queuing model. In G. D. A. Brown & N. C. Ellis (Eds.), *Handbook of spelling: Theory, process and inter- vention* (pp. 365–404). Chichester, UK: John Wiley.

Howard, D., & Patterson, K. E. (1992). *Pyramids and Palm Trees*. Thames Valley Test Company, Bury, St. Edmunds, UK.

Jefferies, E., Sage, K., & Lambon Ralph, M. A. (2007). Do deep dyslexia, dysphasia and dysgraphia share a common phonological impairment? *Neuropsychologia, 45*, 1553-1570.

Jones, A. C., Folk, J. R., & Rapp, B. (2009). All letters are not equal: Sub-graphemic texture in orthographic working memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 35*, 1389-1402.

Jónsdóttir, M., Shallice, T., & Wise, R. (1996). Phonological mediation and the graphemic buffer disorder in spelling: cross-language differences? *Cognition, 59*, 169-197.

Katz, R. B. (1991). Limited retention of information in the graphemic output buffer. *Cortex, 27*, 111-119.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic Assessments of Language Processing in Aphasia (PALPA)*. Hove: Lawrence Erlbaum Associates.

Kohnen, S., Marinus, E., Friedmann, N., Anandakumar, T., Nickels, L., McArthur, G. (2012). *Letter Position Test*. Retrieved from www.motif.org.au.

Kohnen, S., Nickels, L., & Castles, A. (2009). Assessing spelling skills and strategies: A critique of available resources. *Australian Journal of Learning Difficulties, 14*(1), 113-150. doi:10.1080/19404150902783450.

Krajenbrink, T., Nickels, L., & Kohnen, S. (2015). *The role of orthographic neighbourhood in treatment of acquired dysgraphia*. Manuscript submitted for publication.

Kuperman, V., Stadthagen-Gonzales, H., Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behaviour Research Methods, 44*(4), 978-990.

Larsen, L., Kohnen, S., Nickels, L., & McArthur, G. (In Press). The Letter-Sound Test (LeST): A reliable and valid comprehensive measure of grapheme-phoneme knowledge. *Australian Journal of Learning Difficulties.* DOI: 10.1080/19404158.2015.1037323

Lavidor, M., & Ellis, A. W. (2002). Word length and orthographic neighborhood size effects in the left and right cerebral hemispheres. *Brain and Language, 80,* 45-62.

Laxon, V. J., Coltheart, V., & Keating, C. (1988). Children find friendly words friendly too: words with many orthographic neighbours are easier to read and spell. *British Journal of Educational Psychology, 58*, 103-119.

Machtynger, J., & Shallice, T. (2009). Normalizing serial position analyses: The proportional accountability algorithm. *Cognitive Neuropsychology, 26*(2), 217-222.

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

Miceli, G., Capasso, R., Benvegnù, B., & Caramazza, A. (2004). The categorical distinction of vowel and consonant representations: Evidence from dysgraphia. *Neurocase, 10*(2), 109–121.

Miceli, G., Silveri, M. C., & Caramazza, A. (1985). Cognitive analysis of a case of pure dysgraphia. *Brain and Language, 25*, 187-212.

Nickels, L. A. (1995). Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word production. *Language and Cognitive Processes*, 10, 13-45.

Nickels, L., & Cole-Virtue, J. (2004). Reading tasks from PALPA: How do controls perform on visual lexical decision, homophony, rhyme and synonym judgments? *Aphasiology, 18*(2), 103-126.

Olson, A., Romani, C., & Caramazza, A. (2010). Analysis and interpretation of serial position data. *Cognitive Neuropsychology, 27*(2), 134-151.

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.

Posteraro, L., Zinelli, P., & Mazzucchi, A. (1988). Selective impairment of the graphemic buffer in acquired dysgraphia: A case study. *Brain and language, 35,* 274-286.

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

Sanders, R. J., & Caramazza, A. (1990). Operation of the phoneme-to-grapheme conversion mechanism in a brain injured patient. *Reading and Writing: An Interdisciplinary Journal, 2,* 61-82.

Schiller, N. O., Greenhall, J. A., Shelton, J. R., & Caramazza, A. (2001). Serial order effects in spelling errors: Evidence from two dysgraphic patients. *Neurocase, 7*, 1-14.

Swinburn, K., Porter , G., & Howard, D. (2004). *The Comprehensive Aphasia Test*, Hove, UK: Psychology Press.

Tainturier, M. J., & Rapp, B. (2001). The spelling process. In B. Rapp (Ed.), *What Deficits Reveal about the Human Mind/Brain: A handbook of cognitive neuropsychology* (pp. 263-289). Philadelphia: Psychology Press.

Ward, J., & Romani, C. (1998). Serial position effects and lexical activation in spelling: Evidence from a single case study. *Neurocase, 4*, 189-206.

Wechsler, D. (2009). *The Wechsler Memory Scale – Fourth edition (WMS-IV).* USA: The Psychological Corporation.

Wing, A. M., & Baddeley, A. D. (1980). Spelling errors in handwriting: A corpus and distributional analysis. In U. Frith (Ed.), *Cognitive processes in spelling.* London: Academic Press.

Appendix A

Picture description task (CAT, Swinburn et al., 2004)

<u>a) Spoken output</u>

I see a boy and he is (…) going to have a motor car.
And I see coffee (…) and an album, and socks on the table (…) and a coffee table (...) and a tie.
(experimenter: "So, what is happening, in the picture?")
Asleep. *(GEC laughs)* And (...) it is (…) it is a cat, and he is (...) counting the fish and gripping the fish and (…) (*GEC points to falling book in the picture)* one down and scones the fellow. (*GEC laughs)* And I don't know what the other one is (*GEC points to the radio in the picture*). It's (…) rrr (…) speaker, and turntable and recorder, but (…)

(*time*: 2 min 45 sec.)

<u>b) Written output</u>

Kid. Car/Truck.

Coffee. Track (*target: table)*. Cou (*target: cup)* of tea. Pho… (*target: photograph)*. C.

Books. Haed (*target: head).*

Cat wa (*target: was).* Fishing on the water, C …

Plate.

Appendix B

Logistic regression reading data

	B	S.E.	Wald	df	*p*	Odds Ratio	95% CI for Odds Ratio	
							Lower	Upper
Log written frequency	0.355	0.313	1.283	1	.257	1.426	0.772	2.634
Imageability	-0.001	0.002	0.112	1	.737	0.999	0.996	1.003
Number of orthographic neighbours	-0.126	0.047	7.206	1	.007**	0.882	0.804	0.967
Number of letters	-0.116	0.244	0.226	1	.635	0.891	0.552	1.436
Constant	2.957	1.899	2.425	1	.119	19.245		

** $p < .01$

Appendix C

Subsets of words with means (and standard deviations) for Krajenbrink, Nickels, & Kohnen (unpublished)

	Total CELEX freq.	Log written freq.	Imageability	Number of neighbours	Length in letters
Irregular words					
High Im. High Freq. (n=30)	65.4 (70.7)	66.2 (67.5)	596.9 (28.4)	1.3 (2.2)	6 (1.4)
High Im. Low Freq. (n=30)	7.2 (3.3)	7.5 (3.5)	595.2 (27.46)	1.6 (3.4)	5.9 (1.4)
Low Im. High Freq. (n=30)	65.6 (60.8)	65.9 (58.8)	356 (57.3)	1.4 (2.3)	6 (1.5)
Low Im. Low Freq. (n=30)	7.1 (3.9)	7.2 (3.7)	374 (58.9)	1.2 (1.7)	6 (1.3)
Regular words					
High Im. regular (n=20)	52.4 (103)	54.8 (107)	588.2 (36.2)	5.2 (3.2)	4.8 (0.5)
Matched to: high im. irregulars (n=20)	53.1 (90.8)	53.7 (86.5)	591.6 (23.8)	3.3 (4.2)	4.9 (0.91)

Note. Freq = Log written frequency; Im. = Imageability

Appendix D

Means for correct and incorrect items per variable (writing to dictation)

	Accuracy	N	Mean	Std. Deviation	Std. Error Mean	Two sample *t*-test
Log written frequency	Correct	119	1.83	0.58	0.05	$p < .001$
	Incorrect	586	1.38	0.64	0.03	
Bigram frequency	Correct	119	1835.90	1713.40	157.06	$p < .001$
	Incorrect	585	1006.69	844.10	34.90	
Number of letters	Correct	119	4.71	0.89	0.08	$p < .001$
	Incorrect	586	5.83	1.45	0.06	
Number of orthographic neighbours	Correct	119	5.13	4.65	0.43	$p < .001$
	Incorrect	586	2.21	3.42	0.14	
Imageability	Correct	112	482.80	116.52	11.01	$p < .05$
	Incorrect	497	455.35	119.56	5.36	
Age of Acquisition	Correct	119	5.94	2.22	0.20	$p < .001$
	Incorrect	586	7.47	2.54	0.11	

Appendix E

Correlations writing to dictation (letter accuracy), n=609 items

	Log written frequency	Bigram frequency	Length in letters	Number of orthographic neighbours	Imageability	Age of acquisition
Letter accuracy	.220 ***	.219 ***	-.351 ***	.255 ***	.117**	-.237 ***
Log written frequency	.	.379 ***	-.168 ***	.172 ***	-.202* **	-.392 ***
Bigram frequency		.	-.321 ***	.459 ***	-.046	-.232 ***
Length in letters			.	-.542 ***	-.058	.295 ***
Number of orthographic neighbours				.	.091*	-.267 ***
Imageability					.	-.382 ***

* $p < .05$. *** $p < .01$

STUDY TWO

Patterns of Sub-Lexical Impairment in Aphasia

Abstract

The aim of this study was to investigate the nature of sub-lexical spelling processing and impairment. Most theories of spelling consider a phoneme-grapheme conversion (PGC) mechanism as the main component of sub-lexical spelling, however few of these theories are fully specified and it is unclear, for example, how factors such as context play a role. We report data from a case series of 13 people with aphasia and 13 control subjects on two tasks: spelling of sounds in isolation and non-word spelling. We focused on three characteristics: 1) the relationship between spelling PGCs in isolation and in non-words; 2) the effects of consistency and frequency on spelling of PGCs, and 3) the use of context when spelling a vowel in a non-word.

PGCs in isolation and in initial position of non-words were spelled equally accurately, supporting the view that the same mechanism underpins the two tasks. However, accuracy of spelling PGCs was reduced when measured across all positions in non-words, suggesting that other factors may influence PGC accuracy in non-words (e.g., orthographic working memory, segmentation). Consistency of PGC correspondences also influenced accuracy of non-word spelling, but at the group level PGC frequency only influenced spelling PGCs in isolation, implying that PGC consistency influences either strength of mappings or of activation of graphemes. Finally, people with aphasia continued to show sensitivity to context in their vowel spelling, indicating that context must play a role in the sub-lexical mappings.

Introduction

The process of spelling is generally assumed to consist of two distinct procedures or routes: a lexical route is required for correct spelling of irregular words, and a sub-lexical route is required for correct spelling of non-words (e.g., Beeson & Rapcsak, 2002; Tainturier & Rapp, 2001). Evidence for these two distinct processing routes (the dual route theory) has been provided by case studies of acquired dysgraphia showing selective damage to either the lexical route (surface dysgraphia) or the sub-lexical route (phonological dysgraphia) (e.g., Behrmann & Bub, 1982; Shallice, 1981).

Spelling using the sub-lexical route requires a number of processes (Rapp, 2002; Tainturier & Rapp, 2001). First, through the process of phonological segmentation, the spoken input is analysed and segmented into smaller units (e.g., phonemes). Subsequently, phonemes are translated into graphemes (e.g., the phoneme /b/ is spelled using the letter B[1]). Finally, activated abstract grapheme representations are mapped onto modality specific output processes (e.g., written or oral spelling, typing; Tainturier & Rapp, 2001). However, while there is some agreement about the general procedures required for sub-lexical processing, most dual route models are descriptive and do not specify how exactly the sub-lexical route operates. For example, does this phoneme-grapheme conversion process use information about context in which the phonemes occur and if so, how?

The aim of this study is to investigate the nature of sub-lexical spelling and its breakdown in aphasia. The hallmark feature of damage to the sub-lexical spelling route is an impairment in non-word spelling. Poor non-word spelling has been interpreted as a (selective) breakdown of sub-lexical phoneme-grapheme conversion (PGC) procedures (e.g., Rapcszak et al., 2009). Yet it remains unclear what the exact nature of

[1] In this paper phonemes are transcribed using the International Phonetic Alphabet and represented in / /, graphemes are represented in capital letters (B), and word and non-word examples are represented in *italics*.

this breakdown is. For example, PGC conversion could be rendered inefficient and faulty by noise, or knowledge of the correspondences may be inaccessible or lost altogether. However, it is difficult to specify what exactly is meant by 'noise', and 'lost knowledge' as these concepts are theory-dependent. Computational models can be used to help specify the nature of processing and explore the impact of impairment (e.g., Nickels, Biedermann, Coltheart, & Tree, 2007, for sub-lexical reading). Consequently, in order to specify the nature of sub-lexical spelling and enable us to think more precisely about the impact of impairment, we will first consider the computational model of spelling described by Houghton and Zorzi (2003). This is currently the most complete computational model of the spelling process and we use their description of the PGC conversion mechanism to focus our research questions. We begin by describing some general features of the model.

Houghton and Zorzi's (2003) model is a dual route multilayer network where input units (phonemes) activate output units (graphemes) either directly (sub-lexical route) or indirectly via hidden units (lexical route). In their implementation of the lexical route each orthographic word form is represented by a single node in the hidden unit layer. Input phonemes activate corresponding orthographic lexical nodes which in turn activate the corresponding output grapheme nodes. There is competition between output graphemes through inhibitory links within the level.

The sub-lexical (direct) route is implemented as a simple two-layer feed forward network which works as a sound-spelling conversion mechanism between input phonemes and output graphemes, and is activated in parallel with the lexical route. Houghton and Zorzi first describe the features of the sub-lexical route when it is working in isolation. The sub-lexical network was trained on a set of over 3000 uninflected monosyllabic words (sound-spelling pairs). During training the graphemes produced by the model are compared to the correct spelling. The error for each output node (the difference between its target and its actual activation value) is then used to change the weights using the Delta rule error correcting algorithm. Depending on the consistency of the correspondences, the model learns either a direct mapping from one phoneme to a grapheme (e.g., /b/ - B), or one phoneme maps to multiple plausible graphemes (e.g., /k/ - C, K, CK; /f/ - F or PH; Houghton & Zorzi, 2003).

In the model, both phonological input and orthographic output have a syllable structure with onset, vowel and coda positions (a total of seven positions, see Figure 1). Complex graphemes are treated as a single orthographic output. This means that complex graphemes, such as PH, compete for output with a simple grapheme F, to represent the sound /f/, just as simple graphemes compete to represent a sound (e.g., C or K to represent /k/).

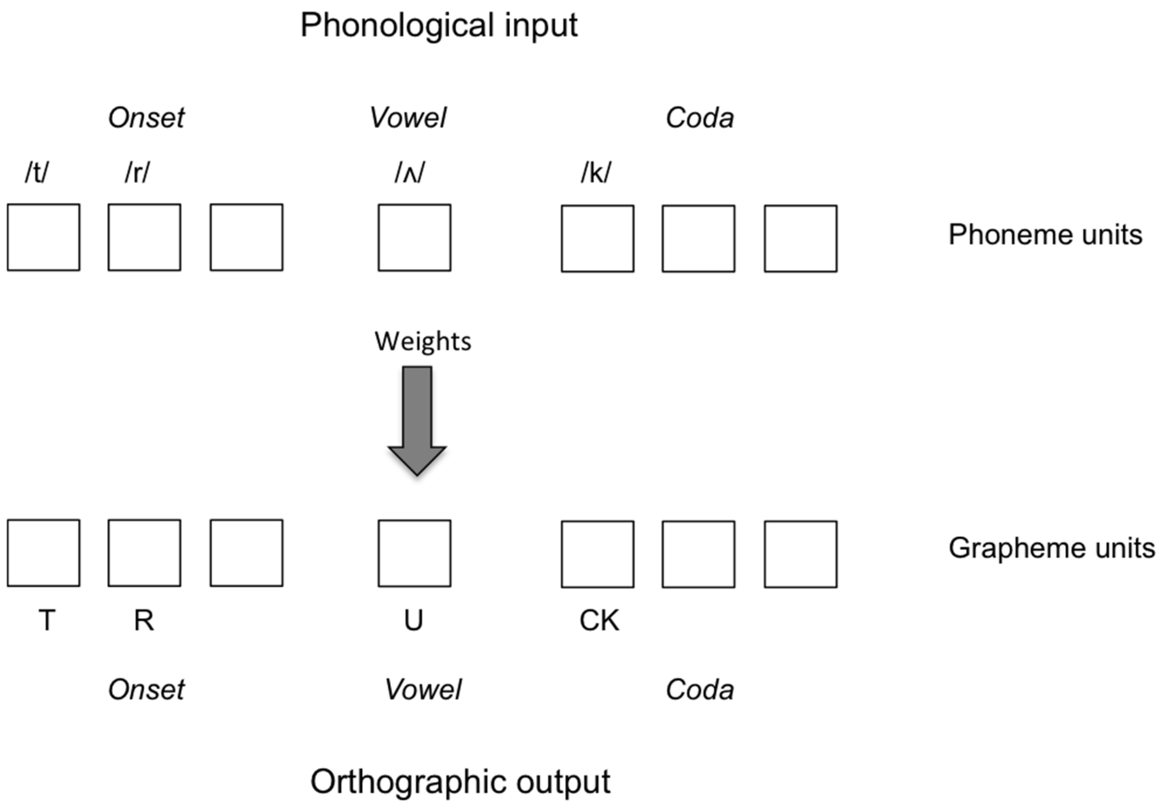


Figure 1. Representation of the route from phonological input to orthographic output (Houghton & Zorzi, 2003).

Houghton and Zorzi (2003) compared the model's spelling to human data to test a number of theoretical claims. First, the spelling of 58 non-words by the sub-lexical route working in isolation was compared with that of MP, an individual with surface dysgraphia (poor irregular word spelling and good non-word spelling; Behrmann and Bub, 1992). The spellings produced by the model were all phonologically plausible, and often identical to those produced by MP, especially in the case of non-words comprising phonemes with highly consistent spellings (e.g. /dɪmp/ -> *dimp*; /wʌʃ/ -> *wush*). When non-words were spelled that included inconsistent PGC mappings (e.g., /dri:s/, /bli:m/, /fri:tʃ/), the model generated spellings that were phonologically plausible and often, but not always, the same as those produced by MP (e.g., MP: *dreece, bleam; freech*; Model: *dreace, bleam, freech*). However, even when the model differed from MP, Houghton and Zorzi reported that MP's output was actually activated, but when competing for output with alternative spellings this response did not 'win'. Moreover, the model varied in the way it spelled inconsistent PGCs from item to item (e.g., sometimes/i:/ was EE, sometimes EA), and MP also showed variation. When the sub-lexical component of the model spelled words with either consistent (e.g., /lɒft/ as *loft*) or inconsistent phoneme-grapheme correspondences (e.g., /bi:n/ as *bean* or *been*), it was less accurate on the inconsistent words and produced phonologically plausible errors (e.g., *bene*). This reflected the sensitivity of the sub-lexical route to the frequencies of PGC mappings (in its training vocabulary). The model also needed more processing cycles to produce a response for a word comprising inconsistent PGCs (e.g., *bean*) compared to a word with consistently spelled phonemes (e.g., *loft*). This is another indication of the effect of sound-spelling consistency on sub-lexical processing: multiple plausible grapheme spellings are activated for an inconsistent (e.g., EA/EE/E.E) phoneme and compete at the level of the output. This competition has two effects: first, it takes longer for any one grapheme to be sufficiently activated to reach threshold for selection and, second, it results in response variation (the same grapheme may not always be selected for a given phoneme). It is a strength of this computational model that PGC consistency affects performance given that this has been shown to influence both normal and impaired spelling (e.g., Perry, Ziegler, & Coltheart, 2002; Treiman, Kessler, & Bick, 2002).

Houghton and Zorzi (2003) also investigated the influence of context when spelling inconsistent phoneme-grapheme correspondences. In many cases where there are multiple graphemes that can represent a single phoneme the choice of a particular grapheme is based on the phonological context. For example, after a short vowel, /k/ is mostly spelled CK (e.g., *lock*), compared to being spelled as K after a long vowel (e.g., *leak*). Houghton and Zorzi suggest that there are links between phonological syllable structure at the input and graphemic output. For example, there is a strong connection from coda /k/ to both K and CK, and short vowels have connections to CK, but inhibit K.

The mappings for vowels are more complicated than for most consonants, as most vowel phonemes have connections to more than one grapheme. For some inconsistent mappings the surrounding context can play a role. Perry et al. (2002) investigated the role of context sensitivity in the spelling of vowels in non-words. For example, the phoneme /eɪ/ at the end of an open syllable is most frequently spelled as AY (*bay*), however, in a closed syllable the most frequent spelling is A.E (*gate*). They found that adult spellers were sensitive to these context-based differences when spelling non-words (see also Treiman et al., 2002).

Given then that context-sensitivity is a feature of the sub-lexical spelling system, we were interested to determine whether this mechanism might be impaired in people with aphasia and if so, how. In particular, we hypothesised that damage to the process of converting phonemes into graphemes may result in decreased sensitivity to context. Moreover, given that inconsistent spellings are more difficult for typical adult spellers (which can also be simulated by Houghton and Zorzi's spelling model), might it be that inconsistent PGCs be especially vulnerable to damage?

The idea that PGCs differ in how difficult they are based on consistency (Perry et al., 2002) is also evident in instructional materials for spelling where often the "easier", consistent, one-to-one mappings are taught before the more complex, inconsistent (e.g., vowels such as /i:/) and context-sensitive mappings. This probable hierarchy of difficulty for PGCs has not been investigated in aphasia. Is it the case that this hierarchy influences PGC spelling in people with aphasia? Will all people with aphasia show difficulties with the same PGCs? More specifically, is there an influence of PGC consistency? Similarly, is there an effect of PGC frequency such that more frequent PGCs are less vulnerable to impairment?

Houghton and Zorzi (2003) do not specify how sub-lexical impairment could be implemented. One possibility is that certain individual mappings (e.g., between /b/ and B) could be impaired. An alternative is an overall impairment to the sub-lexical mechanism - perhaps as a result of noise to the activation of the conversion mechanism. In this case, it is likely that the more frequent or more consistent PGCs will be more resistant to impairment as they are activated more strongly. We investigate this question by comparing the pattern of impairment in people with aphasia at the PGC level of individual PGCs, and investigating the effects of frequency and consistency on spelling of a PGC.

In clinical practice, in addition to non-word spelling, spelling of individual PGSs in isolation is often also assessed (e.g., Roeltgen, Sevush, & Heilman, 1983; Shallice, 1981), particularly when considering treatment (e.g., Luzatti, Colombo, Frustaci, & Vitolo, 2000). In Houghton and Zorzi's (2003) model, spelling PGCs in isolation would mostly be expected to produce similar accuracy to spelling the same PGCs in non-words. Nevertheless, it is possible that performance on the same PGCs differs depending on whether they were spelled in isolation, where there is no effect of surrounding PGCs, or in the context of a non-word.

Intuitively, it seems as though spelling of non-words is a more complex task than spelling of PGCs in isolation. Roeltgen et al. (1981) argued that the phonological (sub-lexical) route consists of two components: segmentation and conversion. Segmentation implies the breakdown of the input into phonemes, which are then converted into graphemes. Roeltgen et al. (1981) reported data from four participants with acquired dysgraphia who were unable to write non-words, but two of these participants were able to write single graphemes when individual phonemes were dictated to them. Roeltgen et al. argued that for these patients the PGC conversion system itself was relatively intact, but the segmentation component was impaired, a pattern also found by Bolla-Wilson, Speedie, and Robinson (1985). This pattern is hard to account for within Houghton and Zorzi's (2003) computational model, which has no requirement for segmentation as phonemes are activated in parallel at input. Given the relative paucity of systematic investigation of this issue, we examined the relationship between spelling PGCs in isolation and in non-word spelling across a case series of people with aphasia and controls.

In sum, the current study investigates the nature of sub-lexical spelling impairment in acquired dysgraphia. We use data from case series of 13 people with aphasia and 13 unimpaired adults to investigate three areas of sub-lexical spelling: 1) The relationship between spelling PGCs in isolation and in non-words; 2) The effects of consistency and frequency on spelling of PGCs; and 3) the use of phonological context when spelling a vowel in a non-word. We will focus on the data from people with aphasia.

Participants

People with aphasia*.* People with aphasia (PWA) were recruited via the database of the Aphasia Research Group at Macquarie University, and through referrals from speech pathologists. We did not adopt specific inclusion criteria regarding the type of (spelling) impairment, and tested people with aphasia who were sufficiently able to read and write to complete the majority of the tasks. All subjects were paid for participation. 18 people with aphasia were tested on a number of lexical and sub-lexical spelling and reading tasks. We excluded five people who were unable to complete the majority of tasks. Data from the remaining 13 people with aphasia (two females) were included in the analyses. Their age ranged from 32-75 years, with a mean age of 58 years and 6 months, and an average of 14 years of education. All had suffered a stroke that resulted in aphasia. All individuals reported no difficulties in learning to read or spell. Further background information can be found below in Table 1. All individuals show characteristics of phonological dysgraphia, with spelling of non-words more impaired than spelling of words.

Control subjects. We tested 13 unimpaired adults (9 females) as control subjects. Controls were partners of the people with aphasia, or individuals recruited via a Macquarie University adult participant database. All subjects were paid for participation. Their age ranged from 35-77 years, with an average age of 64 years and 11 months, and an average of 16 years of education. All individuals reported no difficulties in learning to read or spell.

Table 1

Background information people with aphasia

Partici-pant	Gen-der	Age	Years of educa-tion	Time post onset (years; months)	Aphasia Severity	Fluency	Spoken Compre-hension	Anomia	Repetition on words [1] (%)	Repetition on non-words [2] (%)	Reading words [3] (%)	Reading non-words [4] (%)	Spelling words [5] (%)	Spelling non-words [6] (%)
LIG [a]	m	59	14	1;1	Mild	Fluent	Good	Mild	100	86	98	Disc.	85	70
WNO	m	69	18	1;0	Mild	Fluent	Good	Mild	97	Disc.	Disc.	Disc.	Disc.	30
BRT	m	52	16	3;0	Mild-Moderate	Non-fluent	Good	Mild	85	77	90	20	65	23
RYT	m	73	8	4;6	Mild-mod.	Fluent	Mild imp.	Mild	87	66	75	17	73	16
RAP	m	72	9	9;6	Moderate	Non-fluent	Good	Mild	95	55	92	40	78	9
GEC	m	69	18	4;8	Mild-mod.	Non-fluent	Good	Moderate	97	No data	95	28	23	5
JAC	f	36	13	21;4	Mild	Fluent	Good	Mild	93	81	77	10	75	5
PEH	m	56	16	0;6	Mild	Fluent	Good	Mild	92	No data	85	34	95	2
DEH	m	67	13	8;5	Moderate	Non-fluent	Good	Moderate	No data	No data	45	8	80	2
JOD [b]	m	75	11	16;8	Mild-moderate	Fluent	Good	Mild	93	50	87	10	38	0
JOT	m	43	15	0;5	Severe	Non-fluent	Mild imp.	Moderate	20	12	Disc.	Disc.	75	0
REA	f	32	16	6;1	Mild	Fluent	Good	Mild	95	46	95	11	85	0
HEO [b]	m	61	16	4;2	Mild-mod.	Fluent	Mild imp.	Mild-mod	67	11	78	34	93	Disc.

Note. Classifications based on Speech Pathology reports and subsequent assessments by the Aphasia Research Group at Macquarie University; Mild-Mod=Mild to moderate; mild imp. = mild impairment; Disc = Discontinued due to difficulties with the task. [a] Letter-by-letter reader. [b] JOD and HEO are left-handed.

[1] Items from PALPA 35 Reading words x regularity (n=60) (Psycholinguistic Assessments of Language Processing in Aphasia (PALPA); Kay, Lesser, & Coltheart, 1992); [2] Items from Diagnostic Spelling Test – nonwords (n=74) (DiSTn, Kohnen, Nickels, & Castles, 2009);.[3] PALPA 35 Reading words x regularity (n=60); [4] Diagnostic Reading test for nonwords (DiRT, Colenbrander, Kohnen, & Nickels, 2011) (n=104); [5] PALPA 44 Spelling x regularity (n=40) (Kay, et al., 1992); [6] Diagnostic Spelling Test – nonwords (short version, n=43) (DiSTn, Kohnen, Nickels, & Castles, 2009);

Tests

Participants were administered two sub-lexical spelling tasks as part of a larger reading/spelling battery.

Diagnostic Spelling Test – sounds (DiSTs). The Diagnostic Spelling Test for Sounds (DiSTs; Kohnen, Nickels, & Castles, 2009) tests the ability to translate sounds into letters. This test consists of 32 sounds which are spelled to dictation, starting with mappings to single graphemes (e.g., /b/ as B, /ɪ/ as I), and ending with digraphs (i.e., /ŋ/ as NG, /tʃ/ as CH). A response was scored as correct if the most frequent grapheme used to represent a phoneme. Frequencies are based on a list of phoneme-grapheme correspondences used in a study by Perry et al. (2002). For the majority of items, there was only one correct response, but in the case of multiple high frequency spellings more than one plausible correct response was accepted (e.g., EE and EA for /i:/).

Diagnostic Spelling Test – nonwords (DiSTn). The Diagnostic Spelling Test – non-words (DiSTn; Kohnen et al., 2009) assesses PGCs in the context of non-word spelling. This test consists of 74 monosyllabic non-words. Each of 40 English phonemes occurs at least twice. For example, the sound /ɒ/ is tested in the items *fot, lont* and *ponk*. The test is also available in a shorter version with 43 items (where each PGC is still tested at least twice[2]). As this version was completed by more participants we included data from this version in the majority of analyses.

We scored both item and PGC accuracy, and applied two scoring criteria. The first set of scoring criteria (Scoring A) was based on the original developmental version of the test (Kohnen, Colenbrander, Krajenbrink, & Nickels, in press). Following this scoring system, a response was scored as correct when the spelling was an accurate representation of the sounds in the non-word in the right order. Accuracy of the correspondence between sounds and letters was based on frequency counts of position

[2] The phoneme is /e/ is an exception as it is assessed only once on the short version of DiSTn.

specific sound-letter mappings for English words (Perry et al., 2002). In addition, responses produced by at least 10% of the Kohnen et al. study sample were counted as correct.

We also applied a second, more lenient scoring system, Scoring B. Here, a spelling was considered accurate when the sounds were represented correctly as in Scoring A. However, in contrast to scoring A, frequency counts were not considered. For example, using the norm based criteria from Scoring A the item *vack* has only one correct spelling, whereas with Scoring B the plausible and unambiguous spellings *vac* and *vak* were also scored as correct. Orthographically illegal spellings that do not (or almost never) occur in that context in English (e.g., *bloi* for *bloy*) and ambiguous PGCs were scored as incorrect. A spelling was considered ambiguous when the spelling was more frequently used to represent a different sound. For example, the vowel /u:/ in the item *coove*, is sometimes spelled with the grapheme OU (as in *soup*). However OU is more frequently used to represent the sound /aʊ/ (as in *trout*) and therefore was considered incorrect for /u:/ in *coove*. Responses given by more than half of the control participants in the present study were also counted correct. For example, 54% of the control sample spelled /zaɪ/ (target: *zy*) as *zai*, which is plausible (*thai*) but ambiguous (*rain*).

In addition to item accuracy, the accuracy of individual PGCs was evaluated. For example, the response *fod* would be an incorrect representation of /fɒt/, and therefore be scored incorrect at the item level. Nevertheless, the PGCs F and O are still represented correctly and therefore correct in the PGC scoring. Spelling accuracy of individual PGCs was scored based on scoring B. Hence, to be considered correct, a sound had to be translated into a plausible spelling (but not necessarily unambiguous or contextually accurate). Insertions of incorrect letters did not necessarily affect the accuracy of individual PGCs. For example, *coove* spelled as *croove* is incorrect at the item level, but this spelling resulted in a correct score for all individual sounds in the non-word /c/, /u:/, /v/. However, if the inserted letter resulted in a different PGC, this PGC was not credited: When *coove* is spelled as *choove*, the insertion of the H results in the PGC CH, and therefore spelling of C was counted incorrect.

Each PGC was assessed between two (e.g., /aʊ/) and seven times (e.g., /l/) on the DiSTn. We calculated PGC accuracy as an average of all test items a PGC occurred in. For example, if an individual spelled /ɒ/ correctly on two items, but incorrectly on the third, the mean accuracy for /ɒ/ would be 0.67.

Results

1. Overview of sub-lexical spelling performance: Controls and PWA. Table 2 displays the average group scores for the control subjects and the people with aphasia for spelling sounds in isolation and spelling non-words.

Table 2

Accuracy in percentage on spelling sounds in isolation (DiSTs) and item accuracy for non-words (DiSTn)

Test	Controls n=13		PWA n=13	
	M (SD)	Range	*M* (SD)	Range
Spelling sounds (DiSTs) (n=32)	84.1 (6.0)	71.2-90.1	67.3 (10.6)	37.5-81.3
Spelling non-words (DiSTn)				
- long (n=74)				
scoring A	72.0 (11.2)	51.4-86.5	9.31 (9.51) [a]	0-28.4
scoring B	81.4 (10.1)	62.2-94.6	10.7 (10.7) [a]	0-32.4
- short (n=43)				
scoring A	71.0 (12.1)	46.5-86.0	13.56 (20.30) [b]	0-69.8
scoring B	80.1 (10.7)	60.4-95.3	15.5 (21.02) [b]	0-72.1

[a] n=9 participants. [b] n=12 participants

Spelling sounds (DiSTs). As would be expected, the group of controls were significantly more accurate on this task than the group of people with aphasia (see Table 2) (related t-test, $t(31) = 4.51$, $p < .001$). The majority of people with aphasia (85%; 11/13) scored significantly below the control group (using Crawford's SINGLIMS comparison: Crawford & Garthwaite, 2002). See Table 3 for individual participant data and Appendix B for group data for the individual test items.

Table 3

PWA individual raw scores (accuracy) spelling sounds and spelling non-words

	Spelling sounds(DiSTs) (n=32)			DiSTn (n=74) Scoring A			DiSTn (n=74) Scoring B			DiSTn short (n=43) Scoring A			DiSTn short (n=43) Scoring B		
Control group mean (SD)	*26.92 (1.93)*			*53.31 (8.28)*			*60.23 (7.44)*			*30.54 (5.22)*			*34.77 (4.62)*		
PWA Group mean (SD)	*21.54 (3.38)*			*6.89 (7.04)*			*7.89 (7.93)*			*5.83 (8.72)*			*6.67 (9.04)*		
	score	*t*	*p*	score	*t*	*p*	score	*t*	*p*	score	*t*	*p*	score	*t*	*p*
LIG	25 [a]	-0.959	.178	n/a	n/a	n/a	n/a	n/a	n/a	30 [a]	-0.1	.461	31 [a]	-0.786	.223
WNO	26 [a]	-0.459	.327	n/a	n/a	n/a	n/a	n/a	n/a	13	-3.238	.004	14	-4.332	<.001
BRT	23	-1.957	.037	21	-3.76	.001	24	-4.692	<.001	10	-3.792	.001	12	-4.749	<.001
RYT	21	-2.956	.006	14	-4.575	<.001	16	-5.729	<.001	7	-4.346	<.001	9	-5.375	<.001
RAP	22	-2.456	.015	8	-5.273	<.001	10	-6.506	<.001	4	-4.899	<.001	6	-6.001	<.001
GEC	21	-2.956	.006	4	-5.739	<.001	3	-7.412	<.001	2	-5.269	<.001	2	-6.835	<.001
JAC	23	-1.957	.037	10	-5.04	<.001	10	-6.506	<.001	2	-5.269	<.001	2	-6.835	<.001
PEH	24	-1.458	.085	2	-5.971	<.001	3	-7.412	<.001	1	-5.453	<.001	2	n/a	n/a
DEH	20	-3.455	.002	3	-5.855	<.001	3	-7.412	<.001	1	-5.453	<.001	1	-7.044	<.001
JOD	21	-2.956	.006	1	-6.088	<.001	2	-7.542	<.001	0	-5.638	<.001	1	-7.044	<.001
JOT	21	-2.956	.006	n/a	n/a	n/a	n/a	n/a	n/a	0	-5.638	<.001	0	-7.252	<.001
REA	12	-7.449	<.001	0	-6.204	<.001	0	-7.801	<.001	0	-5.638	<.001	0	-7.252	<.001
HEO	21	-2.956	.006	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Note. Participants ordered by accuracy on DiSTn (short). Individual scores are compared to control sample using Singlims comparisons (Crawford & Garthwaite, 2002). All *p*-values are one-tailed. [a] Score does not differ significantly from control sample. All other scores significantly different at $p < .05$ n/a: score not available, participant did not complete all items on this test.

Spelling non-words (DiSTn). Nine people with aphasia finished the full version of this test. Three more participants completed the short version (43 items) and one participant also discontinued on the short version because of difficulties with the task demands[3]. As would be expected, the group of controls scored higher on this task than the people with aphasia. In addition, all nine people with aphasia who finished the long version of the non-word spelling task (n=74 items) scored significantly below the controls (based on Crawford & Garthwaite's SINGLIMS comparison; see Table 3) for both scoring system A and B. All but one of the 12 people with aphasia who completed the 43-item short version also scored significantly below controls for scoring system A and B.

In summary then, most people with aphasia showed considerable difficulty on both sound and non-word spelling tasks. However, the group averages mask considerable individual variability with some people with aphasia spelling single sounds at control accuracy but almost all people with aphasia spelling non-words below control accuracy.

PGCs in isolation versus in context. First we compare the mean accuracy for the 32 sounds (from the DiSTs) spelled in isolation to the mean score for these same PGCs when spelled in the context of a non-word (on the DiSTn). We used the 43-item version of the DiSTn as more participants with aphasia completed this version than the 74-item version. For each PGC we calculated the mean accuracy per individual and the groups (see Table 4).

Interestingly the two groups showed the opposite pattern: The control group was more accurate at spelling the 32 sounds in the context of a non-word (95%) compared to spelling these in isolation (84%: related *t*-test, $t(31) = 2.62$, $p = .007$, one-tailed),

[3] This participant (HEO)'s data is therefore not reported in group accuracies but the incomplete dataset is used in subsequent analyses.

whereas the people with aphasia as a group were more accurate spelling the sounds in isolation (67%) than in non-words (55%: related *t*-test, $t(31) = 2.62$, $p = .007$, one-tailed).

Why might people with aphasia show a different pattern from controls? Spelling non-words to dictation requires not only knowledge of the sound-letter rules in the non-word, but also requires the graphemes to be kept active in the output buffer while the word is being spelled. Therefore a comparison of performance on DiSTs and DiSTn does not only test the knowledge of PGCs. It may be that people with aphasia performed worse when spelling PGCs in the context of a non-word, because of graphemic output buffer impairments. These impairments should affect spelling of final graphemes as they have to be kept active the longest. In contrast, spelling of initial graphemes should be less affected. Hence, we examined whether people with aphasia still show a difference between spelling sounds in isolation and spelling non-words, when orthographic working memory demands are reduced, by examining only initial graphemes. This would be a more precise reflection of PGC knowledge when tested in the context of a non-word compared to isolation.

PGCs in isolation versus in initial position in non-words.

We included only non-words where the particular sound was tested in initial position (and was not part of a cluster) and used the same scoring criteria as in the previous analysis. In addition, eight vowels and the grapheme NG were excluded because they only occurred in medial or final positions resulting in a reduced set of 23 items compared to the previous analysis (see Table 4).

Table 4

Mean group accuracy for subset of 23 DiSTs sounds compared to spelling these in non-word across all positions or single initial position from DiSTn

			Controls			PWA		
Sound	Target letter	Number of non-words PGC was tested in	DiSTs	DiSTn initial (single item)	DiSTn all positions	DiSTs	DiSTn initial (single item)	DiSTn all positions
b	B	4	1.00	1.00	0.98	1.00	0.92	0.59
d	D	5	1.00	1.00	1.00	0.92	0.92	0.52
g	G	5	1.00	1.00	0.94	1.00	0.77	0.51
m	M	2	1.00	1.00	1.00	0.92	0.83	0.65
l	L	7	1.00	1.00	0.98	0.92	0.69	0.63
p	P	3	1.00	1.00	0.97	0.92	0.85	0.54
n	N	4	1.00	1.00	1.00	0.85	0.85	0.59
f	F or PH	4	1.00	1.00	0.98	0.92	0.92	0.52
t	T	5	0.92	1.00	0.98	1.00	0.92	0.59
s	S	4	0.85	1.00	0.95	0.92	0.69	0.56
h	H	2	1.00	1.00	1.00	0.85	0.92	0.88
z	Z	5	0.92	1.00	0.94	1.00	0.54	0.43
j	Y	3	1.00	1.00	1.00	0.46	0.69	0.64
v	V	3	1.00	1.00	1.00	0.77	0.77	0.59
r	R	4	1.00	1.00	1.00	1.00	0.77	0.73
ʤ	J	3	0.77	0.92	0.90	0.69	0.42	0.42
k	C or K	3	1.00	1.00	0.97	1.00	0.77	0.62
ð	TH	2	0.77	1.00	0.96	0.38	0.67	0.65
ʃ	SH	5	1.00	1.00	0.78	0.54	0.46	0.28
w	W or WH	3	0.92	1.00	0.95	0.85	0.92	0.85
kw	QU	2	0.31	1.00	0.88	0.15	0.77	0.42
θ	TH	4	0.92	1.00	0.98	0.62	0.54	0.37
tʃ	CH	3	0.92	1.00	0.88	0.38	0.62	0.38
Mean			*0.93*	*1.00*	*0.96*	*0.79*	*0.75*	*0.56*

The data show that for the group of people with aphasia, PGC spelling accuracy in a non-word is higher when based on initial than when based on all positions (0.75 compared to 0.55 when based on all 32 items: related *t*-test comparing initial instance with all positions: $t(22) = 6.99$, $p < .001$). Spelling the sounds in isolation and spelling the same sounds in initial position in a non-word is now comparable (related *t*-test $t(22) = 0.79$, $p = .220$).

At the individual level, two participants were significantly worse at spelling sounds in initial position in a non-word compared to sounds in isolation: REA (McNemar's test exact, $p = .016$) and JOT (McNemar's test exact, $p = .020$; see Appendix B). The other participants did not show a significant difference between the two tasks.

To summarise, controls were better at spelling sounds in the context of a non-word compared to spelling them in isolation. Interestingly, people with aphasia showed the opposite pattern. Follow-up analysis showed that when the possible impact of orthographic working memory is reduced, people with aphasia show significantly better PGC spelling in non-words. As a group, they also show a smaller, no longer significant, difference between spelling sounds in isolation and in the context of a non-word, however two individuals still showed poorer performance for spelling in the context of a non-word. It seems then that the impact of holding graphemes in the buffer prior to spelling contributes to the reduced performance on non-word spelling compared to spelling sounds in isolation. We will return to this issue in the Discussion.

2. Detailed analysis of PGC spelling in people with aphasia. The second part of the study focuses on the performance of people with aphasia when spelling non-words, and in particular PGC accuracy within non-words. As mentioned in the Introduction, the nature of impairment in the sub-lexical process is often unclear: does impairment affect certain PGCs while others remain (relatively) intact, or is there a general impairment in the conversion process that affects all PGCs?

Group performance. Because we wanted to investigate how impairment affects PGCs, we only included PGCs that are accurately spelled by most adults. Hence we used a set of 19 PGCs that had a control group average accuracy of at least 95% when spelled in the context of non-words (DiSTn). We excluded three PGCs that are inconsistent, that is, where multiple spellings are possible for the sound (/f/: F or PH; /w/: W or WH, /k/: C or K; /i:/ EA or EE) and the vowel /e/ which was assessed in only one DiSTn item (while all other PGCs were assessed in two or more items). As Figure 2 shows, group accuracy on these 19 PGCs spelled in non-words ranges between 0.35 and 0.88, with/h/ attracting the most and /ŋ/ and /ɑ:/ the least accurate spellings. Note, however, that group means mask a great deal of variation in individual patterns (which we will return to below).

Individual performance PWA. Figure 2 shows PGC accuracy for the 19 PGCs for each individual with aphasia spelled in non-words (DiSTn).

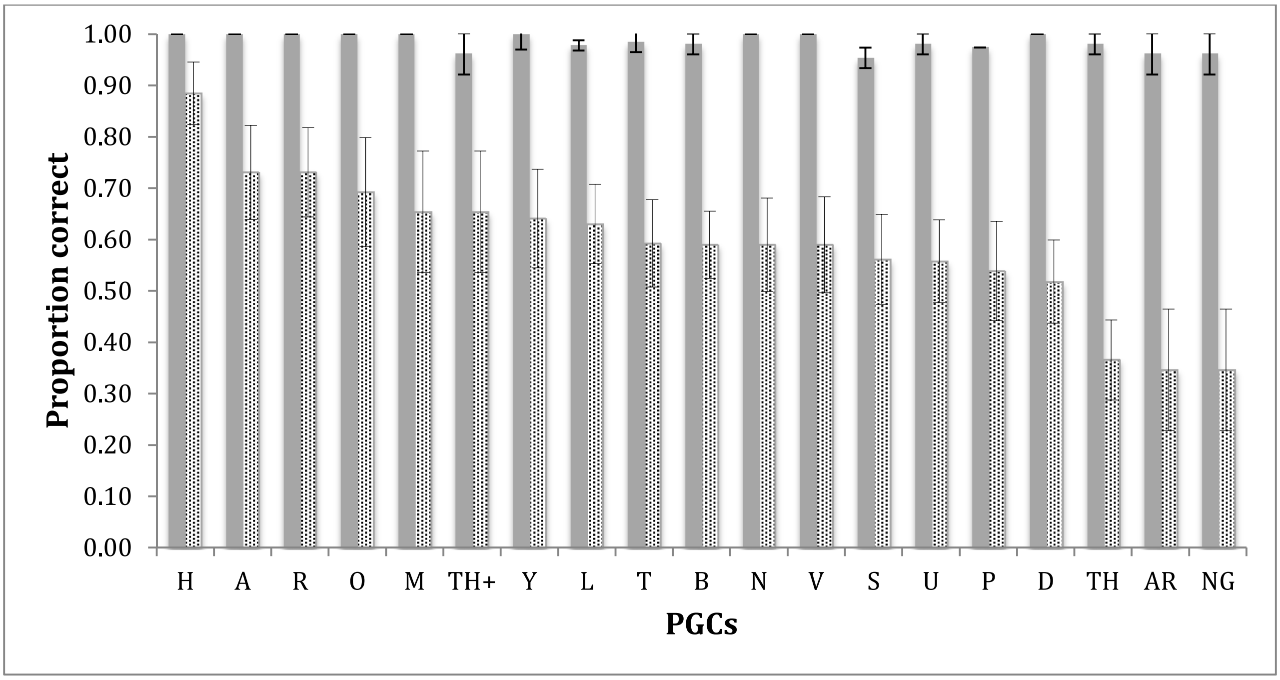
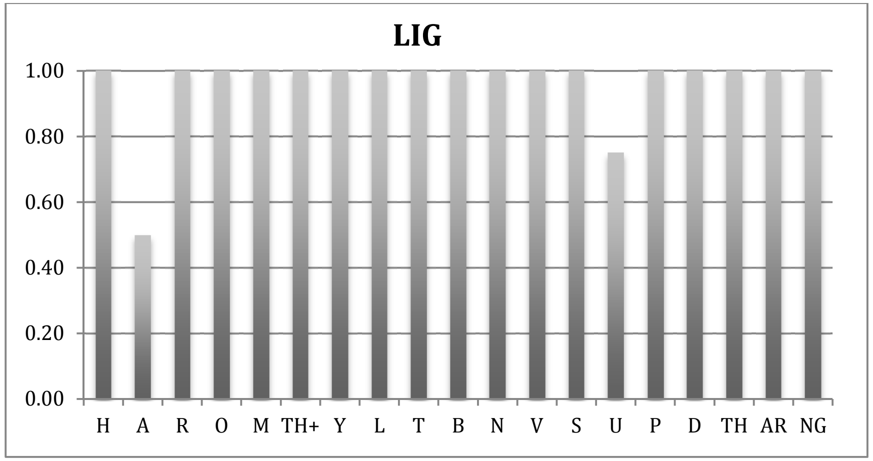
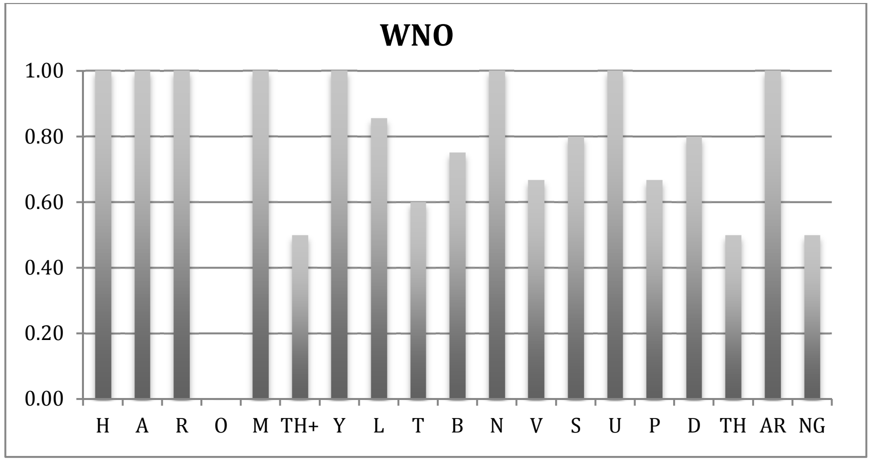
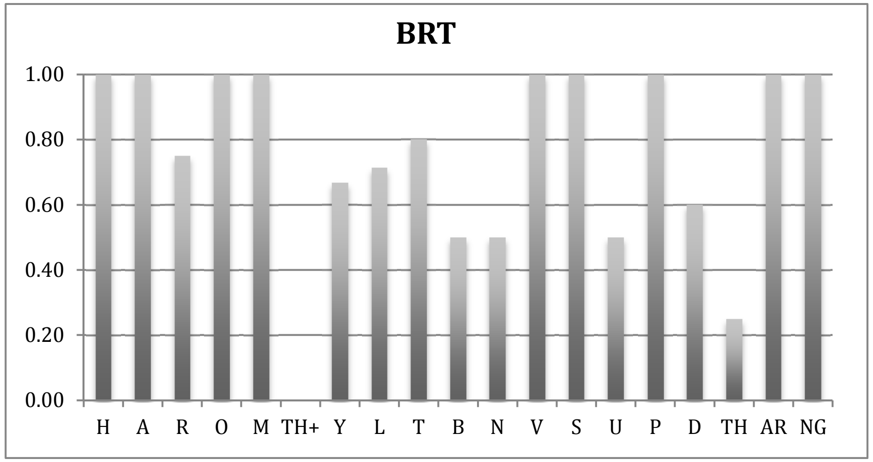


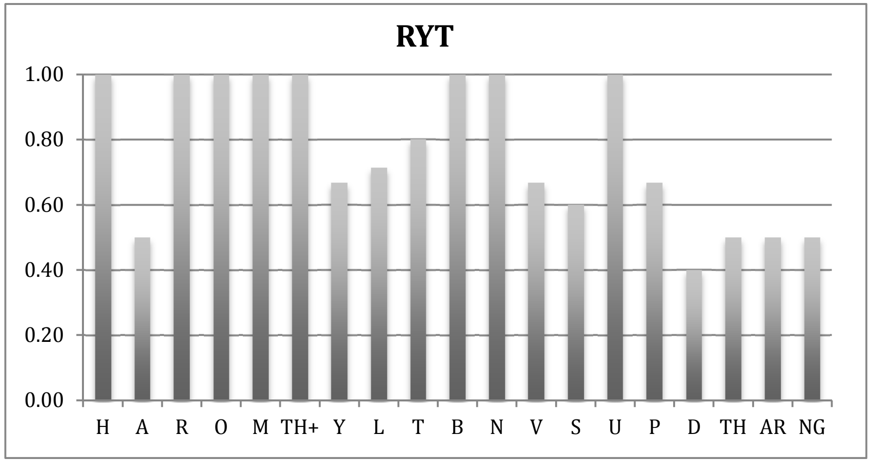
Figure 2. Group comparison on subset of 19 PGCs with ≥95% control accuracy (solid bars) on DiSTn, ordered by PWA accuracy (dotted bars) (highest to lowest accuracy). Error bars display standard errors.

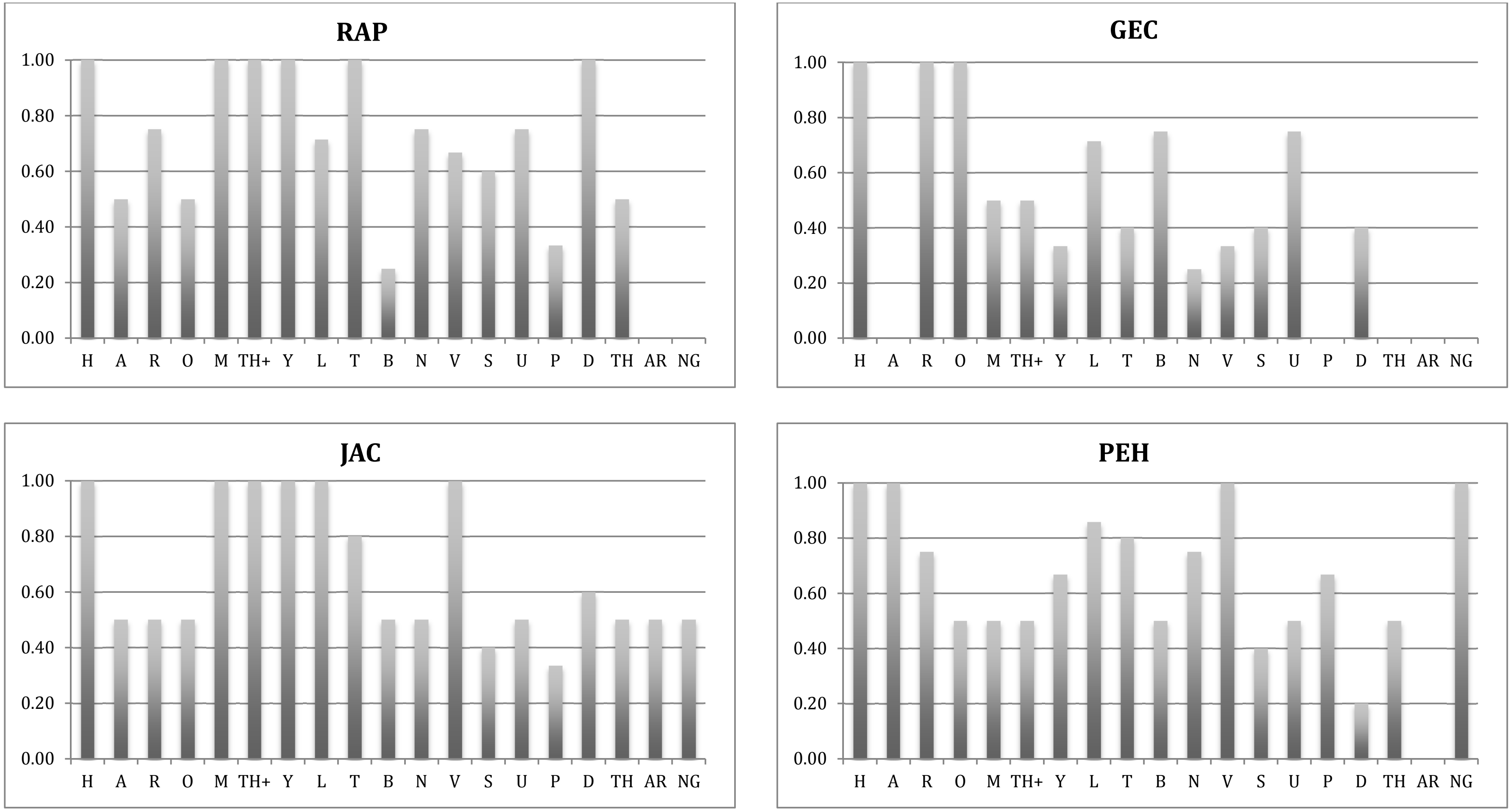
Note. The voiced /th/ sound (e.g., *that*) is marked as 'TH+', the voiceless /th/ (*thin*) as 'TH'.

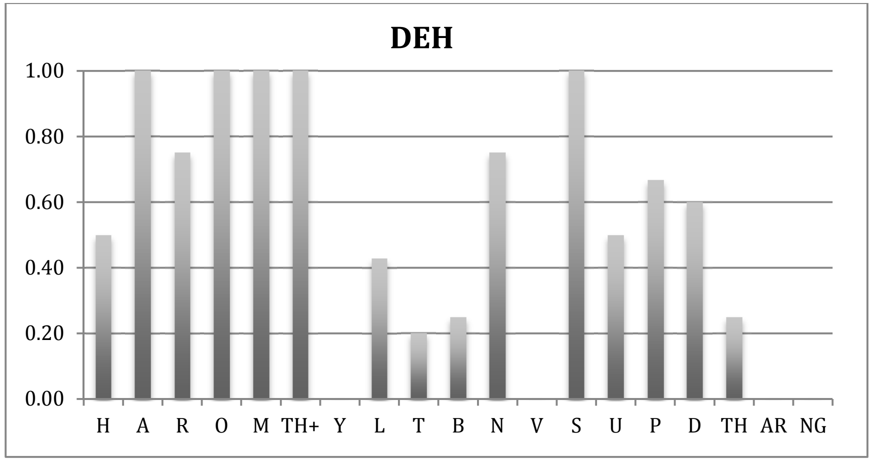

LIG
1.00
0.80
0.60
0.40
0.20
0.00
H A R O M TH+ Y L T B N V S U P D TH AR NG

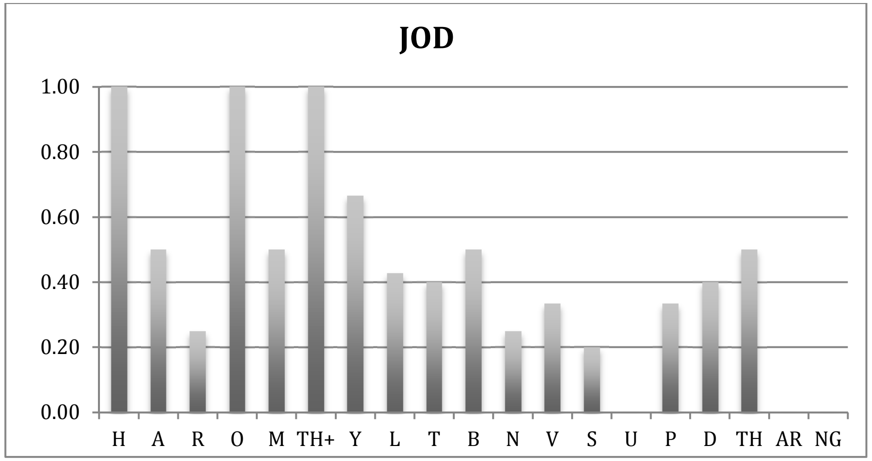

WNO
1.00
0.80
0.60
0.40
0.20
0.00
H A R O M TH+ Y L T B N V S U P D TH AR NG

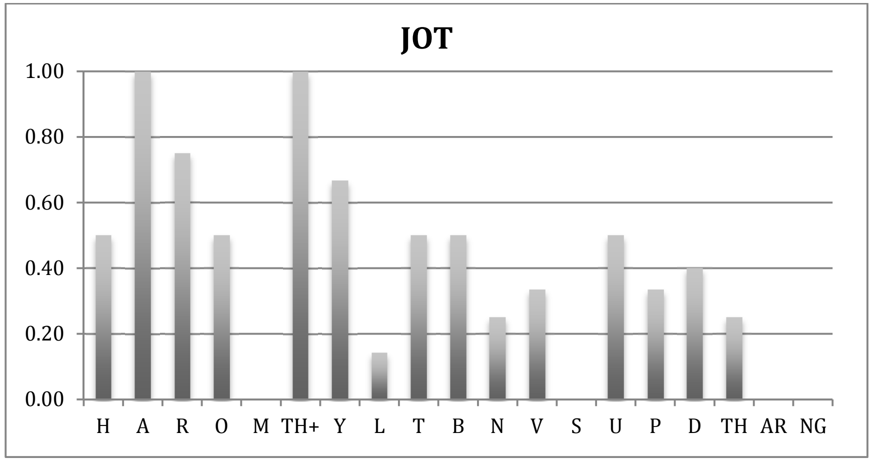

BRT
1.00
0.80
0.60
0.40
0.20
0.00
H A R O M TH+ Y L T B N V S U P D TH AR NG

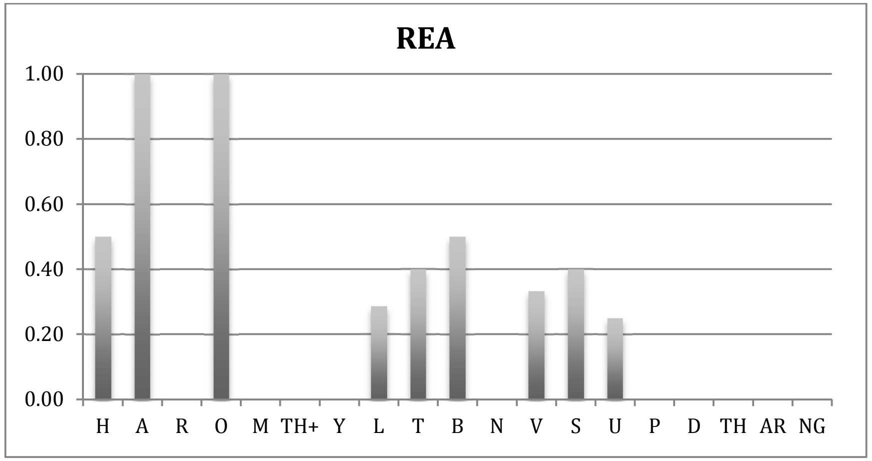

RYT
1.00
0.80
0.60
0.40
0.20
0.00
H A R O M TH+ Y L T B N V S U P D TH AR NG


RAP
GEC
JAC
PEH
1.00
0.80
0.60
0.40
0.20
0.00
H A R O M TH+ Y L T B N V S U P D TH AR NG


DEH
1.00
0.80
0.60
0.40
0.20
0.00
H A R O M TH+ Y L T B N V S U P D TH AR NG


JOD
1.00
0.80
0.60
0.40
0.20
0.00
H A R O M TH+ Y L T B N V S U P D TH AR NG


JOT
1.00
0.80
0.60
0.40
0.20
0.00
H A R O M TH+ Y L T B N V S U P D TH AR NG


REA
1.00
0.80
0.60
0.40
0.20
0.00
H A R O M TH+ Y L T B N V S U P D TH AR NG

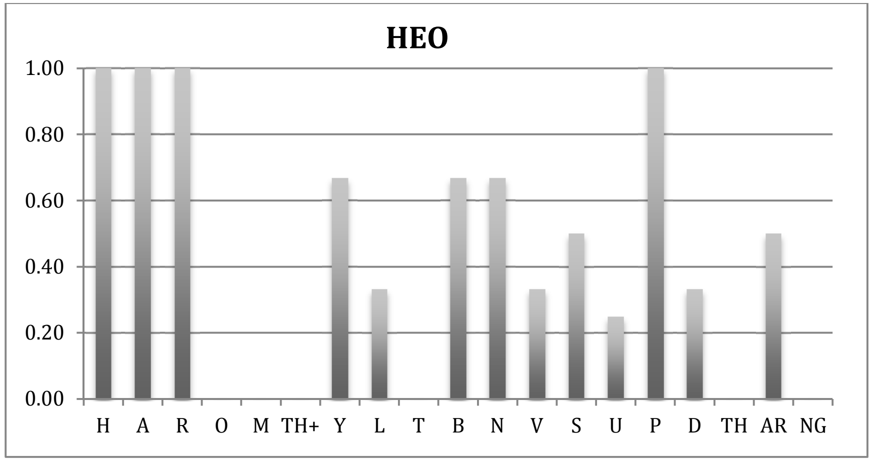


Figure 3. Individual PWA mean accuracy scores in non-words for 19 PGCs spelled with ≥95% control accuracy. PGCs are ordered by PWA mean group accuracy (highest to lowest).

Scanning Figure 3 shows the large amount of individual variation in PGC accuracy. For example, LIG scores almost at ceiling on PGCs in non-word spelling, compared to REA who scores at floor for many PGCs. Furthermore, there is a difference in the pattern of impaired PGCs across individuals. For example, as a group the people with aphasia are least accurate on the PGCs NG and AR with a mean accuracy of 0.35. However, individual scores reveal that three out of 13 people with aphasia still score 100% on NG, and four out of the 13 score 100% accurate on AR. Moreover, there is no individual who shows a pattern where all PGCs are either completely accurate or completely impaired as might be expected if impairment was characterised by rules either being 'lost' or 'retained'.

Effects of consistency and frequency on PGC spelling. Houghton and Zorzi (2003) found that frequency and consistency of PGCs influenced the performance of their computational model. Consequently we examined the influence of these factors in our data: was there a correlation between spelling accuracy and frequency and consistency of the PGC, when spelled in isolation and in the context of a non-word?

We obtained token frequency values based on consonant onset spellings and medial vowel spellings[4] for the 32 PGCs from Perry et al. (2002). In the case of multiple spellings being acceptable in our scoring (e.g., /k/ as C or K) we combined the frequency counts from both spellings.

We then calculated a consistency value for each PGC. Using the frequency counts from Perry et al. (2002), we calculated the proportion of a certain grapheme based on the total token frequency for the phoneme. For example, in initial position /g/ spelled can be spelled as G (e.g., *got*) or GH according to the Perry list. G has a token count of 123,499, and GH (e.g., *ghost*) has a token count of 576. Therefore, out of the total count

[4] There is no specific frequency count for vowels in initial position in Perry et al. (2002). However, on the DiSTn vowels were only assessed in medial and final non-word positions.

of 124,075, 99.5% of the tokens with the sound /g/ are spelled as G, and therefore /g/ -> G has a consistency value of 0.995. In the case of multiple plausible responses we combined the tokens for both spellings (e.g., F and PH). See Table 5 for group and individual data.

Table 5

Correlation coefficients PGC accuracy and consistency and frequency

	Consistency		Frequency	
	Sounds in isolation	Non-words	Sounds in isolation	Non-words
Controls (group)	.668 ***	.262	.257	.219
PWA (group)	.580 **	.248	.494 **	.640 ***
LIG	.418 *	-.034	.052	.109
WNO	.377 *	-.241	.350 *	.388 *
BRT	.468 **	-.113	.266	.279
RYT	.429 *	.019	.363 *	.361 *
RAP	.384 *	.217	.464 **	.158
GEC	.443 *	.122	.432 *	.453 *
JAC	.548 **	.387 *	.192	.193
PEH	.546 **	.378 *	.321	.226
DEH	.348	.219	.426 *	.434 *
JOD	.481 **	.398 *	.570 **	.235
JOT	.235	.159	.318	.481 **
REA	.193	-.113	.412 *	.425 *
HEO	.490 **	.118	.305	.251

* $p < .05$. ** $p < .01$. *** $p < .001$. All values are two-tailed.

Controls. As a group, controls showed an effect of PGC consistency when spelling sounds in isolation (DiSTs), but not when spelling non-words. They did not show an effect of PGC token frequency on either spelling sounds in isolation or in the context of a non-word.

People with aphasia. As a group, people with aphasia showed a correlation with consistency for PGCs in isolation, which was also significant for 10 individuals. In non-words, there was no correlation with consistency as a group, but this was significant for three individuals.

As a group, people with aphasia showed a correlation with token frequency for PGCs in isolation, and in non-words. This was significant for seven individuals (out of 13) in isolation and six individuals in non-words.

Hence, people with aphasia, PGC accuracy is affected by token frequency for both when spelling PGCs in isolation and in non-words. PGC accuracy was also associated with PGC consistency for spelling sounds in isolation but not for spelling non-words, although three individuals did show a significant effect. We will return to this in the Discussion.

3. Effects of context on sub-lexical spelling. Our final analysis investigated context dependent spelling rules. We followed previous research conducted on unimpaired individuals (Perry et al., 2002). Specifically, we investigated whether vowel spellings varied as a function of position in open versus closed syllables, and whether people with aphasia showed the same pattern as controls.

Method. To examine this question we selected vowels that differ in spelling depending on whether the vowel is in an open or closed syllable. For example, words ending in the vowel /eɪ/ are spelled AY (e.g., *say*), whereas in a closed syllable this vowel is most frequently spelled A.E (e.g., *save*). The DiSTn contains items for three vowels with this pattern (see Table 6), with two or three items per condition. We hypothesised that if participants are sensitive to position-specific spellings of the vowel, they will most commonly produce a spelling that adheres to position-specific constraints. For example, they will be most likely to produce A.E in medial and AY in word final positions, rather than, for example, spelling /eɪ/ as AY in middle positions.

Table 6

Selection of vowels in open and closed syllables

Vowel	Word examples		Items DiSTn	Items DiSTn
	Open	Closed	Open	Closed
/eɪ/	say	save	chay	chate
			snay	hafe
			tay	
/aɪ/	try [1]	bite	gly	jise
			shly	thripe
/ɔɪ/	toy	boil	bloy	thoing
			droy	zoish

[1] We chose to use the most frequent spelling using token frequency counts, as the control data seemed to suggest this was important (the majority of people spelled /glaɪ/ as *gly*). However, based on type frequency, -IE would be the most frequent spelling in this context (e.g., *pie)* (Perry et al., 2002).

Scoring. We scored accuracy of the vowel PGC in the 13 targets that included the critical vowels (see Table 6), regardless of overall accuracy of the item. Six different response categories were considered:

1) Correct: The correct and the most frequent spelling (based on token frequency) for this context (e.g., A.E in closed vowel: *hafe*)

2) Plausible (high frequency): A high frequency alternative spelling that is correct for context (e.g., *haif*);

3) Incorrect for context: A plausible spelling but one that is incorrect for this particular context (e.g., AY in closed vowel: *hayf*);

4) Plausible (low frequency): A plausible but very low frequency spelling (type frequency < 8 or token frequency < 100: e.g., *haef*)

5) Implausible: An implausible representation of the sound (e.g., *hif*);

6) Omission: An omission of the vowel (e.g., *hf*).

In the case of ambiguous responses with more than one vowel (especially from individuals with aphasia) we scored the first vowel in the response (e.g., *hafto*: the A was scored as the vowel representing /eɪ/).

A pattern of more responses of category 1 (the correct spelling and the most frequent for this context) and Category 2 (a high frequency alternative spelling that is correct for context) than Category 3 (a plausible spelling but one that is incorrect for context) would reflect context sensitive performance.

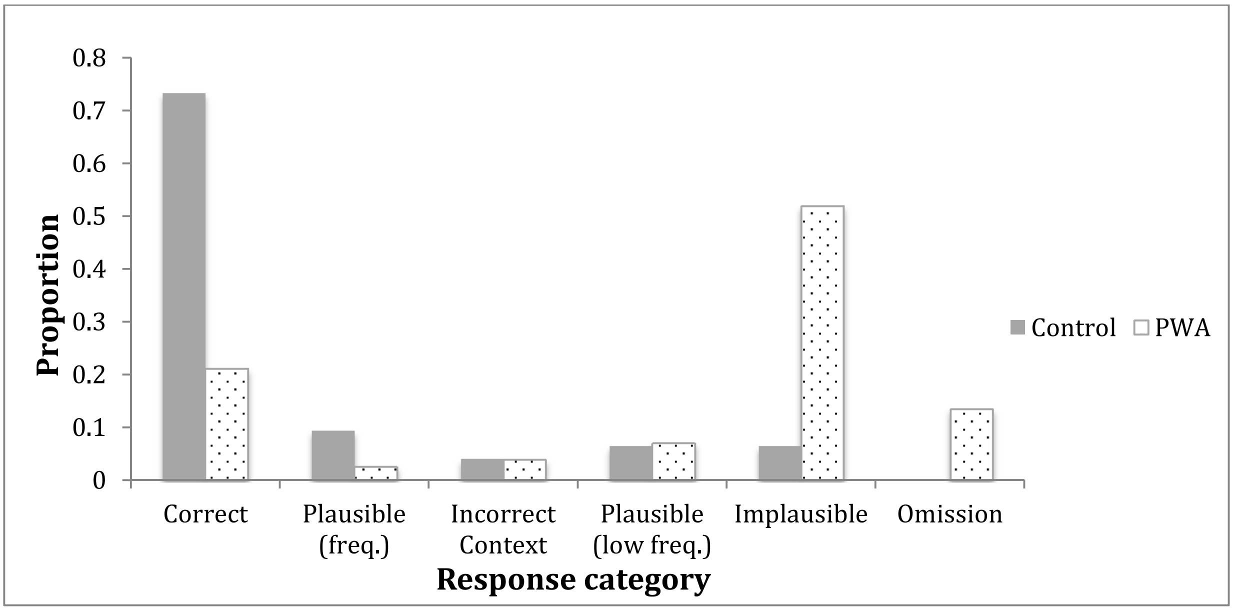


Figure 4. Distribution across different response categories for 13 vowel spellings.

Note. Freq. = frequency.

Results. The distribution of responses in Figure 4 shows that controls wrote few implausible spellings (Category 5), and most often used the spelling for the vowel that is most accurate for the context (Category 1). People with aphasia had more difficulty spelling and most frequently produced an incorrect representation of the vowel (Category 5). In order to compare whether context plays a role in the spelling of a vowel, we only looked at the responses that indicated a high frequency, plausible response for the sound (Categories 1, 2 and 3). We then compared the proportion of these three categories (see Table 7).

For controls the most frequent correct response is the spelling that is most frequent for the particular context (Category 1). The next most frequent response type is Category 2: a spelling that is less frequent but is also correct for context, followed by the response that is incorrect for context (Category 3).

The pattern for people with aphasia is similar to controls (χ (2) = 1.69, p = .430): the most frequent spelling that is correct for context makes up the majority of correct spellings, followed by Category 3 and Category 2. For the people with aphasia the number of plausible representations is smaller and therefore the difference between Category 2 and 3 is negligible. Most importantly, the most frequent correct response for both groups of participants is the spelling that is appropriate for the context. Hence, it seems that both groups are sensitive to the context of a vowel when spelling that vowel.

Table 7

Proportion of responses in Category 1, 2 and 3 of total number of plausible, high frequency spellings produced

	Total number of plausible spellings	Category 1: Correct	Category 2: Plausible	Category 3: Incorrect context
Controls	147	0.84	0.11	0.05
PWA	43	0.77	0.09	0.14

It could be the case that the choice of vowel spelling is not driven by the particular context, but by overall frequency of occurrence. For example, the spelling A.E is higher in frequency (in closed syllables; Type: 368, Token: 141,131) compared to AY (in open syllables; Type: 73, Token: 63,320). Therefore *hafe* may be used most frequently to spell /heif/ simply because A.E occurs more frequently. Is it that what appears to context-sensitive knowledge for the spellings of open and closed syllables is actually an effect of frequency?

In the case of /ɔɪ/, the spelling pattern does not seem to be driven by overall frequency. Both type and token frequency counts for OI (in closed syllables; Type: 66, Token: 19,868) are higher than the frequency counts for OY (in open syllables; Type: 15, Token: 8,083), yet controls used the spelling OI to represent the syllable-final/ɔɪ/vowel in *bloy* and *droy* only twice*,* compared to OY in 25 responses. The people with aphasia used a plausible representation for the vowel in the items *bloy* and *droy* nine times. In eight of these cases the spelling was correct for context (i.e., OY). Only once was the spelling for the closed syllable used (*bloi*). This suggests that context is sufficient to overcome any effects of PGC frequency on accuracy.

In sum, it seems that even though people with aphasia make many errors on vowels, they remain sensitive to the context of the vowel, similar to the control participants. This effect cannot be explained by overall frequency of the graphemes.

Discussion

This study into the nature of sub-lexical impairment in people with aphasia focused on three questions: 1) the performance of people with aphasia on spelling sounds in isolation and spelling sounds in the context of a non-word, and the relationship between the tasks; 2) the effects of consistency and frequency on spelling of PGCs; and 3) the use of phonological context when spelling a vowel in a non-word.

While controls performed more accurately at spelling sounds in non-words than in isolation, people with aphasia showed the reverse pattern. However, further analyses showed that when the comparison was based on initial sounds in non-words only (e.g.,

/b/ in *buv*) there was no longer an advantage for spelling sounds in isolation for the group of people with aphasia and for 11 of the 13 individuals.

We hypothesised that one of the reasons for this difference might be the increased demands on orthographic working memory required in non-word spelling compared to the spelling of sounds in isolation. Sounds in final position of non-words would be affected more than those in initial position, independent of the accuracy of the mapping of the phoneme-grapheme correspondence. Indeed, we found that when only taking spellings of initial sounds in non-words into consideration, the difference between sound and non-word spelling disappeared. Therefore, for people with aphasia the disproportionate difficulty when spelling PGCs in non-words seems at least in part to be due to the increased demand on orthographic working memory.

However, there were two individuals who continued to show more errors when spelling sounds in the context of non-words than in isolation, even when only initial position was considered. Roeltgen et al. (1983) proposed two different components within the sub-lexical route: segmentation and conversion. They suggest that impaired spelling of sounds in non-words in the context of intact spelling of individual sounds indicates an impairment to the segmentation component, with a relatively intact conversion component. It follows then that perhaps these two individuals were impaired in the segmentation component of non-word spelling, Administration of an oral segmentation task might have provided further evidence for this hypothesis, however this was not possible in the context of this study.

In summary then, our data show that people with aphasia can show poorer non-word spelling than would be predicted based on their PGC knowledge which is proposed to be due to difficulties with orthographic working memory and segmentation. This reinforces the need for assessment of both a non-word spelling task, and also spelling of individual sounds when assessing the sub-lexical route.

The second part of the study investigated individual PGC accuracy in non-word spelling on a subset of 19 PGCs which were accurately spelled by the controls. Across individuals with aphasia there was large variability in performance on individual PGCs. For example, PGCs that were poorly spelled based on group accuracy (e.g., mean accuracy of 0.35 for grapheme AR) were nevertheless spelled correctly by four individuals. Moreover individuals with a similar overall accuracy score (e.g., RYT and WNO, see Figure 3) showed impairment to different PGCs. The data clearly show that group averages mask important individual variability. Of course, this is a point that has often been used to argue for single case and case series investigations (e.g., Caramazza & McCloskey, 1988). Furthermore, the pattern of accuracy across the individuals did not seem to be consistent with a general pattern of loss of individual PGCs – most individuals showed variable accuracy within a PGC rather than a bimodal distribution of a PGC either always being correct or always incorrect. Admittedly this is not entirely straightforward, as a 'retained' PGC could still be error prone because of an additional impairment (e.g., to orthographic working memory). Nevertheless, it is hard to reject a hypothesis that the pattern of variability is consistent with noise affecting PGC processing.

Furthermore, it may be the case that the range of individual performance could explain part of the heterogeneity in the group. For example, the participants differ in the size of the difference between accuracy for spelling words compared to non-words (lexicality effect, see Table 1). It seems that participants with more damage to sub-lexical spelling processes show an impairment affecting more individual PGCs compared to individuals with a milder impairment.

To further investigate the factors influencing PGCs we examined whether accuracy of spelling sounds in isolation or in the context of a non-word was correlated with the frequency and consistency of the PGC, consistent with predictions from studies with adult spellers (Perry et al. 2002) and the Houghton and Zorzi model (2003). Consistency was significantly correlated (both at group and individual level) with spelling sounds in isolation. Three individuals also showed a significant association between consistency and spelling of PGCs in non-words but for the group overall consistency did not correlate with performance. Frequency was also correlated (both group and individual level) with both spelling sounds in isolation and spelling non-words. This suggests that that more consistent and frequent PGCs may be more resistant to damage. The data is therefore consistent with Houghton and Zorzi's (2003) model, that also showed sensitivity to the frequency of PGC mappings and the model needed more processing cycles to produce a response for a word that included an inconsistent PGC.

It is hard to explain why PGC spelling in non-words did not show a significant effect of consistency, however. It may be the case that as spelling non-words relies on additional processes, for example working memory processes, that perhaps these may mask effects of consistency on performance. Another possibility is our measure of consistency. We used token frequencies based on initial position for consonants and medial positions for vowels as listed by Perry et al. (2002). These were chosen as the best approximation for sounds in isolation, and then also used to examine the consistency in non-words. However, for the spelling of non-words position in the non-words may influence consistency and therefore also examining consistency using different measures for the PGC in each position may have been appropriate and perhaps have influenced the results.

The final part of our investigation concerned the use of context in vowel spelling. We found that even though people with aphasia make a larger number of errors on vowels than controls, when the vowel is correctly represented they will often use the vowel spelling that is the most frequent for the particular context. Control participants showed the same pattern. However, the choice of spelling overall was not purely due to frequency - a context-appropriate vowel of lower frequency would be preferred to a context inappropriate vowel of higher frequency (e.g., OY for word-final spelling of /ɔɪ/). The data imply that context specific spelling mechanisms were largely intact in the group of people with aphasia in this study.

Conclusion

This study investigated the nature of sub-lexical impairment in a group of people with aphasia, through analyses of spelling of sounds in isolation and in non-words. A computational model of spelling proved helpful to direct investigations and to form specific hypotheses regarding impairment. The data found in the study support the model of a sub-lexical processing route that is used for spelling sounds in isolation and in non-words, and is sensitive to frequency and consistency of a PGC, and its phonological context. Results from the study also suggest that both sound spelling and non-word spelling are important tasks for assessment of sub-lexical processing.

References

Beeson, P. M., & Rapcsak, S. Z. (2002). Clinical diagnosis and treatment of spelling disorders. In A. E. Hillis (Ed.), *The handbook of adult language disorders: Integrating cognitive neuropsychology, neurology and rehabilitation* (pp. 101-120). New York: Psychology Press.

Behrmann, M., & Bub, B. (1992). Surface dyslexia and dysgraphia: Dual routes, single lexicon. Cognitive Neuropsychology, 9, 209–252.

Bolla-Wilson, K., Speedie, L. J., & Robinson, R. G. (1985). Phonological agraphia in a left-handed patient after a right-hemisphere lesion. *Neurology, 35*, 1778-1781.

Caramazza, A., & McCloskey, M. (1988). The case for single patient studies. *Cognitive Neuropsychology, 5,* 517–528.

Colenbrander, D., Kohnen, S., & Nickels, L. (2011). *Diagnostic Reading Test for Nonwords (DiRT).* Retrieved from www.motif.org.au.

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.

Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology, 20*(2), 115-162. DOI: 10.1080/02643290242000871.

Kay, J., Lesser, R., & Coltheart, M. (1992). PALPA: Psycholinguistic Assessments of Language Processing in Aphasia. Hove: Lawrence Erlbaum Associates.

Kohnen, S., Nickels, L., & Castles, A. (2009). Assessing spelling skills and strategies: A critique of available resources. *Australian Journal of Learning Difficulties, 14*(1), 113-150. DOI:10.1080/19404150902783450.

Kohnen, S., Colenbrander, D., Krajenbrink, T., Nickels, L. (in press). Assessment of lexical and nonlexical spelling in students in Grades 1-7. *Australian Journal of Learning Difficulties*.

Luzzatti, C., Colombo, C., Frustaci, M., & Vitolo, F. (2000). Rehabilitation of spelling along the sub-word-level routine. Neuropsychological Rehabilitation, 10, 249–278.

Nickels, L., Biedermann, B., Coltheart, M., Saunders, S., & Tree, J. J. (2008). Computational modeling of phonological dyslexia: How does the DRC model fare? *Cognitive Neuropsychology, 25*(2), 165-193.

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, 55*(A), 897-915.

Rapcsak, S. Z., Beeson, P. M., Henry, M. L., Leyden, A., Kim, E., Rising, K., Andersen, S., & Cho, H. (2009). Phonological dyslexia and dysgraphia: Cognitive mechanisms and neural substrates. *Cortex, 45*, 575-591.

Rapp, B. (2002). Uncovering the cognitive architecture of spelling. In A. E. Hillis (Ed.), *The handbook of adult language disorders: Integrating cognitive neuropsychology, neurology and rehabilitation* (pp. 47–69). New York: Psychology Press.

Roeltgen, D. P., Sevush, S., & Heilman, K. M. (1983). Phonological agraphia: Writing by the lexical-semantic route. *Neurology, 33*, 755-765.

Shallice, T. (1981). Phonological agraphia and the lexical route in writing. Brain, 104, 413– 429.

Tainturier, M. J., & Rapp, B. (2001). The spelling process. In B. Rapp (Ed.), *What deficits reveal about the human mind/brain: A handbook of cognitive neuropsychology* (pp. 263 – 289). Philadelphia, PA: Psychology Press.

Treiman, R., Kessler, B., & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language, 47*(3), 448-468.

Appendix A

Mean group scores for individual sounds (DiSTs, n=32)

Sound	Target letter	Controls	PWA
b	B	1.00	1.00
k	C or K	1.00	1.00
g	G	1.00	1.00
r	R	1.00	1.00
t	T	0.92	1.00
z	Z	0.92	1.00
æ	A	0.69	0.92
d	D	1.00	0.92
f	F or PH	1.00	0.92
l	L	1.00	0.92
m	M	1.00	0.92
p	P	1.00	0.92
s	S	0.85	0.92
h	H	1.00	0.85
n	N	1.00	0.85
w	W or WH	0.92	0.85
ɛ	E	0.77	0.77
v	V	1.00	0.77
ʤ	J	0.77	0.69
ɒ	O	1.00	0.69
θ	TH	0.92	0.62
ɪ	I	0.85	0.54
ʃ	SH	1.00	0.54
j	Y	1.00	0.46
tʃ	CH	0.92	0.38
ð	TH	0.77	0.38
i:	EA or EE	0.62	0.15
ŋ	NG	0.69	0.15
kw	QU	0.31	0.15
a:	AR	0.15	0.08
ʌ	U	0.23	0.08
u:	OO	0.62	0.08
Mean		*0.84*	*0.67*

Note. Items are listed in order of highest to lowest accuracy for the PWA.

Appendix B

Individual PWA data (raw scores) spelling sounds in isolation compared to in initial position of single non-word (n=23)

	Sounds (DiSTs)	Non-word initial position (DiSTn)
LIG	20	23
WNO	20	17
BRT	19	21
RYT	17	21
RAP	18	20
GEC	18	18
JAC	21	19
PEH	21	20
DEH	16	15
JOD	18	19
JOT [a]	18	11
REA [a]	10	4
HEO [b]	16	13

[a] REA and JOT showed a significant difference between the two scores (McNemar's test exact $p < .05$). All other participants showed no significant difference (McNemar's test exact $p > .05$)

[b] This score is based on a subset of 19 items as HEO did not complete the non-word spelling task.

STUDY THREE

Generalisation after Treatment of Acquired Spelling Impairments: A review

This paper has been published as:

Krajenbrink, T., Nickels, L., & Kohnen, S. (2015). Generalisation after Treatment of Acquired Spelling Impairments: A review. *Neuropsychological Rehabilitation, 25*(4), 503-554. DOI: 10.1080/09602011.2014.983135

Abstract

This paper provides a comprehensive review of treatment studies of acquired dysgraphia and the occurrence of generalisation after this treatment. The aim of this review is to examine what determines the occurrence of generalisation, by investigating the link between the level of impairment, the method of treatment, and the outcome of therapy. We present the outcomes of treatment with regards to generalisation in 40 treatment studies. We derive general principles of generalisation which provide us with a better understanding of the mechanism of generalisation: 1) Direct treatment effects on representations or processes; 2) Interactive processing and summation of activation; 3) Strategies and compensatory skills. We discuss the implications of these findings for our understanding of the cognitive processes used for spelling. Finally, we provide suggestions for the direction of further research into this important area, as a better understanding of the mechanism of generalisation could maximise treatment effects for an individual with acquired dysgraphia.

Introduction

Treatment studies can play an important and often undervalued role in informing cognitive theories of language processing (Nickels, Kohnen, & Biedermann, 2010). The cognitive neuropsychological approach to treatment of language impairments employs model-based assessment, where the type of treatment is chosen after a thorough analysis of the impairment, within a cognitive theoretical framework (Nickels et al., 2010). This approach can enable us to examine the link between the effects of intervention and a certain cognitive function (Hillis & Caramazza, 1994). Furthermore, a better understanding of the factors that contribute to the success of a particular type of treatment will help improve the effectiveness of rehabilitation programmes (Rapp, 2005).

One outcome of treatment that can particularly inform theory is generalisation. We use the term generalisation when treatment leads to improvements not only for treated items, but also for untreated control items, or untreated modalities (e.g., Behrmann, 1987; Rapp & Kane, 2002; Rapp, 2005; Sage & Ellis, 2006). Generalisation effects are thought to occur because improvement of a certain part of the processing system can lead to improved performance on other tasks that require this same process. For example, treating semantic processing should improve performance on both comprehension and production tasks. A better understanding of the mechanisms that drive generalisation may help us maximise treatment effects, and should contribute to our understanding of the cognitive processes and their interrelationships.

In this paper we give an overview of the cognitive neuropsychological approach to acquired written language impairments (dysgraphias[1]). A review of lexical and sub-lexical remediation studies for (central) dysgraphias is provided, with a focus on

[1] Both "agraphia" and "dysgraphia" are used throughout the literature of acquired impairments of written word production. In the current paper the term "dysgraphia" is chosen, as it is most frequently used.

generalisation. The aim of this review is to examine what determines the occurrence of generalisation, by investigating the link between the level of impairment, the method of treatment, and the outcome of therapy. We review the different explanations that have been put forward, and discuss the implications for our understanding of theories of language processing and for remediation programmes.

Cognitive Models of Spelling

Most cognitive theories of spelling to dictation propose two main processing routes (e.g., Tainturier & Rapp, 2001; Beeson & Rapcsak, 2002): the lexical route and the sub-lexical route (see Figure 1). These types of theory are the most commonly used in the acquired dysgraphia literature (e.g., Rapp, 2002; Whitworth, Webster, & Howard, 2005) and are often used as a framework for treatment studies of spelling impairment (e.g., Rapp & Kane, 2002; Schmalzl & Nickels, 2006). Consequently, this model forms the basis of the discussion here. Other theories of language and more specifically reading, such as 'the triangle model' of reading (e.g., Seidenberg, 2005), encode information from semantics, phonology and orthography, and could also apply to the process of spelling. However, these types of models are generally less clearly specified in their descriptions of the spelling process. We will address these theories in the Discussion.

The lexical route for spelling to dictation relies on the retrieval and activation of representations of whole words, and allows correct spelling of all familiar words (both regularly and irregularly spelled). Within this route, an orally presented word activates its representation in the phonological input

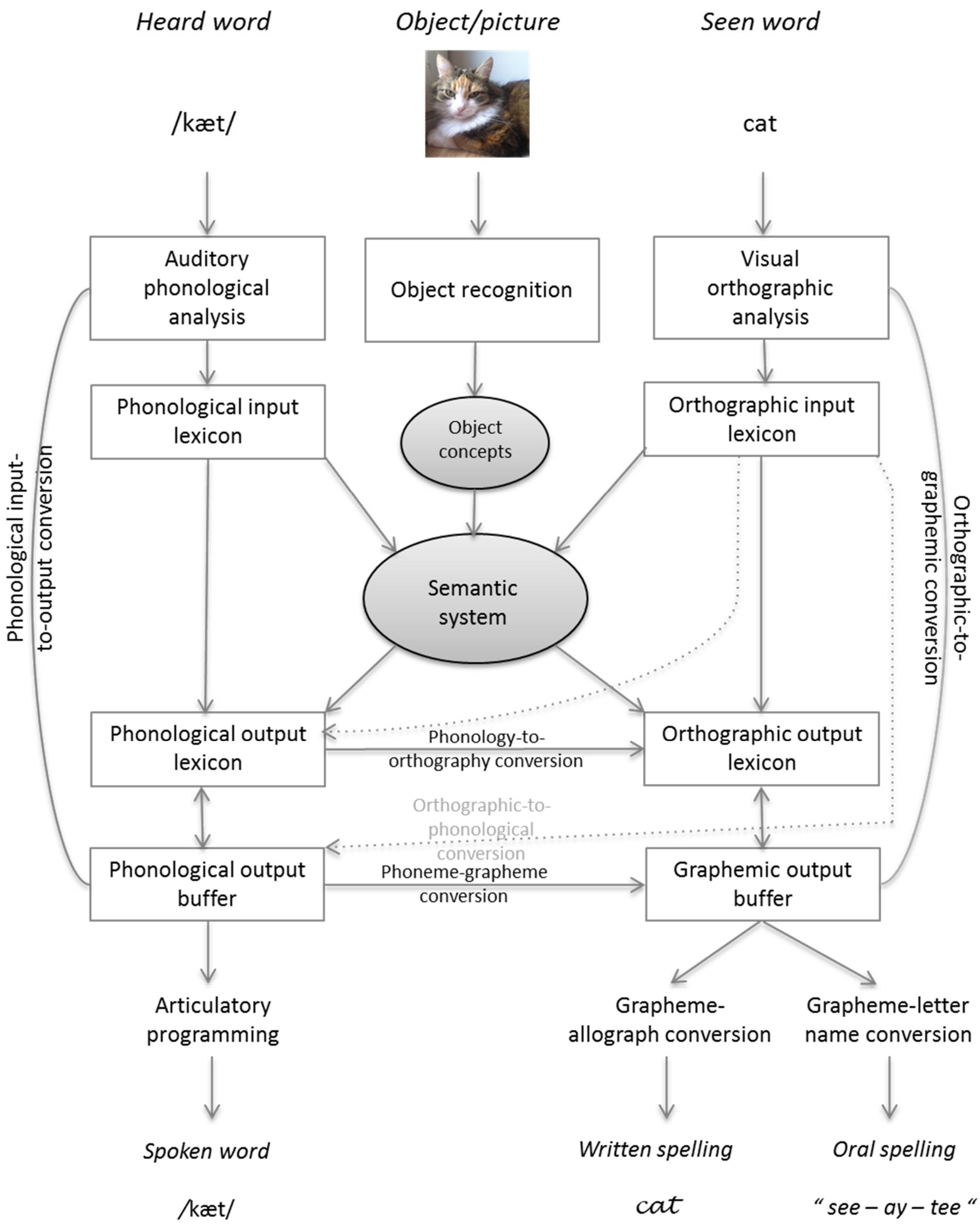


Figure 1: Schematic representation of the spelling system (adapted from Whitworth, Webster & Howard, p. 80).

Note: The solid lines depict the routes and components relevant for written word production (i.e., writing to dictation, written naming and copying). Dotted lines represent additional components involved in the process of language production.

lexicon, which consequently allows the listener to gain access to the representation of the word's meaning within the semantic system. This concept can in turn activate the representation of the word's spelling within the orthographic output lexicon, a long-term memory store of the spellings of familiar words. Written naming and spontaneous writing also uses the lexical route. In written naming, an object or picture is first visually processed and its corresponding concept is activated in the semantic system. In turn, the concept activates the representation of the word form in the orthographic output lexicon (OOL).

The sub-lexical route allows correct spelling for regular (familiar and unfamiliar) words and non-words by applying phoneme to grapheme conversion (PGC) rules. When working in isolation, this route yields erroneous but phonologically plausible spellings for irregular words (e.g. "yacht": *yot*). The abstract graphemic representations generated by either route are temporarily stored in the graphemic output buffer. This working memory component keeps the representations active while the physical process of writing is completed and can be considered the interface between central and peripheral processes. Subsequently, at a post-graphemic level, the abstract graphemic representations are converted into visual letter shapes or stroke features by the allographic conversion system (Rapp & Caramazza, 1997). Finally, peripheral writing is performed through the coordination of motor processes that allow sequences of strokes to be formed into each written letter (Goodman & Caramazza, 1986b).

In addition to the lexical and sub-lexical route, some authors have proposed a third "lexical non-semantic route" (Ellis & Young, 1988; Whitworth, et al., 2005). This route links the phonological input lexicon and the orthographic output lexicon via the phonological output lexicon by bypassing semantics.

The Nature of Spelling Impairments

As Rapp (2002) concluded, the study of spelling impairments has contributed to our understanding of how the spelling process is structured. Written word production[2] calls upon a multitude of cognitive, linguistic, and perceptual-motor processes. Disruption of these processes can be caused by damage to central components (i.e., semantics, orthographic lexicon, phoneme to grapheme conversion, graphemic output buffer) as well as more peripheral components of the spelling process (e.g., allographic conversion, letter name conversion).

Level of spelling impairment. A comprehensive assessment of a person's spelling abilities includes examination of spontaneous writing, written naming, writing to dictation, copying and oral spelling. It requires the use of controlled word lists that allow for independent evaluation of various lexical features such as word frequency, imageability, grammatical class, spelling regularity, word length, morphological complexity and, in addition, non-word spelling. Assessment of oral spelling, typing, spelling with anagram letters and copying may be indicated to discern whether central or peripheral spelling processes are impaired (Beeson & Hillis, 2001). Table 1 provides an overview of the subtypes of spelling impairments with their most important features.

[2] In the literature on this topic, the terms spelling and writing are both used to describe different aspects of "written word production". Throughout the current paper the overarching term "spelling" is chosen to describe this process.

Table 1

Subtypes of central spelling impairments

Subtype	Cognitive Impairment	Behavioural Characteristics	Example references
Surface dysgraphia	Impaired lexical route: damage to orthographic output lexicon or access to this lexicon; Sub-lexical route intact	Impaired spelling of irregular words relative to regular words; Relatively spared non-word spelling.	Beauvois & Derousne, 1981; Behrmann & Bub, 1992; Goodman & Caramazza, 1986a; Weekes & Coltheart, 1996
Phonological dysgraphia	Impaired sub-lexical route: damage to sub-lexical spelling procedures; Intact lexical route	Impaired non-word spelling; Relatively spared word spelling	Shallice 1981; Roeltgen, Sevush, & Heilman, 1983
Deep dysgraphia	Both lexical and sub-lexical route impaired; Spelling via impaired lexical-semantic route	Impaired picture naming, writing to dictation with semantic errors; Poor non-word spelling.	Bub & Kertesz, 1982; Hillis, Rapp, & Caramazza, 1999
Graphemic Output Buffer dysgraphia	Impaired graphemic output buffer: other components of lexical and sub-lexical routes relatively spared	Word and non-word spelling both impaired with similar error types; Effect of length on spelling accuracy. Error types are letter substitutions, deletions, additions, transpositions.	Caramazza, Miceli, Villa & Romani, 1987; Miceli, Silveri, & Caramazza, 1985

However, it should be noted that "pure" cases of dysgraphia, that clearly fit the criteria for one of the subtypes of impairment mentioned above, are relatively rare. Moreover, each subtype is in fact a syndrome, where the same symptoms can be the result of different underlying impairments. For example, surface dysgraphia can be the result of different impairments, including (1) impaired access the orthographic lexicon,

(2) an impairment of the representations in the orthographic lexicon itself, or, (3) poor connections from the semantic system to the orthographic lexicon. Thus, when defining an individual's spelling impairment, labels such as "surface dysgraphia" are shorthand terms, and it is often more appropriate to describe which components of the spelling system are impaired.

Treatment Approaches. The identification of the pattern of impaired and intact components of the spelling system (and often also the status of reading and spoken language) is important when planning an intervention. Treatment can focus on compensation mechanisms or relay strategies (e.g., Hatfield, 1983), or can focus on improving impaired processes, either lexical or sub-lexical or shared (e.g., graphemic output buffer). See Table 2 for an overview of different treatment techniques. It is important to note that other treatments, that do not have spelling as their main focus, may nevertheless improve spelling. For example, a treatment that successfully improved semantic processing would result in improved written (and spoken) naming. However, in this review we restrict ourselves to studies where there was an explicit focus on spelling.

Table 2

Overview of different types of treatment used with people with a central dysgraphia (adapted from Beeson, 2004)

Impairment	Possible treatment	Rationale	References
Lexical route	Repeated training of specific set of words, in writing to dictation or copying (e.g., CART: Copy and Recall Treatment)	1) Improve access to specific orthographic representations; 2) improve quality of representations; 3) restore representations	Behrmann, 1987; Schmalzl & Nickels, 2006
	Strategic use of sub-lexical skills	Using preserved sub-lexical route to facilitate retrieval of correct spelling, and training to self-correct phonologically plausible errors (possibly in combination with spell-checker)	Beeson, Rewega, Vail, & Rapcsak, 2000
Both lexical and sub-lexical routes	Repeated training of specific set of words, in writing to dictation or copying (e.g., ACT: Anagram and Copy Treatment)	1) Improve access to specific orthographic representations; 2) improve quality of representations; 3) restore representations	Beeson, Hirsch, & Rewega, 2002; Beeson, 1999
	Combining lexical and sub-lexical skills	Maximising the use of residual orthographic information combined with phoneme to grapheme correspondences	Beeson, Rewega, Vail, & Rapcsak, 2000
	Strengthening or relearning PGC rules (may include personalised key words to learn PGCs ("K as in Kate"))	Sub-lexical training aims to facilitate use of PGCs to help retrieve correct (lexical) spelling	Hillis Trupe, 1986; Luzzatti, Colombo, Frustaci, & Vitolo, 2000; De Partz, Seron, & Van der Linden, 1992
Graphemic output buffer	Copy and recall	Strengthening lexical representations to become more resistant to impact of impaired graphemic output buffer (similar to CART / ACT protocols	Rapp 2005; Rapp & Kane 2002
	Segmentation	Segmenting long words into smaller units (to circumvent length effect and decrease working memory load)	De Partz, 1995
Various impairments	Voice recognition dictation software	Using intact spoken language abilities to assist writing process	Bruce, Edmundson, & Coleman, 2003

Note. PGC = Phoneme-to-Grapheme Correspondence.

Generalisation after Treatment

Whatever the focus of the treatment, if it is successful, at the very least, the items practiced during treatment should have improved. When improvements are restricted to these treated items, this is referred to as an "item-specific effect". In lexical treatments, this improvement would be interpreted to reflect either (1) improved access to the treated words' representations in the orthographic output lexicon, or (2) stronger or more complete representations, or (3) re-learning of representations.

All but one of the treatment studies that we review below resulted in improved spelling of treated items for all participants. This means that it is possible to strengthen or improve access to specific orthographic representations in people with acquired spelling impairments, even for those with very limited writing skills or many years post onset (e.g., 24 years, Kiran, 2003). Sometimes, in addition to the improvement of specifically treated items, the effects of treatment are found to generalise. Generalisation effects can be informative for a number of theoretical issues which will be addressed below.

Theoretical issues under debate. Cognitive models similar to the one in Figure 1 have been widely used as a framework to describe patterns of language impairment in reading or spelling. Even though there is a relative consensus about the basic levels of processing, some issues concerning the nature of specific processes or the interaction between processes have been raised and are still open to debate. In their discussion, Tainturier and Rapp (2001) addressed a number of critical questions concerning the nature of the spelling process. It is beyond the scope of this paper to discuss all of these in detail. However, we will examine those issues where data from treatment studies have been a source of evidence and, in particular, those where studying generalisation after treatment can shed more light.

A first question raised by Tainturier and Rapp was whether reading and writing are entirely separate or whether they share processes. Evidence from patients shows that similar impairments can occur in spelling and reading, for example surface dysgraphia can co-occur with surface dyslexia (e.g., Weekes & Coltheart, 1996). Both could result from an impairment to a single, shared orthographic lexicon (Behrmann and Bub, 1992; but see Weekes and Coltheart (1996) for a different view). However, a particular impairment in spelling is not always accompanied by the same impairment in reading. For example, individual RG (Beauvois & Derouesné, 1981) showed an impairment in spelling of irregular words, but was almost perfect at reading these words. While this pattern might argue against a shared lexicon, Tainturier and Rapp (2001) suggest that data from dissociations in reading and spelling performance can still be compatible with a shared-processing account. They conclude that it is likely that at least some components, such as the orthographic lexicon, are shared where others, such as sub-lexical conversion processes, are not. A treatment study that results in generalisation across modalities could further test this claim. For example, an improvement of spelling performance after treatment for reading would support a shared-processing account.

Another question is whether oral and written spelling use two different output buffers, or whether they rely on one single buffer. Rapp and Caramazza (1997) and Pound (1996) described individuals who showed a superior performance for oral spelling compared to written spelling with regard to letter-level errors, and who showed an effect of regularity only in the written modality. These data seem to support independent buffers for the two output modalities. Once again, generalisation across modalities after treatment can provide further evidence to inform this issue. A treatment study resulting in generalisation from the oral modality to the written modality supports either a shared buffer account, or the existence of links between the two buffers (Pound, 1996).

A third issue raised by Tainturier and Rapp (2001) is that while two distinct spelling routes are assumed (lexical and sub-lexical), there is debate about whether these two routes operate fully independently, or whether information from the two routes is shared during the process of spelling. Several studies have provided data suggesting that lexical and sub-lexical processes share information during processing (Folk, Rapp, & Goldrick 2002; Folk & Rapp, 2004; Rapp, Epstein, & Tainturier, 2002). For instance, the non-word spelling of subjects without spelling impairment is influenced by previously having heard rhyming words. For example, the non-word /zi:f/ was more likely to be spelled “zeef” when preceded by the lexical prime “beef”, but as “zeaf” following the prime “leaf” (Campbell, 1983).

A possible mechanism for this influence of lexical information on sub-lexical processing is interaction between these two processes. Even though more often addressed in studies of spoken word production (e.g., Goldrick & Rapp, 2002), interaction between levels of representation is also discussed in written word production (e.g., McCloskey, Macaruso, & Rapp, 2006). It is possible that activation in spelling simply flows forward from one level of representation to the other. For example, in written naming, first a semantic representation is activated, which then leads to activation flowing to the orthographic level, which in turn activates the graphemes. In contrast, an interactive, bi-directional view assumes that information from the grapheme level can feed back to the orthographic level. In this way sub-lexical processes can activate grapheme representations, which feed activation back to the level of the lexicon, thereby influencing lexical selection.

Treatment studies can provide support for this feedback mechanism. An example would be if, after treating sub-lexical spelling, a patient with deep dysgraphia shows a reduced number of semantic errors in spelling. A spelling attempt for the target word "panther" might lead to the response "leopard". After treating phoneme-to-grapheme correspondences, there might be feedback from the sub-lexical route for the initial grapheme "p", which means the entry for "panther" receives a higher level of activation compared to competing entries, which will lead to the correct selection of the target word (cf., Nickels, 1992, for a similar explanation in reading).

In sum, for all of the debates regarding representation and processing for spelling that we have described in this section, carefully designed treatment studies could prove useful in further investigating the questions at hand. In particular, the occurrence of generalisation effects after treatment can be a potential source of evidence. Hence, to better understand the nature of generalisation effects, we now review treatment studies for written language impairment, with a focus on generalisation. The next section will outline the logic of this review and subsequently discuss the results.

Generalisation Effects: a Review of Treatment Studies

In this review we focus on spelling impairments resulting from damage to one or more of the central spelling components. We exclude impairments of language processes that are not specific to written word production but may impact on writing (e.g., auditory discrimination). We have searched for treatment studies by using the following online search engines: Google Scholar, the Macquarie University Library catalogue and the Aphasia Treatment Website from the Academy of Neurologic Communication Disorders and Sciences (http://aphasiatx.arizona.edu/written_writing). The search terms entered were: "acquired dysgraphia, written language impairment, treatment and generalisation". We also considered the reference lists of studies found by the search.

Articles were excluded if they did not provide a detailed assessment of the participants' spelling impairment, as this prevented us from studying the role of the type of impairment in relation to generalisation effects. As noted above, treatment studies with a primary focus outside the spelling domain were excluded. In addition, other studies have combined treatments with different foci (e.g., reading and spelling treatment, Orjada & Beeson, 2005). Because it is difficult to determine the exact contribution of the spelling treatment in these cases, we have included treatments with a focus on spelling only. This led to the selection of 40 treatment studies.

Measuring generalisation. A variety of types of generalisation can be measured, including generalisation to untreated control items, to standardized language tasks, to untreated modalities (e.g., from typing to handwriting) or to the functional use of writing as a means of communication. We will briefly discuss each of these different types of generalisation.

In order to be able to test for generalisation to words other than those directly treated (generalisation across items), it is important that treatment materials include a list of untreated control items. Ideally these control items should be matched to the treated items for difficulty, and they should be probed before treatment (on two or more occasions) and after treatment, but they should not be practiced as part of the treatment (Howard, 1986; Willmes, 1990). However, a number of studies have measured generalisation by looking at improvements to untreated items administered pre and post treatment in the context of a standardised test. This could, for example, be a spelling task from an aphasia battery such as the Western Aphasia Battery (WAB: Kertesz, 1982; e.g., Raymer, Cudworth, & Haley, 2003). The untreated[3] items in these cases are not matched to the treated items for difficulty, and it may therefore be a less reliable measure for evaluating generalisation of treatment effects. Nevertheless, an

[3] It should be noted that it is vital to check that the standardised tests do not include any of the treated items, however, this is often not verified.

improvement on such spelling tasks can still be viewed as generalisation of improvement across items.

Some studies have aimed to investigate whether an individual can acquire written vocabulary to facilitate communication, or whether the communicative use of writing can be encouraged (e.g., Robson, Pring, Marshall, Morrison, & Chiat, 1998). In these cases the main goal is to train a personalized list of items, and, generalisation is not generally investigated, although there is no reason in principle why it could not be.

While a change in accuracy of spelling of untreated items is the most common measure of generalisation, a change in error types can be another source of evidence. For example, participant EMF (Beeson & Egnor, 2006) showed very little change in her accuracy scores on untreated items on standardized assessment tasks administered pre and post treatment. However, before treatment EMF produced mainly non-words as error responses on a written naming task (Psycholinguistic Assessments of Language Processing in Aphasia (PALPA), Kay, Lesser, & Coltheart, 1992, task 53), whereas after treatment her errors consisted of visually similar real words, suggesting improved access to orthographic representations. A similar qualitative change in error types was reported for HR (Murray & Karcher, 2000), who had a graphemic output buffer impairment. Before treatment, HR produced mainly non-words and other unrelated errors in written naming, but following treatment, these errors were more often target-related (e.g., semantically related), regardless of whether the stimulus had been treated. This qualitative improvement across both treated and untreated items suggested that treatment had facilitated the functioning of the graphemic output buffer in general. Clearly, how generalisation is measured depends on the questions asked in a particular study.

We now discuss groups of treatment studies organised by the type of treatment. We give the characteristics of the studies that could play a role in the manifestation of generalisation (level of impairment, type of treatment) and discuss patterns of any type of generalisation found. If a study included several treatment methods, it may appear in more than one table. Some studies provided more detail about the participant's underlying impairment compared to others. For example, some studies indicated their participant suffered a form of lexical impairment, while others localised the impairment more specifically within the lexical route to the orthographic output lexicon. Therefore, the descriptions in the tables vary in the specificity of the impairment (e.g. "Lexical (OOL)" vs. "Lexical"). When the severity of the impairment made an in-depth assessment impossible, the term "global dysgraphia" is used. Similarly, not all studies specified the type of items that were used to test for generalisation (i.e., whether these were an untreated control items or a standardised assessment such as PALPA, or for example whether the words were regular or irregular). Therefore, in the tables the details of these items vary depending on the information provided in the studies.

As will become apparent, the same type of treatment (e.g., compensatory) can have different results, and treatments with different aims (e.g., compensatory vs. lexical) can have the same outcome: For example, a treatment which aimed to provide a compensatory strategy, may for one individual also have a direct effect on the accessibility of treated stimuli (item specific improvement). Hence, in the structure of this review, this compensatory treatment (Category 1) would have an outcome consistent with the aims of a treatment for lexical processes (Category 3). This pattern is not restricted to studies of spelling, in the context of anomia treatments, Nickels and Best (2000; Nickels, 2002) also note that multicomponent treatments may work in different ways for different individuals. In order to simplify the description of the treatment, we have classified by the aims of treatment, rather than the results. In the Discussion section we then draw together the studies with the same results.

Category 1: Using compensatory strategies to assist writing. In this section four types of compensatory strategies are discussed. Two treatment studies focused on assisting impaired writing by the use of a computer or other types of writing aids (Behrns, Hartelius, & Wengelin, 2009; Bruce et al., 2003). The next two studies (Mortley, Enderby, & Petheram, 2001; Pound, 1996) taught their participants to use their intact oral letter naming abilities to improve their written spelling. In the other two studies (Beeson et al., 2000; Hillis & Caramazza, 1987) participants were taught to use their relatively intact residual knowledge of phoneme-to-grapheme correspondences to improve their ability to self-correct spelling errors. Finally, De Partz (1995) taught participant AM to divide long words into shorter segments. Table 3 summarises the characteristics of these studies.

More studies have used computers in their treatment, and this type of treatment has proven successful in a number of studies (see Beeson, Rising, Kim, & Rapcsak, 2010). However, there remains a distinction between using the computer to simply administer the therapy and/or provide an output modality (e.g., Deloche, Dordain, & Kremin, 1993) and interactive studies where the computer is used as a writing aid (e.g., Bruce et al., 2003). The former are integrated into the appropriate sections below, and the latter are discussed in this section.

Level of impairment. The participants described by Behrns et al. (2009) all had severe writing impairments, and it was therefore reasoned a writing aid would assist their spelling by making the revision process easier. In other cases, individuals had relatively good spoken language skills but poor written skills and speech recognition software was successful in improving their ability to express themselves in writing (e.g., Bruce et al., 2003). The participants described by Mortley et al. (2001) and Pound (1996) both had relatively intact oral spelling abilities compared to their impaired written spelling. The studies by Beeson et al. (2000) and Hillis & Caramazza (1987) focused on their participant's relatively intact sub-lexical skills, in the context of a lexical impairment and impairment to the output buffer, respectively. Finally, De Partz (1995) treated a participant who mainly suffered an impairment to the graphemic output buffer.

Type of treatment and items. For these studies designed to facilitate spelling using a strategy, the use of this strategy or writing aid is the focus of treatment rather than accurate spelling of a specific list of words (e.g., Behrns et al., 2009). For example, Behrns et al.'s participants were provided with a book with pictures and were asked to use the computer to write about what they saw in the pictures. Picture description was simply a means to get the participants to use and experiment with the word predictor and spell-checker. During intervention, the participants were asked to write weekly notes in a diary. Writing performance was measured by a number of variables, including the total number of words in the final text, and proportion of correctly written words.

De Partz (1995) used words that contain a lexical segment within the word, e.g., the French word 'cravache' ('crop') contains the lexical segment 'vache' ('cow'). In a delayed copying task these segments were underlined, to draw the participant's attention to parts of the word, in order to train a segmentation strategy of dividing words up into smaller chunks.

Table 3

Category 1: Using compensatory strategies to assist writing.

Participant	Impairment	Method of treatment	Results of treatment	Gen, across items?	Gen, across tasks / modalities?	Authors' interpretation of results
SV (Beeson et al., 2000)	Lexical (OOL) and sub-lexical	Using residual knowledge of PGCs, plus use of electronic speller	+ (improved self-correction and use of electronic speller)	+ (control words (JHU); Error change)	No data	Partial information from lexical and sub-lexical routes strengthens representations: Evidence that lexical and sub-lexical routes interact.
SW (Beeson et al., 2000)	Lexical (OOL) and sub-lexical			+ (control words (PALPA))	No data	Participant's "problem-solving" technique improved to self-correct spelling attempts.
Anders (Behrns et al., 2009)	Sub-lexical and some semantic impairment	Writing with word prediction	+ (e.g., more words written, more words written correctly)	+ (some improvement maintained when aid was not in use)	No data	Computerised writing aid made revision process more efficient and facilitated word retrieval process in general.
Bo (Behrns et al., 2009)	Sub-lexical and lexical					
Carol (Behrns et al., 2009)	Sub-lexical and lexical and some semantic impairment	Writing with spell-checker				
MG (Bruce et al., 2003)	GOB and allographic processes (impaired letter shape selection)	Voice recognition system (VRS) as a writing aid	+ (longer texts)	+ (more diverse vocabulary; writing became means of communication (e.g., started a diary))	No data	Writing by voice can benefit people with writing impairment who have good oral language skills.
AM (De Partz 1995)	GOB and some additional lexical (OOL) and sub-lexical impairment	Lexical segmentation strategy in delayed	+ (most improvement for decomposable words)	+ (control words (decomposable and non-	No data	Participant learned to use a strategy of decomposing words into smaller segments, to compensate for

		copying with lexically decomposable and non-decomposable words		decomposable) but only after second treatment phase); decomposable non-words		dysfunctioning GOB.
DH (Hillis & Caramazza, 1987)	GOB	Using residual knowledge of PGC to train spelling self-corrections	+	- (no decrease in errors but self-correction improved)	No data	Self-monitoring strategy to detect errors improved by using intact OOL.
MF (Mortley et al, 2001)	Lexical, sub-lexical, and GOB: Written spelling more impaired than oral spelling	Improving written spelling via oral spelling: letters written as self-dictated from oral spelling	+ (learnt to use oral spelling strategy independently)	+ (control words and non-words)	+ (written picture naming, letter naming; functional setting (personal letters up to sentence level))	Compensatory strategy built on residual oral spelling abilities.
JA (Pound, 1996) *	Lexical (OOL) and sub-lexical and GOB: Written spelling more impaired than oral spelling	Improving written spelling via oral spelling	+ (learnt to use strategy successfully)	+ (control words (reg. and irreg.), non-words)	+ (written picture description)	Treatment enabled (sub-lexical) oral writing skills to become available for written spelling.

Note. “Results of treatment” refers to whether there was successful use of the strategy that was taught in treatment. “Generalisation across items” refers to whether the strategy was also used outside of treatment (e.g., in spontaneous writing), or whether spelling was also improved after treatment when the spelling aid was not in use, or an improvement of untreated items.

Gen = Generalisation; GOB = Graphemic Output Buffer; OOL = Orthographic Output Lexicon; + = improvement or generalisation; - = no improvement or generalisation; Control = to Control items (regular / irregular words, or non-words); Error change = change of error types; JHU = Johns Hopkins University Dysgraphia battery (Goodman & Caramazza, 1986c); PALPA = Psycholinguistics Assessment of Language Processing in Aphasia (Kay, Lesser, & Coltheart, 1992); * = study occurs in more than one table.

Outcome. The idea behind teaching a strategy (e.g., by using a compensatory writing aid) is that such a strategy could then be applied to any word that is to be spelled, which means the treatment should automatically lead to generalisation. The studies in this category proved to be successful, both in terms of improvements when a strategy or writing aid was introduced and with evidence of functional benefits in other situations, and in some cases this improvement was maintained when spelling without the aid after treatment (Behrns et al., 2009). Behrns et al. pointed out that when the revision process is made easier by the spell-checker, the working memory load is reduced and therefore more resources may now be available for the process of writing.

Using intact letter naming abilities to retrieve the written spelling can provide a useful strategy for some individuals who show a difference in oral compared to written spelling abilities. Such a strategy indicates that treatment can improve the transmission of information between lexical and sub-lexical components of the spelling process. For example, participant JA (Pound, 1996) would correctly read aloud an exception word (e.g., "wolf") she could not write correctly ("wold"). After treatment she learned to spell aloud the word one grapheme at a time from the intact representation in the lexicon, and then via the sub-lexical route convert these sounds to letters for written spelling. This means treatment can make certain routes or connections in the spelling process available that were not used before and might have been "unnatural" for normal writing processes (Pound, 1996; however, this strategy might not be effective for all individuals: see Lesser, 1990). De Partz (1995) argued individual AM learned to use a strategy of segmenting words into smaller parts that he could then write. This strategy was successful to compensate for graphemic output buffer deficit.

Category 2: Treating sub-lexical processes. The sub-lexical spelling system is based on phoneme-to-grapheme conversion, and provides an alternative to (or supplement for) the lexical spelling process (Beeson & Rapcsak, 2002). In sub-lexical treatments the goal is for an individual to relearn phoneme-to-grapheme correspondences, in order to improve spelling. The next 7 studies (for 11 participants) adopted such an approach, sometimes combined with a form of lexical treatment (e.g., Cardell & Chenery, 1999; De Partz et al., 1992. The results are summarised in Table 4.

Level of impairment. Predictably, the majority of the treatments described here (6 of the 7 studies) included individuals who (amongst other impairments) suffered from impairments to the sub-lexical spelling route. However, sub-lexical treatments have also been administered to individuals whose main impairment was of a different type. De Partz et al. (1992) reported a case with a more impaired lexical route, compared to a relatively spared sub-lexical route. The authors argued that sub-lexical treatments can aim to support an impaired lexical route, at least for regular words (e.g., hand) and regular parts of irregular words (e.g., the y and t in yacht). It should be noted that the more regular an orthography is the more successful a sub-lexical treatment is expected to be in generalising to word spelling. Treating sub-lexical processes can furthermore be used to support spelling in the case of graphemic output buffer impairment (Hillis & Caramazza, 1987).

Table 4

Category 2: Sub-lexical processes: re-teaching specific phoneme-to-grapheme correspondences

Participant	Impairment	Method of treatment	Results treated items	Gen, across items?	Gen, across tasks/ modalities?	Authors' interpretation of results
P1 (Beeson et al., 2010) P2 (Beeson et al., 2010)	Lexical and sub-lexical	Phonological cueing to retrain 20 consonant and vowel PGCs with keywords plus using problem-solving approach, with use of electronic speller	+	+ (control words (reg. + irreg.) and non-words)	+ (phonological tasks (e.g., rhyme judgement), lexical decision, and case conversion, and to reading)	Strengthened phonological skills supported spelling. Reflection of combined output of lexical and sub-lexical routes.
CV (Cardell & Chenery, 1999) *	Lexical and sub-lexical, link between semantics and OOL, GOB	Focus on relationship between letters and sounds, using non-word anagrams, letter-sound mismatch detection	+ (non-words)	+ (control non-words)	+ (written naming)	Both PGCs and functioning of GOB improved. Evidence for feedback between levels of representation.
LP (De Partz, et al., 1992) *	Lexical	Teaching context-sensitive graphemic rules (e.g., phoneme /o/ is grapheme EAU at end of word: "bateau")	+	+ (control words (reg.); error change: over-generalisation rule)	No data	Treatment optimised the relatively spared sub-lexical route in writing.
JS (Hillis Trupe, 1986)	Semantic and sub-lexical	Relearning 30 PGCs with keyword cues	+	+ (error change; control words)	No data	JS was able to use partial information from both sub-lexical route and lexical system to derive correct spelling.
SJD (Hillis &	Lexical (OOL)	Relearning 30	+	+ (verb	No data	Treated PGCs served as self-cueing

Caramazza, 1994)	and sub-lexical	PGCs with keyword cues.		accuracy and number of sentences written)		strategy for retrieval of correct spelling of words.
P1 (Kiran, 2003)	Lexical (OOL) and sub-lexical	Improving set of regular words by training the PGCs	+	+ (control words; error change)	+ (written naming of treated and untreated regular words)	Treatment strengthened the link between GPCs, GOB and OOL. Strengthened representations accessible in different modalities.
P2 (Kiran, 2003)			- (no improvement on treated items, only some error responses more closely matched to target)	+ (error change; modest improvements control words)	-	
P3 (Kiran, 2003)	Semantic and sub-lexical		+	+ (control words; error change)	+ (written naming and oral spelling of treated and untreated regular words)	
RO (Luzzatti et al, 2000)	Global dysgraphia (only able to write single letters)	Decomposition of words and non-words into syllables and phonemes to relearn PGCs	+ (on CV and bi-syllabic CVCV words)	+ (control words (reg. + irreg.), non-words)	+ (reading; naming task)	Near-normal performance on writing shows it is possible to restore sub-lexical route in an orthographically transparent language.
DR (Luzzatti et al, 2000)	Lexical and sub-lexical, including GOB			+ (control words (reg. + irreg.), non-words)	No data	

Note. Control = to Control items (regular / irregular words, or non-words); Error change = change of error types; Gen = Generalisation; GOB = Graphemic Output Buffer; GPC = Grapheme-to-Phoneme Correspondence; OOL = Orthographic Output Lexicon; PGC = Phoneme-to-Grapheme Correspondence; + = improvement or generalisation; - = no improvement; *study occurs in more than one table.

Treatment method and items. Re-teaching phoneme-to-grapheme correspondences (PGCs) can be achieved in various ways. Luzzatti et al. (2000) devised a treatment programme that started by teaching decomposition of spoken words and non-words into syllables and syllables into phonemes. Hillis Trupe (1986) linked a target grapheme with a key word (e.g., "/b/ is for baby"), in order to access the grapheme via the use of this key word.

Outcome. Relearning individual PGCs should enable the participant to spell regular words containing this PGC. For the papers reviewed here, all except one participant (10 of the 11 participants) were able to use their sub-lexical knowledge to improve writing. Furthermore, four different types of generalisation were found: (1) generalisation to the spelling of untreated control words (e.g., on a writing to dictation task: De Partz et al., 1992); (2) a change of error types (a lower proportion of 'no responses'; Kiran, 2003 (P2, P3); (3) generalisation across different input modalities (from writing to dictation to written naming; Kiran, 2003); and (4) across other modalities (improved performance in reading words and non-words; Luzzatti et al., 2000).

Change at the cognitive level. It seems that a sub-lexical training programme can improve specific phoneme to grapheme correspondences in the sub-lexical route. However, as described above, sub-lexical training can also result in lexical improvements. Different explanations for lexical improvement have been put forward. One possibility is that better word spelling reflects an improved strategic problem-solving approach to spelling, for example shown by an increase in self-corrected spelling errors (Beeson et al., 2010).

Secondly, improvement of word spelling can be explained by a mechanism of bidirectional feedback between different levels of representation. Focussing on the phoneme-to-grapheme correspondences strengthens the link between the PGCs and the graphemic output buffer. Untrained words will then improve as a result of reinforced links between PGCs, the buffer and up to the orthographic output lexicon (OOL).

Beeson et al. (2010) tried to achieve this kind of generalisation by combining a lexical treatment to retrain item-specific spellings, a phonological treatment to strengthen sub-lexical skills, and an interactive phase to maximize the use of residual lexical and phonological knowledge. After treatment a change in error types was reported, with errors now showing evidence of a combination of lexical and sub-lexical information. For example, the error "anteeque" reflects a sub-lexical locus for the generation of "ee" while "que" has to be generated lexically. Beeson et al. argued that these error types show it is possible to facilitate feedback between lexical and sub-lexical processing with an interactive intervention focussing on both lexical and sub-lexical processes.

Furthermore, improvement across input modalities (e.g., writing to dictation to written naming) may occur because both modalities access the same single representation in the orthographic output lexicon, and therefore training one modality of representation can have beneficial effects on retrieving the same representation in another modality (from writing to dictation to written naming, Kiran, 2003). In a regular orthography, training the rules for converting phonemes to graphemes in writing can generalise to the grapheme-to-phoneme rules used in reading, shown by generalisation from writing to reading (Luzatti et al., 2000).

Finally, sub-lexical treatment can be beneficial for individuals with a lexical impairment who have relatively spared sub-lexical skills. For example participant SJD (Hillis & Caramazza, 1994) presented with mainly a lexical spelling impairment, and showed preserved oral naming abilities. Treatment focused on a set of PGCs enabling SJD to support her written naming with her knowledge of the spoken names of the words she wished to write. Following treatment SJD was able to use PGCs to derive the first letter of the written word from the spoken word. This first letter acted as a cue for retrieval of the written word form.

Category 3: Treating lexical processes. When spelling via the lexical route is impaired, treatment may focus on strengthening (access to) orthographic representations or underlying semantic knowledge (Beeson & Rapcsak, 2002).

Category 3.1: Lexical strategies. The use of item specific mnemonics or a cognitive relay strategy has proven successful in three studies (Hatfield, 1983; De Partz et al., 1992; Schmalzl & Nickels, 2006), with six participants, as summarised in Table 5.

Level of impairment. The studies described were conducted with individuals with impaired lexical retrieval (e.g., De Partz et al., 1992). Four individuals also had co-occurring impairments: a sub-lexical impairment (deep dyslexia; Hatfield, 1983), or an impairment to both the graphemic output buffer and a semantic impairment (Schmalzl & Nickels, 2006).

Treatment method and items. The rationale for Hatfield's (1983) treatment was to improve writing by using a relay strategy. Prior to treatment, it was noted that the participants were able to write concrete words, even when they were unable to write a corresponding homophonic function word (e.g., "our" (incorrect) compared to "hour" (correct)). The treatment method therefore consisted of linking function words to homophonic content words. The preposition "in", for example, was linked to the noun "inn" (pub), as a strategy to write the function word.

De Partz et al. (1992) and Schmalzl and Nickels (2006) used a visual imagery strategy to facilitate access to orthographic representations (combined with another form of, respectively, sub-lexical treatment and copy and recall). The idea was that the use of mnemonics could aid the retrieval of semantic representations and as a consequence facilitate the access of these entries in the orthographic output lexicon. FME (Schmalzl & Nickels, 2006) was asked to generate an image that would remind her of the meaning of the target words. An example of such a cue is the word "look" written down, with a picture of glasses drawn around the letters "oo" in the picture. This semantically related and letter-framed cue aimed to facilitate FME's access of the target form when evoking the semantically related image.

Outcome. In all of these cases the use of a strategy was taught successfully. All six patients improved spelling for trained words. Furthermore, Hatfield (1983) and De Partz et al. (1992) argued for some form of generalisation of the treatment. Hatfield (1983) reported successful use of the strategy to facilitate writing of function words. However, only limited reassessment data is available to fully support this claim. De Partz et al. (1992) reported some transfer to spontaneous writing, after they had trained their individual LP to detect the treated items in spontaneous written production and apply the strategy of drawings. Additionally, LP showed generalisation to derivational forms of treated items (writing "inflammation" after training the word "flamme" (flame)). Furthermore, untreated words improved for LP, however the authors argued an improved score after treatment could partly be the reflection of variability in LP's performance, but it could also be the case that LP learnt to spontaneously apply the strategy of visualizing an image when writing a word. Schmalzl and Nickels (2006) demonstrated improved retrieval of orthographic representations after the lexical copy and recall treatment with mnemonics, however generalisation was not observed across items.

Change at the cognitive level. Visual imagery strategies or relay strategies result in item specific improvement, as they target specific representations in the lexicon. Thus, representations for treated words may have been strengthened as a result of the training or access to these words has become easier due to the practice. It is less clear that these strategies reliably result in generalisation. De Partz et al. (1992) argued if an individual has learned to generate images in relation to words, one cannot prevent this strategy being used for other sets of words. However, because the authors show their participant's performance was somewhat variable across testing sessions, we cannot provide strong evidence for generalisation with this strategy. Weekes & Coltheart (1996) also pointed out that an imagery strategy like the one used by De Partz et al. (1992) focuses truly on visual properties of the words, rather than orthographic properties such as in a lexical treatment, and is therefore possibly less likely to show generalisation.

Table 5

Category 3.1: Improving lexical retrieval with lexical strategies

Participant	Impairment	Method of treatment	Results treated items	Gen, across items?	Gen, across tasks / modalities	Authors' interpretation of results
BB, DE, PW (Hatfield, 1983)	Sub-lexical and semantic	Linking function words with homophonic content words	+	?	?	Generalised use of strategy reported (however limited reassessment data available).
TP (Hatfield, 1983)	Lexical					
LP (de Partz et al., 1992) *	Lexical	Improving irregular words by using visual imagery strategy	+	?	?	Trained strategy used in spontaneous writing, however possibly due to inconsistency in performance.
FME (Schmalzl & Nickels, 2006) *	Lexical (OOL), semantic system and GOB	Copy and recall with use of mnemonic cues	+ (only with mnemonic cues)	-	-	Mnemonic cues treat item specific representations only.

Note. Gen= Generalisation; GOB = Graphemic Output Buffer; OOL = Orthographic Output Lexicon; + = improvement or generalisation; - = no improvement or generalisation; ? = limited data available or pattern unclear; * = study occurs in more than one table.

Category 3.2: Treatment to improve semantics. The semantic system is assumed to be a central component that is linked to different modalities of language processing (e.g., spoken input, spoken output, written input, written output; see Figure 1, Whitworth et al., 2005). Hence, impairments of this central component can lead to impaired functioning in written and spoken language. By this same logic, successful treatment of the central semantic component can benefit across different components of language processing. Treatment aimed at improving semantic processing has been applied in four studies (six participants), see Table 6.

Type of impairment. Almost all individuals (five out of six) described in this section suffered from an impairment to the semantic system. This deficit leads to the occurrence of semantic errors across lexical tasks (e.g., written naming and oral naming). It is assumed that semantic representations consist of a number of functional and perceptual features that determine the meaning of a word (Hillis, 1992). For example, the representation of "tulip" consists of information such as <flower>, <bulb>, <spring blossoms>, and <upright petals>. Semantic errors are hypothesised to arise if specific information (e.g., <upright petals>) is damaged or reduced, which can result in erroneous selection of a representation with similar information, such as "daffodil" (Hillis, 1992). One participant (Patient 2, Hillis, 1989) made semantic errors in oral naming but not in written word production, and therefore Hillis argued the semantic errors arose from a difficulty accessing the phonological word forms.

Treatment method and items. Semantic treatment usually focuses on relearning the specific characteristics of items in order to reduce semantic errors. This is often achieved by using tasks that are set up to distinguish between semantically similar items (Hillis, 1991). This type of treatment is proposed to address the semantic system directly. This can be achieved via written naming, either with a cueing hierarchy (participant KE, Hillis & Caramazza, 1994) or by explicitly teaching semantic distinctions (participant HG, Hillis, 1991), or more implicitly via written and/or spoken word to picture matching (Nickels & Best, 1996; participant JJ, Hillis & Caramazza, 1994).

Outcome. All treatments reported here resulted in improved writing of the treated items. Furthermore, in some studies generalisation was found for untrained words within the same semantic category as the trained words for two individuals (HG, Hillis, 1991; KE, Hillis & Caramazza, 1994). Nickels and Best (1996) reported generalisation to spoken naming.

Change at the cognitive level. If a single deficit (i.e., to the semantic system) is the source of semantic errors in, for example, both oral and written naming (e.g., KE, Hillis & Caramazza, 1994), an improvement in semantic processing is likely to result in improved performance across modalities. Generalisation may also be expected to semantically related words that share certain specific features that were the target of treatment. For example, treating the specific information for flowers (<flower>, <bulb>, <spring blossoms>, and <upright petals>) would be predicted to strengthen the representations of items sharing one or more of these characteristics (both tulip and daffodil).

Generalisation has been found both when the distinctions between semantically related items were explicitly taught (e.g., HG, Hillis, 1991) and when they were being taught more implicitly (e.g., in word-picture matching (Hillis & Caramazza, 1994). For example, when asked to point to a *tulip* with a daffodil as a distractor, JJ (Hillis & Caramazza 1994) was able to determine some aspects of

Table 6

Category 3.2: Treatment to improve access to orthography from semantics

Participant	Impairment	Method of treatment	Results treated items	Gen, across items?	Gen, across modalities?	Authors' interpretation of results
HG (Hillis, 1991)	Semantic	Teaching distinctions between semantically related items (e.g., tiger and lion) in written naming	+	+ (written naming of untreated items semantically related to trained items)	+ (oral naming, repetition, writing to dictation, word-picture matching of trained items)	(Re)establishment of distinctions between related items within semantic system.
Patient 1 (Hillis, 1989)	Semantic	Written naming of nouns with cueing hierarchy	+	+ (untreated nouns in same semantic categories, no gen to untrained verbs)	+ (oral naming of both trained and untreated items)	Treatment improved Patient 1's semantic impairment and Patient 2's access to phonological representations. Contrasting results in generalisation attributed to differences in underlying deficit.
Patient 2 (Hillis, 1989)	Impaired access to POL; GOB	Written and oral naming of nouns with cueing hierarchy	+	-	-	
JJ (Hillis & Caramazza, 1994)	Semantic	Written word to picture matching	+	No data	+ (oral naming of untreated, semantically related items; written naming of treated items)	Semantic representations were strengthened even though treatment did not target specific semantic distinctions between items.
KE (Hillis & Caramazza, 1994).	Semantic	Written naming with cueing hierarchy	+ (reduction of semantic errors)	+ (untreated nouns in same semantic category)	+ (oral naming of treated items)	Cueing and reinforcement of correct response strengthened semantic representations. Single deficit responsible for impaired oral and written naming.

TRC (Nickels & Best, 1996)	Semantic (and link to POL)	Written and spoken word to picture matching	+	+ (untreated items after written word-picture matching)	+ (treated and untreated items both in spoken and written naming after written word-picture matching)	Therapy was effective at level of semantic system: one modality of input (written word-picture matching) produced improvements in both modalities of output (written + spoken naming).

Note. Gen = Generalisation; GOB = Graphemic output buffer; POL = Phonological output lexicon; + = improvement or generalisation; - = no improvement or generalisation the distinction between tulips and daffodils (Hillis, 1992) resulting in improvement for both the treated item (tulip) and the untreated item (daffodil).

Nickels and Best (1996) argued that generalisation from written to spoken naming for TRC also shows that treatment using word-picture matching is effective at the level of the semantic system. However, more recently, Nickels (2002) suggested that word-picture matching is more likely to be effective by strengthening the links between the semantic system and phonology. Consequently, TRC's generalisation to spoken naming may instead be due to the activation of the phonology of treatment words during the treatment task (see also Patient 2, Hillis, 1989).

Category 3.3: Treatment methods aiming at directly improving lexical representations or access. In previous sections we discussed studies that had the ultimate aim of improving word spelling. They did so either using compensatory strategies (e.g., spell-checker), or by targeting components of the spelling system that are not directly involved in word spelling but may circumvent the problem of poor lexical knowledge (e.g., treating PGCs or directly treating semantics). This last section reports studies that directly treated spellings of words. The majority of studies reviewed in this paper (24 studies, 48 individuals) report treatments that aim to strengthen orthographic representations and/or improve access to these representations (sometimes combined with a sub-lexical treatment component). In most of the studies, more than one participant took part in the treatment, and often results are not consistent across participants. We therefore group the results according to whether and what type of generalisation occurred and we report this by participant (rather than by study).

Table 7 features the participants who showed item-specific effects only. Table 8 lists individuals for whom treatment resulted in generalisation of spelling across items, either untreated control items, or other tasks (e.g., PALPA). In Table 9 we detail participants for whom treatment generalised across modalities (e.g., improved verbal repetition; Beeson, Rising, & Volk, 2003). Table 10 reports participants for whom the treatment resulted in a change of error types.

In the tables in this section we use a slightly different format, in order to specify the type of generalisation in more detail. If treatment resulted in more than one type of generalisation (e.g., across items and across modalities), the participant's results will appear in more than one table. These classifications were used to investigate if there are common participant, treatment or other characteristics that determine the occurrence of generalisation. We will discuss each of these in turn including potential mechanisms for generalisation.

Table 7

Category 3.3: Lexical treatments showing item-specific improvements only, no generalisation across items, spelling tasks or other language modalities.

Participant	Impairment	Method of treatment	Results treated items	Generalisation?	Authors' interpretation of results
JES (Aliminosa et al. 1993)	Lexical and sub-lexical impairment, and GOB	Delayed copying + writing to dictation	+	- (n.s across items, no data across tasks)	Strengthening of specific graphemic representations in OOL, might indicate incorrect assumption of GOB deficit.
ST (Beeson, 1999)	Lexical impairment and GOB	ACT + CART	+	- (but newly acquired vocabulary used to assist spoken communication)	Treatment improved access to specific orthographic representations, and also shows functional impact of writing treatment.
FD, AD (Beeson, Hirsch, et al., 2002)	Global dysgraphia	ACT + CART	+	- (n.s. across tasks. Increased use of writing as means of communication)	Treatment was item-specific, however participants started using writing as communication modality.
LG, ED (Beeson, Hirsch, et al., 2002)	Lexical and sub-lexical impairments	CART			
MR, MB, GP, JF (Beeson et al., 2003)	Global dysgraphia	CART	? Some word specific improvements but did not reach / maintain criterion	- (n.s. across items or tasks)	CART strengthens item specific representations and access to these representations.
WK (Beeson et al., 2003)			+	- (n.s. across items or tasks)	
NEM (Beeson & Egnor, 2006)	Lexical impairment, reliance on (impaired) phonological skills	CART written and spoken naming combined with spoken repetition	+ (largest improvement for CART spoken naming)	- (n.s. on PALPA tasks (writing + oral naming + repetition))	CART + repetition proves beneficial if phonological skills are intact to enhance links between orthography and phonology.
DR (Clausen & Beeson, 2003; all reported in	Lexical and sub-lexical impairment	CART with group treatment	+	? Use of learned (written) items to	Specific orthographic representations were

Beeson et al. 2002 / 2003)				support (spoken) conversation in group treatment, though learned items needed considerable practice	targeted successfully. Some generalisation to group setting shows that conversational training as a more natural context can complement individual spelling treatment.
SL (Clausen & Beeson, 2003; all reported in Beeson et al. 2002 / 2003)			+		
WD (Clausen & Beeson, 2003; all reported in Beeson et al. 2002 / 2003)			+		
AD (Clausen & Beeson, 2003; all reported in Beeson et al. 2002 / 2003)			+		
MMD (Rapp & Kane, 2002; and Rapp 2005)	Lexical (OOL) and sub-lexical impairment	Spell-study-spell / delayed copying	? (however not long lasting)	- (n.s. on control words)	Treating lexical deficit seems more likely to only show item specific effects.
CB (Rapp & Glucroft, 2009)	Lexical (OOL), GOB, possibly sub-lexical (primary progressive aphasia)	Spell-study-spell	? (modest improvement of treated items, however fragile)	- (n.s. on control words)	No generalisation, however treated words seemed protected from deterioration 6 months post onset.
P4 (Raymer et al. 2009)	Lexical and sub-lexical impairment, semantic impairment, including GOB	Errorless vs. errorful learning (adaptation copy and recall)	+ (both errorful and errorless condition)	- (n.s. on control words or WAB)	Other factors (reading, executive functions) might contribute to outcome of treatment.
CM, DY (Robson et al, 2001)	Global dysgraphia	Variety of tasks (e.g., delayed copying, word-picture matching, anagram sorting) followed by communicative therapy phase	? More items correct but no significant gains	- (n.s. on written naming of control words; no generalisation to functional use (conveying messages))	Naming might be item specific only, or writing as a means of communication is unnatural.
SW, AM, LT, DH (Robson et al, 2001)	Global dysgraphia		+		
RMM (Robson et al, 1998)	Global dysgraphia	Variety of tasks (e.g., delayed copying, word-picture	+	? Some generalisation to use of treated items in second treatment	Improved access to specific orthographic representations, but no

		matching, anagram sorting) followed by communicative therapy phase + 'message therapy' to encourage functional communication		phase but not to functional use	generalisation to writing as alternative form of communication.
FME (Schmalzl & Nickels, 2006) *	Lexical (OOL), semantic impairment, and GOB	Copy and recall	- (improvement only with additional use of mnemonics)	-	In presence of an additional semantic impairment, targeting orthographic representations alone might not be sufficient.
NW (Weekes & Coltheart, 1996)	Lexical impairment	Spelling to dictation of homophones with mnemonic cue	+	- (n.s. across items or tasks)	Lack of generalisation is in favour of two orthographic lexicons (reading/writing) account.

Note. ACT = Anagram and Copy Treatment; CART = Copy and Recall Treatment; GOB = Graphemic Output Buffer; OOL = Orthographic Output Lexicon; PALPA = Psycholinguistics Assessment of Language Processing Aphasia (Kay, Lesser, & Coltheart, 1992); WAB = Western Aphasia Battery (Kertesz, 1982); + = improvement or generalisation; - = no improvement or generalisation; ? = limited data available or pattern unclear; n.s = not significant; * study occurs in more than one table.

Table 8

Category 3.3: Lexical treatments showing generalisation across items (either untreated control items or other standardised spelling lists)

Participant	Impairment	Method of treatment	Results treated items	Generalisation?	Authors' interpretation of results
CCM (Behrmann, 1987) *	Lexical impairment	Writing to dictation and a sentence completion task with homophones	+ (treated homophones)	+ (control words (irreg.))	Both improved lexical procedure and improved "checking mechanism".
CV (Cardell & Chenery, 1999) *	Lexical and sub-lexical impairments	Writing to dictation of low imageability nouns + semantic discrimination task	+	+ (synonyms)	General process was targeted, not just individual items. Evidence for interaction or feedback between levels of spelling process.
RB (Deloche et al., 1993) *	Lexical, some semantic	Written naming with cues (e.g., semantic, orthographic)	+	+ (written naming of untreated items)	Treatment improved naming process in itself.
GC (Deloche et al., 1993) *	Lexical, some semantic	Written naming with cues (e.g., semantic, orthographic)	+		
JF (Harris et al. 2012) *	GOB	ACT	+	+ (untreated neighbours with shared middle letters); - (negative impact on untreated neighbours with changed middle letters)	Top-down support from learned lexical representations, which facilitates neighbours with similar letters but interferes with neighbours with changed middle letters.
Ray (Panton & Marshall, 2008)*	GOB	Spelling to dictation and copying	+	+ (control words)	Improved functioning of buffer and strengthened links between buffer and OOL (allowing strengthened lexical entries to support fading buffer level activation).
JRE (Rapp, 2005)	GOB	Spell-study-spell	+	+ (although no maintenance)	Underlying impairment determines response to treatment: buffer impairments likely to generalise as this component is used for spelling of all words.
RSB (Rapp, 2005; and Rapp & Kane, 2002)	GOB	Spell-study-spell	+	+ (control words)	

NM (Raymer et al., 2003)	Lexical (OOL) and GOB	Copy and recall	+	+ (some untrained unrelated words, but mainly to items with similar beginnings (treated "*rac*ket", untreated improvement to "*rac*coon"); to writing to dictation (WAB)).	Improvement to unrelated words mediated by GOB, improvement of "part words" (partly related in spelling to trained words) mediated by OOL.
P1 (Raymer et al., 2009)	Both lexical and sub-lexical, including GOB	Errorless vs. errorful learning (adaptation of copy and recall)	+ (both errorful and errorless)	+ (writing task WAB).	Different improvements for patients with different dysgraphia patterns shows that other factors might play role (e.g., reading, executive functions).
P2 (Raymer et al., 2009)			+ (errorful condition)		
P3 (Raymer et al., 2009)			+ (errorless condition)		
BH (Sage & Ellis, 2006)	GOB	Errorless learning tasks, e.g., pairwise comparison and word search grids.	+	+ (neighbours of treated items)	Top down support from lexical units to graphemic output buffer. Evidence for interaction or feedback between levels of spelling process.
NR (Spencer et al. 2000)	Lexical (OOL) and GOB	Rhyme therapy: oral and written naming of rhymed word pairs of different semantic categories (e.g. *moon* (cue) – *spoon* (target)	+ (both written and oral naming)	+ (both written and oral naming within and across semantic categories	Improved access to phonological and orthographic lexicon. Interactive model of word retrieval: suggests feedback from phonological lexicon to lemma.

Note. ACT = Anagram and Copy Treatment; GOB = Graphemic Output Buffer; OOL = Orthographic Output Lexicon; WAB = Western Aphasia Battery (Kertesz, 1982); + = improvement or generalisation; - = no improvement or generalisation; * = study occurs in more than one table.

Table 9

Category 3.3: Lexical treatments showing generalisation across modalities

Participant	Impairment	Method of treatment	Results treated items	Generalisation?	Authors' interpretation of results
SL (Beeson et al., 2003)	Global dysgraphia	CART	+	+ (verbal repetition task (PALPA))	Generalised improvement in repetition due to increased exposure to written and spoken form of treated items: CART stimulates spoken and written language.
WD (Beeson et al., 2003) *				+ (verbal repetition task (PALPA); use of target words in spoken group sessions)	
DR (Beeson et al., 2003)				+ (written word-picture matching (PALPA) and cross case copying; use of target words in spontaneous speech reported)	
EMF (Beeson & Egnor, 2006) *	Lexical and sub-lexical	CART written + spoken naming combined with spoken repetition	+ (strongest for written treatment)	+ (oral reading of treated items)	Improved access to phonological representation of treated items.
CV (Cardell & Chenery, 1999) *	Lexical and sub-lexical	Writing to dictation of low imageability nouns + semantic discrimination task	+	+ (related tasks such as synonym judgement)	General spelling process was targeted, not just individual items. Evidence for bi-directional language processing: feedback between levels of processing.
RB (Deloche et al., 1993) *	Lexical, some semantic	Written naming with cues (e.g., semantic, orthographic)	+	+ (treated and untreated items in oral naming)	Results reflect underlying impairment: Treatment improved naming process in itself. This improved oral naming, because RB could use PGCs to transcode spoken words, written naming of untreated words was also found.
GC (Deloche et al., 1993) *	Lexical, some semantic	Written naming with cues (e.g., semantic, orthographic)	+	+ (treated items in oral naming)	Results reflect underlying impairment: Because GC's oral naming relied on written naming, oral naming improved. The improvement on untreated written

					naming was argued to be insufficient to allow generalisation.
HR (Murray & Karcher, 2000) *	Lexical (OOL) and GOB	Word picture naming with cueing hierarchy to support word retrieval (word + sentence level)	+	+ (spoken naming of treated items; some written and spoken discourse skills: e.g., more variety in word classes used, longer utterances)	Improved access to spoken forms due to spoken form being available in treatment. Generalisation to discourse is the result of higher level improved processing of grammatical forms (before "split" to different modalities).

Note. CART = Copy and Recall Treatment; GOB = Graphemic Output Buffer; OOL = Orthographic Output Lexicon; PALPA = Psycholinguistic Assessment of Language Processing in Aphasia (Kay, Lesser, & Coltheart, 1992); PGC = Phoneme-to-Grapheme Correspondence; WAB = Western Aphasia Battery (Kertesz, 1982); + = improvement or generalisation; * = study occurs in more than one table.

Table 10

Category 3.3: Lexical treatments showing a change of error types

Participant	Spelling Impairment	Method of treatment	Results treated items	Generalisation?	Authors' interpretation of results
CCM (Behrmann, 1987) *	Lexical	Writing to dictation and a sentence completion task with homophones	+	Errors changed from non-words to opposite member of treated homophone pair	Reinstated lexical procedure. Change of error type result of increased awareness and checking mechanism.
EMF (Beeson & Egnor, 2006) *	Lexical and sub-lexical	CART written + spoken naming combined with spoken repetition	+ (strongest for written treatment)	Errors in written naming changed from non-words to visually similar words	Improved access to orthographic representations.
JF (Harris, et al., 2012) *	GOB	ACT	+	Errors changed from deletion to transposition errors in "shared middle neighbour set" and to substitutions in "changed middle neighbour set"	Interfering effect from neighbours: letters from treated items incorrectly appeared in untreated items.
HR (Murray & Karcher, 2000) *	OOL and GOB	Word picture naming with cueing hierarchy to support word retrieval (word and sentence level)	+	Errors in naming changed from non-words to semantically related errors for treated and untreated items	Improvements in both OOL (item specific improvement) and GOB (generalisation to untrained words).
Ray (Panton & Marshall, 2008) *	GOB	Writing to dictation and copying	+ (writing to dictation)	Change in mean letter position of first error (from error in letter position 2.8 to position 3.5 post treatment)	Improved functioning of GOB.: larger window in buffer available for retaining letter information. Strengthened links between OOL and GOB (allowing strengthened lexical entries to support fading buffer level activation).
JA (Pound, 1996) *	OOL and sub-lexical and GOB: Written spelling more impaired	Improving written spelling via residual oral spelling	+	More phonological plausible spellings in errors	Treatment made route used for oral spelling available for written spelling.

	than oral spelling.

Note. ACT = Anagram and Copy Treatment; CART = Copy and Recall Treatment; GOB = Graphemic Output Buffer; OOL = Orthographic Output Lexicon; + = improvement; * = study occurs in more than one table.

Type of impairment. In this section, we address the question of whether generalisation is only shown by patients with a particular kind of impairment. Tables 8 and 9 illustrate that the individuals who show each form of generalisation have a variety of lexical route impairments in writing: an impaired orthographic lexicon (or access to that lexicon, e.g., Behrmann, 1987), impairment to the graphemic output buffer (e.g., Rapp and Kane, 2002; Rapp, 2005; Sage and Ellis, 2006) or a deficit affecting (at least) both the orthographic lexicon and the buffer (e.g., Cardell and Chenery, 1999; Raymer et al., 2003; Spencer et al., 2000). In many cases the impairment in written word production was accompanied by additional difficulties in reading and spoken language.

It is clear from the range of impairments in Tables 8 and 9 that we cannot conclude that one type of impairment is particularly associated with generalisation compared to another impairment. For example, generalisation has been reported for some participants with an impairment to the buffer (e.g., JF, Harris et al., 2012; JRE, Rapp, 2005) but not for others (e.g., DH, Hillis & Caramazza, 1987).

Previous papers have argued the level of impairment should play a crucial factor in predicting generalisation. Consequently, some studies aim to investigate the role of the underlying impairment in determining the efficacy of treatment. Rapp and Kane (2002) and Rapp (2005) applied the same treatment programme to participants with different levels of impairment in spelling. They found generalisation effects only in individuals with buffer impairments and not in those with orthographic output lexicon deficits. These studies seem to indicate that the buffer plays a role in generalisation effects. Indeed, the lack of generalisation effects has led some authors (Aliminosa et al., 1993; Beeson 1999) to argue that their patient was unlikely to have a buffer impairment.

Different explanations have been put forward to explain the role of the buffer in generalisation. Firstly, generalised improvement can arise from an improvement at the level of the graphemic output buffer. One possibility is that treatment can improve the capacity of the buffer. For example Panton and Marshall (2008) reported a change in the position of the error in the word for their participant Ray. Before treatment, the first error occurred around the third letter of the word (letter position 2.8), compared to letter position 3.5 after treatment. The authors argue that this shift in position shows that after treatment a larger window in the buffer must have become available for retaining letter information. Secondly, other processes that are specific to the buffer might have improved, such as scanning speed or speed of transfer to letter-shape conversion processes (Rapp & Kane, 2002). Thirdly, strengthened orthographic representations can be more resistant to error in the context of graphemic output buffer damage (Rapp & Kane, 2002).

Furthermore, generalisation in an individual with buffer impairment does not necessarily mean that buffer functioning has improved. For example, De Partz (1995), argued that for individual AM functioning of the buffer did not improve but instead AM started using a strategy of lexical segmentation of words which would circumvent buffer impairment.

Clearly diagnosed and selective buffer impairments are not reported in many of the studies reviewed here, which may be one reason that not many studies show clear generalisation effects to control items. However, individuals with an impairment to both the lexicon and the buffer do also show generalisation (e.g., participant NM, Raymer et al., 2003). Raymer and colleagues (2003) explained the generalisation found in their patient with a lexical and buffer impairment as an interaction of training effects both in the lexicon and the buffer. They suggested the lexical improvement consisted of an improvement of trained words in the lexicon. These lexical improvements were argued to have a knock-on effect on the buffer helping to better maintain activation of words orthographically similar to the target. Raymer et al. also argued that the severity of the impairment to the orthographic lexicon could limit further generalisation effects, which may explain why not all individuals with an impairment to both the lexicon and the buffer show generalisation after treatment (e.g., Raymer et al., 2009, Beeson, 1999).

Additionally, other cognitive or linguistic functioning can influence the outcome of treatment and generalisation in particular. Treatment tasks call upon other cognitive processes than just those involved in writing. Copying a word in treatment for example requires processing visual information and reading whether the spelling is written correctly (Beeson et al., 2003; Raymer et al., 2009). Beeson et al. (2003) argued for a link between a limited response to treatment and poor performance on tasks of visual processing such as visual lexical decision task and the Raven's progressive matrices (Raven, Court, & Raven, 1990). Therefore, it seems there could be a link between severity of impairment and generalisation. Individuals with more severe damage or an impairment affecting several different levels of processing may be less likely to show generalisation. An improvement of untreated items is usually smaller than an improvement of treated items. Consequently, when a more severe deficit leads to a smaller improvement of treated items, an additional significant improvement for untreated items is less likely to occur. For example, some individuals that showed item specific improvements only (no generalisation, Table 7) also seemed to show small or short lasting gains for treated items (e.g., MR, MB, GP, & JF, Beeson et al, 2003; MMD, Rapp & Kane, 2002; CM & DY, Robson et al, 2001).

Furthermore, lexical-semantic processing impairments can have an effect on the learning of new verbal information (Rapp, 2005; Martin, Fink, & Laine, 2004, cited in Rapp, 2005). Beeson et al. (2003) also found a positive correlation between outcomes on semantic tasks (e.g., word-picture matching) and the benefits of treatment. In addition, monitoring performance by providing feedback in the event of an error has been linked to executive functions (Raymer et al., 2009).

To summarise, it seems that even though the graphemic output buffer has been argued to play an important role in generalisation, the variety of impairments found in this review shows there is no clear link between the type of impairment and the occurrence of generalisation, in part because many participants show an impairment to multiple components of the spelling process. Consequently, type of impairment cannot be the only contributing factor in generalisation. Furthermore, although a number of studies point out the possible role of other cognitive factors in treatment, few studies report what the exact contributing linguistic and cognitive individual factors are and how they might play a role.

Treatment method. The second factor that may play a role in generalisation is the type of lexical treatment given to the individual. A number of lexical treatments involve repetitive presentations of a word's written form (copying, e.g., CART, ACT). This form of treatment is argued to be beneficial as repeated exposure strengthens representations or access to the representations of treated items in the long-term memory (Rapp & Kane, 2002; and see Kohnen, Nickels, Coltheart, and Brunsdon, 2008, for a similar explanation in a developmental study). Copying could furthermore be effective in producing generalisation considering this method focuses on a word's orthography, and certain parts can be common to more words. We will explain how generalisation can occur to items with similar orthography via a mechanism of feedback in the next section on treatment items.

Nevertheless, not all participants treated with a copying task show generalisation effects (e.g., participant MD, Rapp & Kane 2002; FME, Schmalzl & Nickels, 2006). Additionally, treatments that did not use copying tasks have produced generalisation. For example, Behrmann (1987) used writing to dictation and sentence completion tasks that led to generalisation to untreated irregular words. Overall, the *type of treatment* cannot be the critical factor.

Treatment items. A variety of stimulus types both for treated and control items were used in treatment studies that led to generalisation. First, generalisation has been reported for words that are semantically related to the treated word: to synonyms of treated items (Cardell & Chenery, 1999) or to untreated words in the same semantic category as the treated words (e.g., animals) (Spencer et al., 2000). This improvement can be explained by assuming that synonyms or semantically related items co-activate related concepts, which in turn will activate their orthographic forms. We have also addressed this in the section on semantic treatment studies (Category 3.2, Table 6).

Secondly, orthographic similarity between items seems to play a role in generalisation. Raymer et al. (2003) used two sets of words and non-words, and a list of untreated words that shared half their letters with the treated word either at the beginning or the end of the word. Generalisation was found for words with a similar orthography to the treated items, where for example the item *"rac*coon" improved after treating the item "*rac*ket". Furthermore, Sage and Ellis (2006) matched items for written frequency and orthographic neighbourhood size, and generalisation to neighbours of a set of treated items was reported. Harris et al. (2012) also investigated the role of orthographic neighbourhood in remediating a case of graphemic output buffer impairment. They reported an improvement of untreated neighbours with the same medial position letters (e.g., clock – block).

In a recent paper, Goldrick, Folk, and Rapp (2010) showed the importance of dimensions of lexical frequency and neighbourhood density for the activation of non-target words in both spoken and written production. Looking at data from the developmental dysgraphia literature, studies also show that neighbourhood size and frequency of items are important factors influencing generalisation (Brunsdon, Coltheart, & Nickels, 2005; Kohnen et al., 2008). In one case of developmental surface dysgraphia generalisation was best predicted by neighbourhood size and frequency (Kohnen et al., 2008).

Studies that have shown generalisation to orthographic neighbours (Harris et al., 2012; Sage & Ellis, 2006) or orthographically similar words (Raymer et al., 2003) allow particular insight into how improvement of specific items can generalise as a result of interaction in processing. Sage and Ellis (2006) explained how a word in the graphemic output buffer can receive support from the activation of its neighbours in the orthographic lexicon. A word with many neighbours (e.g., cask) will activate the letter units c, a, s, k, in the graphemic output buffer. This activation is bidirectional, which means the activation in the buffer feeds back up to the lexicon. In the lexicon other words that share letters with the target word "cask", such as its neighbours (cash, case, task, etc.), will in turn also be activated. These representations will provide extra top-down support to the letter units in the buffer. As a result, the letters in a word with many neighbours like "cask" receive more support compared to a word like "edit", which has fewer neighbours (Sage and Ellis, 2006). The number of orthographic neighbours has also been found to have an influence on spelling performance in unimpaired individuals: words with a large number of orthographic neighbours are spelled aloud faster and more accurately compared to words with a small neighbourhood size (Roux & Bonin, 2009).

It could be that the acquired cases described here showed generalisation to control items that were high in frequency, neighbourhood size, or were direct neighbours of the target. However, not many of the studies discussed here specified all these item characteristics for the treatment or control items. Consequently, it is not possible to ascertain the exact influence of the type of items on the occurrence of generalisation effects. Remarkably, only the study described by Behrmann (1987) seemed to have controlled for regularity of the items.

Finally, an individual's item specific performance can determine whether untreated items are likely to generalise. Brunsdon et al. (2005) reported generalisation to untreated irregular items after treating irregular word spelling in a case of developmental surface dysgraphia. It was found that misspellings that were closer to being correct before treatment were more likely to generalise. Assuming the process of relearning exploits residues of original learning, it seems likely that items with larger "residual knowledge" are more likely to reach improvement (Weekes & Coltheart, 1996).

Discussion

Having reviewed around 40 treatment studies that aimed to improve a participant's spelling performance, it is clear that a variety of treatment methods and items have been administered to a large series of participants. This has led to a variety of outcomes and interpretations.

Not surprisingly, with such a large number of studies, there are differences in type of participants, items and treatment. We organised the sections above by the different methods of treatment that were administered (e.g., sub-lexical and lexical treatments). As noted above, the results of treatment may differ even with the same method and may be the same with different methods. For example, both a lexical and a sub-lexical treatment can result in generalisation across items. Hence, in drawing together the studies, we now focus on the different results of treatment and attempt to extract general principles of rehabilitation and generalisation that may underlie the different treatment results.

1. Direct treatment effects on representations or processes. While the focus of this review has been on generalisation, it is the case that many of the treatments we reviewed do not result in generalisation but are successful in the sense that they improve spelling of the items or skills that were treated. These can be both lexical or sub-lexical improvements, we will discuss each in turn.

Lexical treatment consisting of repetitive practice of a set of words, often significantly improves the spelling of those specific words treated (e.g., for 80% of the individuals that received treatment aimed at strengthening lexical representations (Category 3.3)). However, treatments that have as their aim a compensatory approach can also result in improvements on the stimuli used in treatment (e.g., Hillis & Caramazza, 1987). For example, when a spell-checker is used in combination with a certain set of words, repetitive practice can also have an item specific effect improving those items that have been practiced in treatment (participant SV, Beeson et al., 2000). The mechanism underlying this item specific improvement is often suggested to be increased accessibility of the orthographic form, perhaps through priming of the representation, strengthening of the connections from the meaning to the form or from the form to graphemes in the buffer, or by 'reteaching' the orthographic form (e.g., Beeson, 1999; Beeson et al., 2003; Panton & Marshall, 2008; Rapp & Kane, 2002). While these effects are by nature item specific, when the same set of representations are accessed from several modalities, generalisation may occur. For example, Kiran (2003) demonstrated generalisation from writing to dictation to written naming, and argued that after treatment the strengthened representations were accessible from both input modalities.

In a similar way, treatment that focuses on teaching a set of phoneme-grapheme correspondences is generally successful for the specific PGCs that are taught (e.g., Beeson et al., 2010). However, there is usually generalisation of the use of those PGCs beyond the particular stimuli (words or non-words) which were used in treatment (e.g., De Partz et al., 1992).

In addition, it has been suggested that there can be generalisation beyond the treated items as a direct effect of treatment resulting in a general improvement of processing. Weekes and Coltheart (1996) argued that this general effect has to do with the procedure by which representations are accessed: if this procedure is damaged, and if treatment improves this procedure, this will benefit the spelling not only of treated, but also other, untreated words. For example, an improvement for irregular words was explained by Behrmann (1987) as a consequence of improvement in general lexical processing and an improved 'lexical check' mechanism to reject incorrect spellings of words. Similarly, improved functioning of the graphemic output buffer should generalise to all words (e.g., Panton & Marshall, 2008; Rapp, 2005). Moreover, if the improved process is a more central process (e.g., access to lexical representations or the graphemic output buffer) generalisation will occur to all types of spelling: written naming and writing to dictation, written and oral spelling. However, given that a large proportion of studies do not show generalisation of treatment effects, we believe that there is some doubt regarding the extent to which general access procedures can indeed be improved by treatment. Indeed it is also possible that some apparent generalisation is in fact a result of repeated probing during treatment (cf., Nickels, 2002 for improvements in spoken naming as a result of repeated probing).

2. Interactive processing and summation of activation. In addition to direct effects of treatment on processes and representations, generalisation has also been proposed to occur as a result of the interactive and summative nature of the language processing system. These characteristics allow an untreated item to 'benefit' from the activation of a treated item during treatment.

One type of interactive processing that has been proposed to underlie generalisation is the interaction between lexical and sub-lexical information in spelling, and the summation of these two sources of information. For example, a strategy of using oral spelling (spelling each letter aloud) to arrive at written spelling has been suggested to improve the transmission of information between lexical and sub-lexical spelling processes (Pound, 1996). When lexical processing is impaired, strengthened sub-lexical processing can assist lexical spelling by providing additional activation in two ways: First, the sublexical information can provide additional activation of the letters in the buffer if they are already partially activated by the lexical route. In addition, this activation can feedback to the lexicon, which can provide further support to the lexical activation, or, if the lexical entry is insufficiently activated for selection could provide enough additional activation for this to be possible (De Partz et al., 1992). The results obtained by Beeson et al. (2000) further support the claim that lexical and sub-lexical spelling routes interact. Their participant's error types reflected a combination of lexical and sub-lexical information, and the authors argued their treatment facilitated this interaction between lexical and sub-lexical processes, improving spelling accuracy (Beeson et al., 2000).

Feedback between the lexicon and the graphemic output buffer, alluded to above, has been discussed in a number of studies (e.g., McCloskey et al., 2006; Roux & Bonin, 2009). A number of treatment studies reviewed here have provided evidence that generalisation might be based on this interaction.

As mentioned in the Introduction, McCloskey et al. suggest that in spelling, feedforward and feedback activation create an activation loop between the lexicon and the level of the buffer (McCloskey et al., 2006). When a word is activated in the lexicon, this will activate the corresponding letter units in the graphemic output buffer. These letter units feed their activation back up to the lexicon to strengthen the lexical representation, and as a result not only the target representation will receive activation but also orthographically similar lexical representations that share letter units with the target. These co-activated representations will in turn send activation to the letter units in the buffer.

This mechanism of feedback of activation leads to two predictions regarding spelling and treatment of dysgraphia. First, in the process of spelling a word can receive support from the co-activation of its orthographic neighbours. Hence, words with more neighbours are predicted to respond better to treatment. Indeed, as support for the feedback mechanism, it has been found that un-impaired adults spell aloud words with many neighbours faster and more accurately compared to words with a small neighbourhood size (Roux & Bonin, 2009).

Secondly, this mechanism of feedback and co-activation of related lexical representations, predicts that neighbourhood size plays a role in generalisation. When words are presented in treatment, untreated items will also be activated via this same mechanism of feedback: untreated words with many neighbours or words that are orthographically related to the treatment targets will benefit from this co-activation the most.

A number of treatment studies in the acquired and developmental dysgraphia literature have provided evidence supporting the fact that generalisation might be based on this interaction, by demonstrating that neighbourhood is an important factor predicting generalisation. It has been shown that words high in frequency and neighbourhood size are more likely to generalise (e.g., Kohnen et al., 2008), and generalisation has also been found to neighbours of treated words (Harris et al., 2012; Sage & Ellis, 2006).

The principle of feedback is also the basis of interactive models of language production, which assume interactivity between different levels of language processing such as semantics, phonology and orthography (e.g., a framework of lexical processing as described by Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg, 2005). In these models, orthography for spelling would receive input from both semantics and phonology. While phonology may not be an obligatory mediating step between semantics and orthography (see Tainturier & Rapp, 2001, for a discussion of this issue), it may nevertheless be available and when this is the case may support spelling[4]. When phonological processing is unimpaired, this could be an avenue for treatment: When links between semantics and orthography are impaired, phonology could provide an additional source of activation to support spelling. A focus of treatment may therefore be to try and 'promote' the interaction between orthography and phonology, for example by combining written treatment with spoken repetition (Beeson & Egnor, 2006).

This interactive model can furthermore be seen as an example of a 'primary systems' view of language, which argues that the systems of vision, semantics and phonology all interact and contribute to language processing (Jefferies, Sage, & Lambon Ralph, 2007). In their study, Jefferies et al. compared individuals with characteristics of deep dyslexia and deep dysgraphia (who produced semantic errors in reading and/or spelling), and argued for a severe impairment to a common underlying phonological system. Treatment inspired by these interactive models would tap into this interactivity. For example, treatment focusing on phonology (e.g., verbal repetition) could show generalisation to spelling. In this review we have not focused on studies with a treatment method outside the domain of spelling, however treatments designed to explore this interactive nature could further inform our understanding of the mechanisms and effects of interactivity in the process of spelling.

3. Strategies and compensatory skills. In addition to direct effects on representations and generalisation as a result of interactive processing, generalisation

[4] Note that this is not restricted to interactive models of spelling - a route for spelling via the phonological output lexicon is also possible in models like that depicted in Figure 1.

can also occur when a strategy or skill that was taught in treatment is applied to words outside treatment.

For example, treatment focusing on segmentation of words (e.g., into sounds or syllables) can have an item specific effect on the words practiced in treatment, but this strategy of reducing the workload for the buffer can generalise to the spelling other words that were not the target of treatment (De Partz, 1995).

Second, without directly treating poor spelling, individuals can be taught to improve their ability to revise their spelling errors, either by using a spell-checker, or by using problem-solving skills to become more aware of their errors and correct them (e.g., Beeson et al., 2000; Hillis & Caramazza, 1987). For example, participant SW (Beeson et al., 2000) showed a large increase in self-corrected spelling errors after treatment, and the authors argue that for SW treatment probably was not effective by improving access to lexical representations, but rather resulted in improved problem-solving skills, which generalised to untreated words.

Finally, for individuals with better oral spelling than written spelling abilities, the focus of treatment can be to teach the participant to use these oral spelling skills to write a word, by first spelling the word aloud, and then using their sub-lexical skills to write the individual letters to dictation (Pound, 1996). These examples illustrate how treatment can facilitate a procedure of spelling words that might not have been available or have been used before treatment, which can result in generalisation.

Future Directions

In this review we have discussed different factors that could play a role in generalisation, and mechanisms of generalisation in spelling. The next step in improving our understanding of mechanisms of generalisation is to bring all these factors together. We suggest that a series of carefully controlled treatment studies should be used that would allow for systematic testing of the different factors that potentially influence generalisation. For example, treatment studies should control and manipulate the relationship between treated items and untreated controls (e.g., semantic neighbours, orthographic neighbours, high and low frequency controls, words and non-words).

Our review was unable to discern any clear relationship between impairment type and generalisation, when a relationship is predicted to exist by many theoretical accounts. Consequently, it is important to examine the influence of type of impairment by administering the same treatment method to individuals with different underlying impairments (cf., Rapp & Kane, 2002). At the same time, treatment should also be replicated across individuals with the same spelling impairment to allow us to further determine why the same treatment may have different effects for different individuals. It is possible that more subtle aspects of the impairment influence performance (e.g., different types of buffer impairment), or other factors such as executive functioning or oral language skills impact on treatment outcomes.

Furthermore, Hillis and Heidler (2005) pointed out that in gaining a better understanding of the mechanism of treatment, as well as understanding *what* to treat, we also need to focus on *how* to best treat these stimuli. Evidence from studies on mechanisms of recovery and learning could further inform our understanding of other contributing factors (Hillis & Heidler, 2005). For example, it would be important to evaluate the effects of intensity of treatment or the optimal number of items to treat, as has been performed for anomia therapy (Sage, Snell, & Lambon Ralph, 2011; Snell, Sage, & Lambon Ralph, 2010). Results from such analyses can further inform our understanding of the process of learning and rehabilitation in dysgraphia.

Finally, the role of treatment methods should be examined by investigating different treatments (e.g., lexical copying treatment, sub-lexical treatment) within the same individual and across individuals with the same spelling impairment.

Unfortunately, this may not be as simple as it seems, as at a given time in the process of rehabilitation, one type of treatment might be effective whereas another may not until after the first type of rehabilitation. Similarly, impairments to other aspects of the language and cognitive system may impact on the response to spelling treatment even when not directly involved in the spelling process. This reinforces the need for case series where factors influencing the effectiveness of different treatments can be investigated across individuals (Nickels, Howard, & Best, 2011).

Conclusion

We have presented a comprehensive review of the literature on generalisation after treatment of acquired spelling impairment. The aim of this review was to examine what determines the occurrence of generalisation, by investigating the link between the level of impairment, the method of treatment, and the outcome of therapy. We have reviewed different explanations that have been put forward for a mechanism of generalisation.

Our main findings were:

- No clear relationship emerged between the individual's impairment and whether generalisation occurred, despite clear predictions that such a relationship should exist. This could in part be because many participants show an impairment to multiple components of the spelling process. It is also possible that this reflects the spelling system being more interactive than often envisaged in the past. Consequently, impairments at one level of processing can influence and be influenced by activation at other levels of processing.
- Generalisation has been reported after different treatment methods, therefore a particular type of treatment does not seem critical.

- Interactive processing between lexical and sub-lexical processes, and between the lexicon and the graphemic output buffer, may play an important role in generalisation across items.
- When treatment has strengthened representations of treated items, generalisation across modalities can occur when these representations are accessible in different modalities (e.g., written naming and oral naming).

In addition, some studies have pointed out the importance of possible interactions with other factors, such as an individual's general language and/or cognitive abilities, and also principles of learning that might play a role in dysgraphia treatment. We have also provided suggestions for future research to further investigate the different contributing factors in series of carefully controlled case studies.

In sum, a better understanding of the mechanisms that drive generalisation can contribute to our understanding of the cognitive processes in written word production and their interrelationships, and furthermore will help us maximise treatment effects for an individual with acquired dysgraphia.

References

Aliminosa, D., McCloskey, M., Goodman-Shulman, R., & Sokol, S. M. (1993). Remediation of acquired dysgraphia as a technique for testing interpretation of deficits. *Aphasiology, 7*(1), 55-69. doi: 10.1080/02687039308249499.

Aphasia Treatment Website, Academy of Neurologic Communication Disorders and Sciences. Retrieved from: http://aphasiatx.arizona.edu/written_writing

Beauvois, M., & Derousne, J. (1981). Lexical or orthographic agraphia. *Brain*, *104,* 21-50.

Beeson, P. M. (1999). Treating acquired writing impairment: Strengthening graphemic representations. *Aphasiology*, *13,* 767-785.

Beeson, P. M. (2004). Remediation of written language. *Topics in stroke rehabilitation, 11*(1), 37-48.

Beeson, P. M., & Egnor, H. (2006). Combining treatment for written and spoken naming. *Journal of the International Neuropsychological Society*, *12*, 816–827.

Beeson, P. M. & Hillis, A. E. (2001). Comprehension and Production of Written Words. In Chapey, R. (Ed.) *Language Intervention Strategies in Aphasia and Related Neurogenic Communication Disorders,* (4th ed., pp. 572-604). Philadelphia: Lippincott Williams & Wilkins.

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

Beeson, P. M., & Rapcsak, S. Z. (2002). Clinical diagnosis and treatment of spelling disorders. In A. E. Hillis (Ed.), *The handbook of adult language disorders: Integrating cognitive neuropsychology, neurology and rehabilitation.* New York: Psychology Press.

Beeson, P. M., Rewega, M., Vail, S., & Rapsack, S. Z. (2000). Problem-solving approach to agraphia treatment: interactive use of lexical and sub-lexical spelling routes. *Aphasiology*, *14*, 551-565.

Beeson, P. M., Rising, K., Kim, E. S., & Rapcsak, S. Z. (2010). A Treatment Sequence for Phonological Alexia/Agraphia. *Journal of Speech, Language, and Hearing Research, 53*, 450–468.

Beeson P. M., Rising K., & Volk, J. (2003). Writing treatment for severe aphasia: Who benefits? *Journal of Speech, Language, Hearing Research, 46*, 1038–1060.

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

Behrmann, M., & Bub, B., (1992). Surface dyslexia and dysgraphia: Dual routes, single lexicon. *Cognitive Neuropsychology*, *9*, 209-252.

Behrns, I., Hartelius, L., & Wengelin, Å. (2009). Aphasia and computerised writing aid supported treatment. *Aphasiology*, *23*(10), 1276-1294

Bruce, C., Edmundson, A., & Coleman, M. (2003). Writing with voice: an investigation of the use of a voice recognition system as a writing aid for a man with aphasia. *International journal of language & communication disorders, 38*, 131–148. doi:10.1080/1368282021000048258.

Brunsdon, R., Coltheart, M., & Nickels, L. (2005). Treatment of irregular word spelling in developmental surface dysgraphia, *Cognitive Neuropsychology*, *22*(2), 213-251.

Bub, D., & Kertesz, A. (1982). Evidence for lexicographic processing in a patient with preserved written over oral single word naming. *Brain, 105,* 697–717.

Campbell, R. (1983). Writing nonwords to dictation. *Brain and Language, 19,* 153-178.

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: Evidence from a case of acquired dysgraphia. *Cognition*, *26*, 59-85.

Cardell, E. A., & Chenery, H. J. (1999). A cognitive neuropsychological approach to the assessment and remediation of acquired dysgraphia. *Language Testing*, *16*, 353-388.

Clausen, N. S., & Beeson, P. M. (2003). Conversational use of writing in severe aphasia. *Aphasiology, 17*, 625-644.

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

De Partz, M. (1995). Deficit of the graphemic buffer: Effects of a written lexical segmentation strategy. *Neuropsychological Rehabilitation*, *5*(1-2), 129-147.

Deloche, G., Dordain, M., & Kremin, H. (1993). Rehabilitation of confrontation naming in aphasia: Relations between oral and written modalities, *Aphasiology*, *7*(2), 201-216.

Ellis, A., & Young, A. (1988). *Human Cognitive Neuropsychology*. London: Lawrence Erlbaum Associates Inc.

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

Folk, J. R., & Rapp, B. (2004). Interaction of lexical and sub-lexical information in spelling: Evidence from nonword priming. *Applied Psycholinguistics, 25*, 565-585.

Goldrick, M., Folk, J., & Rapp, B. (2010). Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language*, *62,* 113-134.

Goldrick, M., & Rapp, B. (2002). A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology*, *16*(1-2), 20–55.

Goodman, R. A., & Caramazza, A. (1986a). Aspects of the spelling process: Evidence from a case of acquired dysgraphia. *Language & Cognitive Processes*, *1*(4), 263-296.

Goodman, R. A., & Caramazza, A. (1986b). Dissociation of spelling errors in written and oral spelling: the role of allographic conversion in writing. *Cognitive Neuropsychology*, *3*(2), 179-206.

Goodman, R. A., & Caramazza, A. (1986c). The Johns Hopkins Dysgraphia Battery (Baltimore: The Johns Hopkins University, unpublished).

Harris, L., Olson, A., & Humphreys, G. (2012). Rehabilitation of spelling in a participant with a graphemic buffer impairment: The role of orthographic neighbourhood in remediating the serial position effect. *Neuropsychological Rehabilitation: An International Journal*, *22*(6), 890-919.

Hatfield, F. M. (1983). Aspects of acquired dysgraphia and implications for re-education. In C. Code & D.J. Muller (Eds.), *Aphasia therapy* (pp. 157-169). London: Edward Arnold Publisher.

Hillis, A. E. (1989). Efficacy and generalization of treatment for aphasic naming errors. *Archives of Physical Medicine and Rehabilitation, 70*, 632-636.

Hillis, A. E., (1991). *Effects of separate treatments for distinct impairments within the naming process.* In Clinical Aphasiology Conference: Clinical Aphasiology Conference (1989 : 19th : Lake Tahoe, NV : June 1989) / : Pro-Ed(1991), p. 255-265.

Hillis, A. E. (1992). Facilitating written production. *Clinics in Communication Disorders*, *2*(1), 19-33.

Hillis, A. E., & Caramazza, A. (1994). Theories of Lexical Processing and Rehabilitation of Lexical Deficits. In G.W. Humphreys & M.J. Riddoch (Eds.), *Cognitive Neuropsychology and Cognitive Rehabilitation* (pp. 449-484). London: Lawrence Erlbaum Associates.

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

Hillis, A. E., & Heidler, J. (2005). Contributions and limitations of the cognitive neuropsychological approach to treatment: Illustrations from studies of reading and spelling therapy. *Aphasiology, 19*(10-11), 985-993.

Hillis, A. E. & Rapp, B. C., & Caramazza, A. (1999). When a rose is a rose in speech but a tulip in writing. *Cortex, 35*, 337-356.

Hillis Trupe, E. (1986). Effectiveness of retraining phoneme to grapheme conversion. In R. H. Brookshire (Ed.), *Clinical Aphasiology*, (pp. 163-171). Minneapolis, MN: BRK Publishers.

Howard, D. (1986). Beyond randomised controlled trials: the case for effective case studies of the effects of treatment in aphasia. *British Journal of Disorders of Communication, 21*, 89–102.

Jefferies, E., Sage, K., Lambon Ralph, M. A. (2007). Do deep dyslexia, dysphasia and dysgraphia share a common phonological impairment? *Neuropsychologia, 45,* 1553-1570.

Kay, J., Lesser, R., & Coltheart, M. (1992). *PALPA: Psycholinguistic Assessments of Language Processing in Aphasia.* Hove: Lawrence Erlbaum Associates.

Kertesz, A. (1982). *The Western Aphasia Battery.* Orlando, FL: Grune & Stratton.

Kiran, S. (2003). Training phoneme to grapheme conversion for patients with written and oral production deficits: A model-based approach. *Aphasiology*, *19*(1), 53-76.

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.

Lesser, R. (1990). Superior oral to written spelling: Evidence for separate buffers? *Cognitive Neuropsychology, 7*(4), 347-366.

Luzzatti, C., Colombo, C., Frustaci, M., & Vitolo, F. (2000). Rehabilitation of spelling along the sub-word-level routine. *Neuropsychological Rehabilitation, 10*, 249-278.

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., Silveri, M. C., & Caramazza, A. (1985). Cognitive analysis of a case of pure dysgraphia. *Brain and Language, 25*, 187-212.

Mortley, J., Enderby, P., & Petheram, B. (2001). Using a computer to improve functional writing in a patient with severe dysgraphia, *Aphasiology*, *15*(5), 443-461.

Murray, L. L., & Karcher, L. (2000). A treatment for written verb retrieval and sentence construction skills, *Aphasiology*, *14*(5-6), 585-602.

Nickels, L. A. (1992). The autocue? Self-generated phonemic cues in the treatment of a disorder of reading and naming. *Cognitive Neuropsychology, 9,* 155-182.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising and reviewing. *Aphasiology, 16* (10/11), 935-979.

Nickels, L., & Best, W. (1996). Therapy for naming disorders (Part II): Specifics, surprises and suggestions. *Aphasiology*, *10*(2), 109-136.

Nickels, L., Howard, D., & Best, W. (2011). On the use of different methodologies in cognitive neuropsychology: Drink deep and from several sources. *Cognitive Neuropsychology*, 28, 475-485.

Nickels, L., Kohnen, S., & Biedermann, B. (2010). An untapped resource: Treatment as a tool for revealing the nature of cognitive processes. Cognitive Neuropsychology, 27(7), 539-562. doi: 10.1080/02643294.2011.609811.

Orjada, S. & Beeson, P. (2005). Concurrent treatment for reading and spelling in aphasia, *Aphasiology*, *19*(3-5), 341-351.

Panton, A. & Marshall, J. (2008). Improving spelling and everyday writing after a CVA: A single-case therapy study, *Aphasiology*, *22*(2), 164-183.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56-115.

Pound, C. (1996). Writing remediation using preserved oral spelling: A case for separate output buffers, *Aphasiology*, *10*(3), 283-296.

Rapp, B. (2002). Uncovering the cognitive architecture of spelling. In A. E. Hillis (Ed.), *The handbook of adult language disorders: Integrating cognitive neuropsychology, neurology and rehabilitation* (pp. 47-69). New York: Psychology Press.

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. doi: 10.1080/02687030544000209.

Rapp, B., & Caramazza, A. (1997). From graphemes to abstract letter shapes : Levels of representation in written spelling. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1130-1152.

Rapp, B., Epstein, C., & Tainturier, M. J. (2002). The integration of information across lexical and sub-lexical processes in spelling. *Cognitive Neuropsychology*, *19*(1), 1-29.

Rapp, B., & Glucroft, B. (2009). The benefits and protective effects of behavioural treatment for dysgraphia in a case of primary progressive aphasia. *Aphasiology*, *23*(2), 236-265.

Rapp, B., & Kane, A. (2002). Remediation of deficits affecting different components of the spelling process. *Aphasiology, 16*(4-6), 439-454. doi: 10.1080/02687030244000301.

Raven, J. D., Court, J. H., & Raven, J. (1990). *Coloured Progressive Matrices.* London: Oxford Psychologists Press.

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

Raymer, A., Strobel, J., Prokup, T., Thomason, B., & Reff, K-L. (2009). Errorless versus errorful training of spelling in individuals with acquired dysgraphia. *Neuropsychological Rehabilitation*, *20*(1), 1-15.

Robson, J., Marshall, J., Chiat, S., & Pring, T. (2001). Enhancing communication in jargon aphasia: a small group study of writing therapy. *International Journal of Language & Communication Disorders*, *36*(4), 471–488.

Robson, J., Pring, T., Marshall, J., Morrison, S., & Chiat, S. (1998). Written communication in undifferentiated jargon aphasia: a therapy study. *International Journal of Language & Communication Disorders, 33*(3), 305-328.

Roeltgen, D. P., Sevush, S., & Heilman, K. M. (1983). Phonological agraphia: Writing by the lexical-semantic route. *Neurology, 33*, 755-765.

Roux, S., & Bonin, P. (2009). Neighborhood effects in spelling in adults. *Psychonomic Bulletin & Review, 16*(2), 369-373.

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

Sage, K., Snell, C., & Lambon Ralph, M. A. (2011). How intensive does anomia therapy for people with aphasia need to be? *Neuropsychological Rehabilitation, 21* (1), 26-41.

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

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science, 14*:238. DOI: 10.1111/j.0963-7214.2005.00372.x

Shallice, T. (1981). Phonological agraphia and the lexical route in writing. *Brain, 104*, 413-429.

Snell, C., Sage, K., & Lambon Ralph, M. A. (2010). How many words should we provide in anomia therapy? A meta-analysis and a case series study. *Aphasiology, 24*(9), 1064-1094.

Spencer, K. A., Doyle, P. J., McNeil, M. R., Wambaugh, J. L., Park, G., & Carroll, B. (2000). Examining the facilitative effects of rhyme in a patient with output lexicon damage, *Aphasiology*, *14*(5-6), 567-584.

Tainturier, M. J., & Rapp, B. (2001). The spelling process. In B. Rapp (Ed.), *What Deficits Reveal about the Human Mind/Brain: A handbook of cognitive neuropsychology* (pp. 263-289). Philadelphia: Psychology Press.

Weekes, B., & Coltheart, M. (1996). Surface Dyslexia and Surface Dysgraphia: Treatment Studies and Their Theoretical Implications. *Cognitive Neuropsychology*, *13*, 277-315.

Willmes, K. (1990). Statistical methods for a single case study approach to aphasia therapy research. *Aphasiology, 4,* 415–436.

Whitworth, A., Webster, J., & Howard, D. (2005). *A Cognitive Neuropsychological Approach to Assessment and Intervention in Aphasia.* Hove, E. Sussex: Psychology Press.

STUDY FOUR

The Effect of Orthographic Neighbourhood on Treatment of Acquired Dysgraphia

Abstract

Studies on the nature of processing within the spelling system have provided evidence for interactive processing, where activation between different levels of the spelling system flows bidirectionally. In particular, activated letters at the level of the graphemic output buffer feed back activation to the lexicon, supporting selection of the target. As a consequence lexical competitors are activated leading to an effect of orthographic neighbourhood size on spelling (i.e., better spelling for words that are orthographically related to many other words, e.g., grade: trade, grape, etc., compared to words with no neighbours). The aim of the current study was to further examine the nature of this mechanism of interactivity, and in particular the role of orthographic neighbourhood size, in the treatment of two individuals with acquired dysgraphia. To investigate whether neighbours could provide extra support to target words in treatment, two phases of treatment were conducted: in the first phase, treated words had no orthographic neighbours, and in the second phase, treated words had many neighbours. Untreated control sets were used to investigate the influence of neighbourhood size on potential generalisation across items.

Results showed that neighbourhood size did not influence the size of the treatment effect for either participant, and no clear evidence was found for generalisation. It is argued that the amount of feedback in these individuals with graphemic output buffer impairment has decreased, and as a result provides reduced activation of orthographic neighbours.

Introduction

Research regarding the cognitive processes of written word production has identified key components of the spelling process (e.g., Rapp, 2002), and data from acquired spelling impairments have proven important in developing these theories (e.g., Caramazza, Miceli, Villa, & Romani, 1987; Goodman & Caramazza, 1986; Rapp & Caramazza, 1997; Shallice, 1981). A topic that has received recent attention is the nature of the flow of activation between the components of written word production, and, more specifically, whether there is evidence for interactive processing within the spelling system (Buchwald & Falconer, 2014; Falconer & Buchwald, 2013; Folk, Rapp, & Goldrick, 2002; McCloskey, Macaruso, & Rapp, 2006, Rapp, Epstein, & Tainturier, 2002).

Within the spelling system, one possibility is that activation flows strictly in one, feed forward, direction. Under this account, when attempting to write the name of a picture, semantic information leads to the activation of a target (e.g., *cat*) and semantic competitors (e.g., *dog, rat*) at the level of the lexicon. Some 'discrete' processing stage theories (e.g., Levelt, Roelofs, & Meyer, 1999 (speech production); Morton, 1969 (word recognition)) propose that only a single lexical unit (*cat*) is selected. In spelling, this single selected unit would activate its corresponding letters (*c, a, t*; for written production) (Rapp & Goldrick, 2000). However, it is also possible that activation 'cascades' from the lexical to the grapheme level with the result that the letters of not only the target, but also its competitors (*dog*, *rat*) are activated. Whether discrete or cascading, the strictly feed forward flow of activation means that there is no influence of the activated letters on the lexical level in these accounts.

In contrast, in interactive theories activation between the different levels of processing flows bidirectionally. In these theories the target's letters and those of its lexical competitors feed back information up to the lexical level, thereby sending activation to the lexical target and its formal neighbours (Rapp & Goldrick, 2000). See

Figure 1 for a representation of a discrete system and an interactive system including feedback.

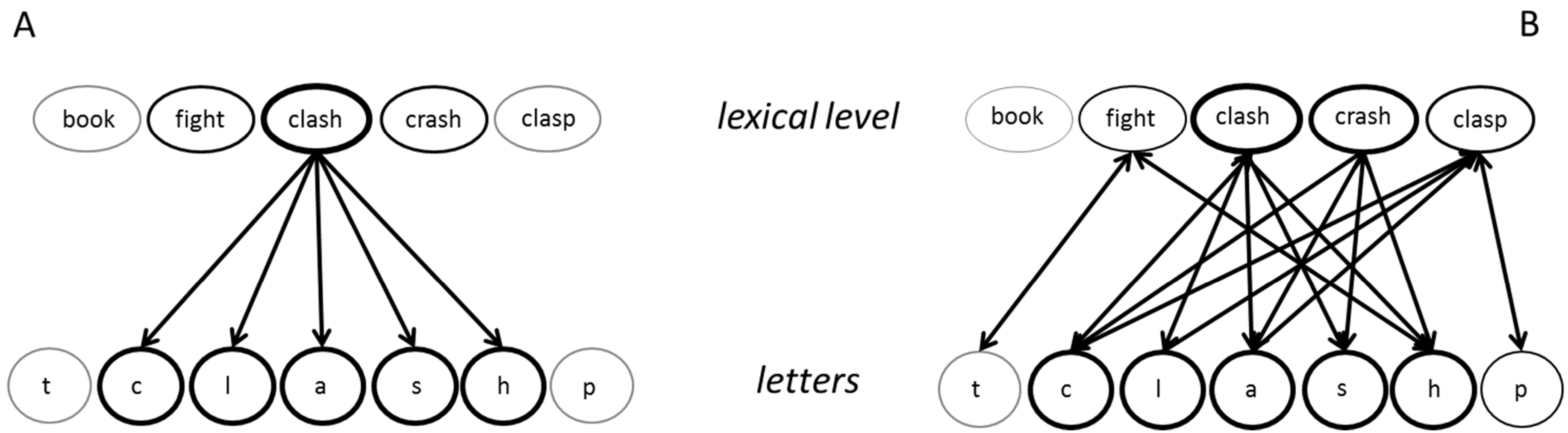


Figure 1: Representation of (A) a discrete account and (B) an interactive feedback account. Bidirectional arrows represent bidirectional flow of activation. The strength of activation of nodes is represented by the thickness of the node outline: target nodes are the most active and hence have the thickest outlines, and competitors have outlines varying in thickness according to their level of activation.

Although more extensively discussed in spoken word production (see Rapp and Goldrick, 2000, for a discussion of theories), the extent of interactivity has also been explored in written production. Studying error patterns of individuals with acquired dysgraphia has provided one source of evidence regarding the nature of activation flow in the spelling system. For example, McCloskey et al. (2006) argued that the error pattern shown by CM, an individual with acquired dysgraphia, was the result of interaction: A large proportion of CM's errors were real word errors that shared letters with the target (e.g., *arm: amber*), and also contained intruded letters from CM's previous responses. Preceding the response *amber,* for example, CM had written *bench*. McCloskey et al. argued that the activated letters *a, r, m* (for the target *arm)* fed activation back up to other lexical items that contained these letters, including the word *amber*. In addition, *amber* received activation from the shared letter nodes from the previous response: *b, e* (from *bench)*. As a result, activation for *amber* exceeded the activation level of the target *arm*. Hence, to explain the error pattern it was required that activation from the grapheme level could be fed back up to the lexicon, in this case resulting in the selection of an incorrect lexical item.

More recently, Falconer and Buchwald also explored the nature of processing between the lexical level and letter level through investigation of the error pattern in acquired dysgraphic individual RMI (Buchwald & Falconer, 2014; Falconer & Buchwald, 2013). RMI's semantic spelling errors showed a higher degree of orthographic overlap between the target and semantic error than expected by chance (e.g., *saint: priest*; Falconer & Buchwald, 2013). This pattern is also easily accounted for in an interactive account where target letters feed activation back to lexical selection processes. For example, when naming *saint,* feedback of activation from the letters *i*, *s* and *t* will not only support *saint*, but also any lexical competitor that shares letters with the target, such as *priest*. Consequently, semantic competitors that share letters with the target will be higher in activation than those that do not (e.g., *pope*) and more likely to be selected (see also *crash* vs. *fight* in Figure 1).

The spelling performance of unimpaired adults has also provided evidence for interactivity within the spelling system. For example, Roux and Bonin (2009) investigated the influence of orthographic neighbourhood size on oral spelling in unimpaired adults. They found that oral spelling of words with a large neighbourhood size, i.e., words for which there are many orthographically similar words (e.g., *clash* has five substitution neighbours: *flash, crash, class*, etc.) was faster and more accurate when compared to words with few or no neighbours (e.g., *script:* no substitution neighbours) (Roux & Bonin, 2009). This was explained as a result of interactive processing between the lexicon and buffer. As activated letters in the buffer feed activation back up to the lexicon, not only the target word will be activated, but words that share orthography with the target (i.e., neighbours) will also receive activation. In turn these neighbours reactivate their letters which will once again activate the target. For example, after a target word *clash* is activated in the lexicon, the corresponding letters *c, l, a, s, h*, are activated in the buffer and will feed activation back up to the lexicon, reactivating the target and activating words that share those letters (e.g., *crash*: see Figure 1 for a representation of this feedback system). Cycling of activation occurs between the lexicon and buffer for the target and its neighbours. Hence, a word with many neighbours receives more activation (is better supported) compared to a word with few neighbours, which predicts an advantage for spelling words with many orthographic neighbours.

However, it is possible that having many orthographic neighbours could be detrimental for spelling. For example, if co-activated, orthographically similar, candidates compete (e.g., through lateral inhibition), having many neighbours might inhibit performance. Even though in the area of visual lexical decision and reading aloud there is some evidence for inhibitory effects of (higher frequency) neighbours (e.g., Grainger, 1990; although see Andrews, 1997, for a review on and discussion of the effects of neighbourhood size), in the area of written word production no inhibitory effects from neighbours have been reported.

The effect of neighbourhood size on spelling has also been explored in treatment studies, in both developmental and acquired dysgraphia, and more specifically in the area of generalisation of the effects of treatment. When treatment is successful in improving untreated items or modalities, this is referred to as generalisation. Even though the relationship between the type of impairment, type of treatment and generalisation remains unclear (Krajenbrink, Nickels, & Kohnen, 2015), evidence from a number of treatment studies has suggested that the type of items used in treatment can influence generalisation.

Sage and Ellis (2006) specifically explored the role of orthographic neighbourhood in priming and treatment studies with individual BH, who showed characteristics of graphemic output buffer impairment. Sage and Ellis reported an effect of neighbours on short-term priming of word spelling: BH was asked to copy three neighbours (e.g., *assert, ascent, absent*) of a target word (*assent*) before writing this target word to dictation. Copying the three neighbours resulted in improved spelling of the target *assent*, compared to copying of three words that were matched to the target on frequency but unrelated in spelling (e.g., *powder, unhook,* and *timber*). Sage and Ellis argued that copying increased activation of the target's neighbours in the orthographic lexicon which provided additional top down support to the target in the buffer when the target was spelled (Sage & Ellis, 2006).

Following this priming experiment, Sage and Ellis (2006) used a treatment study to explore whether this priming could also result in long term improvements of spelling after treatment. Stimuli for the treatment study consisted of three matched sets: a set of treated words, a set of untreated words that each was a neighbour of a treated word, and an untreated set of words unrelated to any treated items (control set). After treatment, the treated items showed the largest improvement in spelling, but the neighbour set also showed significant improvement, whereas no generalisation was reported for the control set of untreated items that were orthographically unrelated to the treated items. Consequently, Sage and Ellis argued that the best way to improve spelling of a target word (e.g., *cask*) is by directly practicing its spelling. However, a word can also receive additional support when a target's neighbour (e.g., *case, bask*) has been primed during treatment. When this neighbour, *case,* is activated during treatment, feedback would provide extra activation to the letters of the target *cask*, and therefore *cask* will benefit more from treatment and be more likely to overcome impairment at the buffer (Sage & Ellis, 2006). Hence, in this study, generalisation to neighbours of treated items was used as evidence for interactivity in processing between the level of the lexicon and the buffer.

The role of orthographic neighbourhood in generalisation of treatment was also investigated by Harris, Olson, and Humphreys (2012), and more specifically, whether the position of the letter change between neighbour pairs influenced generalisation. Harris et al. suggested that letters in medial position of the word were critical for generalisation to neighbours. Following anagram and copy treatment, JF, an individual with graphemic output buffer impairment showed improved spelling of untreated words that shared the medial letters with treated items (e.g.*, couch: pouch*), but treatment had a negative impact on untreated neighbours that did not share middle letters (e.g., *couch: coach*). Harris et al. (2012) argued that, after treatment, treated items supported spelling of untreated neighbours via top-down activation, and this was especially beneficial in the medial position of the word which is more error prone in graphemic output buffer dysgraphia. Although they do not explicitly refer to feedback, implicit in this account is that to support writing of the target, the neighbour must receive activation via feedback from the activated shared letters. We return to this issue in the Discussion, below.

Raymer, Cudworth, and Haley (2003) also reported improvement in spelling of untreated items that shared initial letters with the treated items (e.g., from *rac*ket to *rac*coon) in individual NM, who had severe acquired dysgraphia resulting from an impairment to the orthographic lexicon and the graphemic output buffer. The authors argued that NM might have been able to incorporate segments (syllables) of trained words (e.g., '*rac'* in the example) into untrained words with similar spellings. For example, after treating *racket*, NM could use knowledge of this treated word when spelling untreated words with a similar orthography, like *racoon*.

In addition, two single case studies of developmental dysgraphia demonstrated that neighbourhood size and lexical frequency influenced the likelihood of generalisation of the effects of treatment to untreated irregular words (Brunsdon, Coltheart, & Nickels, 2005; Kohnen, Nickels, Coltheart, & Brunsdon, 2008). Once again, orthographic similarity between treated and untreated words was argued to cause generalisation, through feedback between the grapheme level and the orthographic lexicon (Kohnen et al., 2008). Kohnen et al. (2008) also discussed predictions that followed from their results. First, if orthographic similarity between treated and untreated words is an important factor in generalisation, this would not only have an effect on direct neighbours of treated items, but treatment should also be beneficial for words with a large neighbourhood size. Words with many neighbours might benefit from treatment because (even if not direct neighbours) they are more likely to share orthography with treated items and hence be primed over the course of treatment.

Second, the neighbourhood size of treated items may affect how likely it is that generalisation occurs. Similar to the previous argument, it is possible that because words with a large neighbourhood size share orthography with many words, treatment of these words will result in co-activation of orthographically similar words, whereas words with a sparse neighbourhood would be less likely to produce this generalisation (Kohnen et al., 2008).

Furthermore, it may be the case that the neighbourhood size of treated items not only has an effect on generalisation, but also on the size of the treatment effect. When a word is being practiced during treatment, co-activated letters from neighbours can provide additional support to produce the correct spelling, similar to the facilitative effect of neighbourhood size in spelling for unimpaired adults (Roux & Bonin, 2009). Hence, treated items with many neighbours may respond better to treatment compared to words with a smaller neighbourhood size.

In sum, results from a number of treatment studies have provided further evidence for interactive processing in spelling, specifically from the influence of neighbourhood size on generalisation in individuals with graphemic output buffer impairment.

As a consequence of this feedback, it has been suggested that generalisation may be more likely to occur after treating words with many neighbours. Furthermore, generalisation may not only occur to direct neighbours of treated items, but that untreated words that have a generally large neighbourhood size may show greater generalisation. However, the role of neighbourhood size on the extent to which treated items improve has not been investigated to date. The current treatment study aimed to further investigate the mechanism of interactivity, by exploring the role of neighbourhood size on both treatment effects and generalisation in two individuals with graphemic output buffer impairment. We designed our treatment in order to investigate whether 1) neighbourhood size has an effect on the size of the treatment effect, and 2) whether neighbourhood size influences generalisation of treatment effects to untreated words.

Case Reports

GEC was a 69 year old, right-handed man who suffered a left middle cerebral artery stroke in March 2008. GEC held a university degree in business, and worked as a financial planner prior to his stroke. At the time of this study, five years post onset, GEC presented with non-fluent aphasia with word finding difficulties, and a severe writing impairment. GEC has also been reported in detail in Fieder, Nickels, & Biedermann (2015) and Krajenbrink, Nickels, and Kohnen (2015).

JOD was a 75 year old, left-handed man with 11 years of education who suffered a left hemisphere cerebral haemorrhage over 15 years prior to this study, which resulted in a right sided hemiplegia. He had worked in a rural area as a wool classer, and subsequently as a taxi driver. JOD reported to be left-handed, and that, while forced to use his right hand at school, he preferred writing with his left hand. At the time of treatment, JOD presented with dysarthria and a severe writing impairment, in the context of relatively good expressive and receptive language.

Pre treatment Assessment

Assessment was conducted over multiple sessions to investigate the nature of GEC and JOD's underlying language impairment. These assessments included the Comprehensive Aphasia Test (CAT; Swinburn, Porter, & Howard, 2004), subtests from the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1992) and other tests as appropriate. Here, we summarise the conclusions regarding their level of impairment drawn from these assessments, full results are reported in Appendices A-C. To compare performance across modalities, GEC and JOD were also assessed on oral and written naming, writing to dictation, repetition and reading of the same 60 items (Krajenbrink, Nickels, Kohnen, unpublished) and there was a detailed assessment of written processing.

GEC.

General language processing. GEC showed a mild semantic impairment reflected in performance on tests of conceptual semantics (Pyramids and Palm Trees Test; Howard & Paterson, 1992), and synonym judgements (PALPA; Kay et al., 1992). Input processing was mildly impaired in the visual modality, suggesting difficulties accessing the orthographic input lexicon (visual lexical decision, PALPA; Kay et al., 1992). Assessment of spoken production revealed an impairment in word retrieval, as scores on a word fluency task and picture description were below normal limits (CAT; Swinburn et al., 2004; see Appendix B for the results on a picture description task).

GEC showed a clear impairment in the written modality compared to the other modalities: his written picture naming was severely impaired compared to oral naming (McNemar's test exact, $p < .001$). Written naming and writing to dictation were equally accurate (word accuracy: McNemar's test exact, $p = .125$, two-tailed; letter accuracy: $t(59) = 0.70$, $p = .485$, two-tailed).

Written production: Influence of psycholinguistic variables. Over the course of assessment GEC wrote 705 words to dictation[1] with an overall accuracy of 17%. The different spelling lists that were administered to test for the influence of different variables are shown in Appendix D. A logistic regression was performed on a subset of 609 items (accuracy of 18%) for which data was available on written logarithmic frequency, imageability, number of orthographic neighbours, bigram frequency (average bigram token frequency across the entire letter string), length in letters (all obtained from N-Watch (Davis, 2005) using the CELEX database (Baayen, Piepenbrock, & van Rijn, 1995)) and age of acquisition (Kuperman, Stadthagen-Gonzales, & Brysbaert, 2012). This analysis showed that frequency, imageability and length were significant

[1] Some words were repeated on the different lists that were administered. For the regression analysis we only included the response from the first time the participants were asked to spell the word.

predictors of word accuracy (all $p < .030$), with errors more likely for long compared to short words, and better performance for words that were high in frequency and imageability (see Table 2).

Error analysis. Less than 1% of words resulted in 'no response'. The majority of overt errors GEC made when writing to dictation (79% of errors) resulted in a non-word response with close to 90% of these non-word responses being orthographically related[2] to the target, for example *origin*: *orgoin*). Only two phonologically plausible errors were made (*squad: squod*). GEC made only few morphological and semantic errors, which could also be classified as orthographically related (e.g., *pigeon: pig*).

Non-word spelling. GEC wrote 74 monosyllabic non-words to dictation (the Diagnostic Spelling Test – non-words: DiSTn, Kohnen, Nickels, & Castles, 2009), resulting in only three items written correctly. Most responses were non-words, with only 17% of all errors resulting in a word, of which 85% were orthographically related (e.g., *leet: let*). Spelling of non-words and words was compared using a subset of 20 non-words was matched for length to a set of 20 monosyllabic irregular words (from the Krajenbrink et al. (unpublished) list, see Appendix D). Spelling non-words (0 correct) was more impaired than spelling words (4 correct), which approached significance (Fisher exact, $p = .053$, one-tailed).

Copying. GEC showed intact peripheral processing of retrieving and producing letter shapes (see Appendix C): he performed flawlessly at copying words in sight. However, performance decreased when a delay was introduced, and the longer the delay, the worse his performance (short and long words combined, Jonckheere Trend test, $z = 6.19$, $p < .001$).

JOD.

[2] We defined an error as orthographically related when either at least 50% of target letters were in the response (task: trash), or at least 50% of response letters were target letters (hatred: hit; based on Nickels' (1995) analysis of phonological errors in spoken production).

General language processing. JOD showed intact semantic processing and access to semantics from auditory and visual modalities reflected in performance within normal limits on spoken and written word-picture matching and synonym judgement tasks (PALPA; Kay et al., 1992). JOD's input processing was mildly impaired for the visual modality as shown by scores below the normal range on visual lexical decision (PALPA; Kay et al., 1992), indicating an impairment in accessing the orthographic input lexicon. Spoken production tasks showed an impairment in word retrieval on both a fluency task and a picture description task (CAT; Swinburn et al., 2004; see Appendix B).

Cross modality testing of the same 60 items showed a clear impairment to written production. JOD showed a significant difference between oral and written picture naming (McNemar's exact, $p < .001$). However, no significant difference was found between writing to dictation and written naming (word accuracy: McNemar's exact: $p = .648$, two-tailed; letter accuracy: $t(59) = 1.12$, $p = .269$, two-tailed).

Written production: influence psycholinguistic variables. JOD wrote 585 words to dictation with an accuracy of 25%. Logistic regression analysis on the 531 items (accuracy of 26%) with full psycholinguistic data showed that written frequency, imageability, bigram frequency and length were significant predictors of word accuracy (written frequency, imageability and length all $p < .001$; bigram frequency $p = .039$): accuracy increased for words high in frequency and imageability and words with more frequent bigrams, and long words were more error prone compared to short words.

Error analysis. No response was given to 5% of items. The majority of JOD's overt errors were non-words (82%), with 97% of these non-words being orthographically related to the target. Most errors consisted of letters being deleted or in the incorrect position (*ocean: ocan, radio*: *raido).* No morphological or semantic errors were observed. 5% of errors resulted in a phonologically plausible spelling, however

71% of these also resulted from just a single letter substitution, deletion or transposition (e.g., *salary: salery; scene: sceen*).

Non-word spelling. When asked to spell 74 monosyllabic non-words to dictation (DiSTn; Kohnen et al., 2009), JOD wrote only one item correctly, and almost all responses (95%) were non-words. The majority of JOD's responses (70%) consisted of only one or two letters, and in 82% of errors the initial letter was spelled correctly. When comparing a subset of 20 items from the DiSTn (Kohnen et al., 2009) and 20 monosyllabic irregular words matched for length from Krajenbrink et al. (unpublished) the difference between words (three correct) and non-words (0 correct) was not significant (Fisher exact, $p = .115$, one-tailed).

Copying. JOD was unable to copy cross case from upper case to lower case - he reported he didn't 'know how', and that he always wrote in upper case and always had. However, from lower case to upper case, he was almost flawless at copying words in sight, indicating intact peripheral spelling processes. However, performance decreased when a delay was introduced, and the longer the delay, the worse his performance (short and long words combined, Jonckheere Trend test, $z = 5.34$, $p < .001$).

Summary. GEC and JOD showed good single word comprehension. Both individuals suffered impaired spoken word retrieval in spontaneous speech, and GEC also showed a mild semantic impairment. JOD and GEC were particularly impaired in written word production compared to spoken production. Both participants showed characteristics of an impairment to the graphemic output buffer: 1) poor performance independent of input modality; 2) an effect of length, with more errors in long words compared to short words, 3) in a copying task performance decreased as the delay increased, 4) errors could predominantly be classified as 'letter errors', with a large proportion of errors involving deletion of letters.

GEC and JOD did also show an effect on accuracy of a number of lexical variables, such as frequency and imageability. The interpretation of these effects is unclear: According to some authors, these effects are incompatible with graphemic output buffer impairment and instead reflect an additional orthographic output lexicon impairment (Caramazza et al., 1987). However, other authors argue that this may reflect the effects of lexical support on the buffer (e.g., Sage & Ellis, 2004).

Table 1

Initial assessment of written production and cross modality testing

Task	N of items	GEC %	JOD %
Single Letter processing			
Cross-case copying letters	28	100	57
- lower – upper	14	100	85
- upper – lower	14	100	29
Writing letter names to dictation	26	100	100
Writing letter sounds to dictation (DiSTs)	32	63	66
Writing: sublexical			
Non-word spelling (DiSTn)	74	4	1
Cross modality testing (KNK)			
Oral naming	60	88	95
Written naming	60	13	23
Writing to dictation	60	5	18
Repetition	60	97	98
Reading	60	100	100

Note. DiSTs: Diagnostic Spelling Test – Sounds (Kohnen, et al., 2009); DiSTn: Diagnostic Spelling Test – Non-words (Kohnen, et al., 2009): See www.motif.org.au; KNK: Krajenbrink, et al. (unpublished).

Table 2

Logistic regression predicting word accuracy GEC (n=609 words)

	B	SE	Wald	*df*	*p*	Odds Ratio	95% C.I. for Odds ratio	
							Lower	Upper
Log written frequency	0.815	0.232	12.309	1	<.001 ***	2.259	1.433	3.561
Number of letters	-0.595	0.142	17.589	1	<.001 ***	0.552	0.418	0.728
Imageability	0.003	0.001	5.073	1	.024 *	1.003	1.000	1.005
Bigram frequency	0.000	0.000	3.679	1	.055	1.000	1.000	1.000
Number of orthographic neighbours	0.027	0.032	0.675	1	.411	1.027	0.964	1.094
Age of Acquisition	-0.061	0.064	0.898	1	.343	0.941	0.830	1.067
Constant	-0.982	1.259	0.609	1	.435	0.374		

* $p < .05$. *** $p < .001$.

Table 3

Logistic regression predicting word accuracy JOD (n=531 words)

	B	SE	Wald	*df*	*p*	Odds Ratio	95% CI for Odds Ratio	
							Lower	Upper
Log written frequency	0.840	0.241	12.182	1	<.001 ***	2.317	1.445	3.713
Number of letters	-0.500	0.130	14.859	1	<.001 ***	.606	.470	.782
Imageability	0.005	0.001	13.612	1	<.001 ***	1.005	1.002	1.007
Bigram frequency	0.000	0.000	4.277	1	.039 *	1.000	1.000	1.000
Age of Acquisition	-0.123	0.067	3.427	1	.064	.884	.776	1.007
Number of orthographic neighbours	0.049	0.035	1.948	1	.163	1.050	.981	1.124
Constant	-1.633	1.240	1.736	1	.188	.195		

* $p < .05$. *** $p < .001$.

Treatment Program

The aim of this treatment study was to explore the role of neighbourhood size on 1) the effect of treatment, and 2) on generalisation of treatment effects.

Stimuli

There were two sets of treatment stimuli and three control sets, with 20 words in each set. Choice of these stimuli was related to our research questions. We used the CELEX frequency database (Baayen et al., 1995) and the N-Watch database (Davis, 2005) to calculate the different psycholinguistic variables. See Table 4 and Appendix E for these data for the treated and untreated sets.

1. Does neighbourhood size play a role in the effect of treatment? To answer this question we manipulated the neighbourhood size of the treated items. In the first phase of treatment, the set of treated items (T1) consisted of words that had no orthographic substitution neighbours[3] ('Coltheart's N': Coltheart, Davelaar, Jonasson, & Besner, 1977), for example there are no words that are one letter different to the word *pebble*. In contrast, the words treated in Phase Two (T2) were selected to have many orthographic substitution neighbours (e.g., *rocket* - six substitution neighbours: *socket, racket, rocker*, etc.). We hypothesised that if neighbourhood size plays a facilitating role in treatment, words with many neighbours would respond better to treatment compared to words with no neighbours.

In our consideration of neighbourhood size we took into account the number of neighbours as well as the frequency of the neighbours, as it has been found that both neighbourhood size and neighbourhood frequency are important factors in word processing (e.g., Grainger, 1990). We defined 'many neighbours' as five or more neighbours, or four neighbours including at least one higher frequency neighbour.

---

[3] From this point on, when we refer to neighbours, we mean orthographic substitution neighbours.

Sets T1 and T2 were matched on length in letters and frequency (total frequency, written frequency, and written logarithmic frequency, from the CELEX frequency database (Baayen, et al., 1995)). We were not able to control for phonological neighbourhood size across sets: the words in T2 (many orthographic neighbours) had a larger phonological neighbourhood size compared to the words in T1 (no orthographic neighbours). We hypothesized that if neighbourhood size facilitates improvement as a result of treatment, then set T2 should show greater improvement than set T1.

2. Do the effects of treatment generalise to untreated stimuli and does neighbourhood size play a role in any generalisation? Three sets of control words were used to evaluate whether there was generalisation to untreated words. The words in these sets were tested during pre-treatment baseline assessment and during post-test assessments but were not practiced in treatment.

C1: words with no orthographic substitution neighbours. The words in this set were not neighbours of any treated word, and were matched to the treated items in Phase One (T1) on frequency and length. See Appendix E for more details.

C2: direct neighbours of the treated words in Phase Two (T2). For example, the word *grave* was treated in T2, its neighbour *grade* was administered to test for generalisation.

C3: words with many neighbours in general and that were not orthographic neighbours of any other word in the treated sets.

The sets of words with neighbours (T2, C2, and C3) were matched on frequency (total, written, and written logarithmic frequency, from the CELEX database (Baayen et al., 1995)), number of letters, number of neighbours, and summed and average frequency of neighbours. In addition, they were matched for relative frequency of neighbours (the number of neighbours that are higher or lower in frequency to the target) as this has been found to be a factor influencing neighbourhood effects (e.g., Grainger, O'Regan, Jacobs, & Segui, 1989). For example, the word *flight* has five neighbours but these all are lower in frequency than the target (e.g., *fright*), compared to the word *drown* which also has five neighbours but all are higher in frequency (e.g., *frown*). Therefore, we confirmed the sets did not differ on the number of higher frequency neighbours and lower frequency neighbours.

We hypothesized that if generalisation occurs, and is influenced by neighbourhood size, set C2 and C3 should be more likely to show improvement than C1.

Table 4

Overview of sets of stimuli (see Appendix E for more details)

Treated items (n=20 each set)		Control items (n=20 each set)	
T1	Words with no orthographic neighbours (e.g., *pebble*)	C1	Words with no orthographic neighbours, unrelated to T1 (e.g., *agent*)
T2	Words with many orthographic neighbours (e.g., *rocket*)	C2	Direct orthographic neighbours of T2 (e.g., *socket*)
		C3	Words with many orthographic neighbours in general, and unrelated to T1, T2 and C2 (e.g., *fever*)

Procedure

The treatment consisted of two phases of four weeks of treatment. Performance was tested twice before the treatment began, twice after the first treatment phase and before the second treatment phase, and twice after the second treatment phase. For JOD a third baseline was included. [4]

[4] Initially we included an extra set of untreated words for JOD. With those items combined JOD showed a significant improvement between Baseline 1 and Baseline 2 (McNemar's test exact, p = .047). We therefore decided to include a third baseline to better determine the trajectory of performance. For simplicity and congruence with GEC, these items are not reported, however, these untreated items also showed no significant improvement.

Due to personal circumstances GEC had a nine week break between Phase One and Two of treatment, and therefore two extra baseline assessments were conducted prior to the second phase of treatment. See Table 5 for a timeline of the treatment.

Table 5

Overview of timeline of treatment

Session	Week number GEC	Week number JOD	Task	Sets
Baseline 1.1	1	1	Writing to dictation all sets of treated and control items	T1, T2, C1, C2, C3.
Baseline 1.2	3	3	Writing to dictation all sets of treated and control items	T1, T2, C1, C2, C3.
Baseline 1.3	n/a	4	Writing to dictation all sets of treated and control items	T1, T2, C1, C2, C3.
Treatment Phase One	4-7	5-8	Copy and recall treatment	T1
Post-test 1.1	8	9	Writing to dictation all sets of treated and control items	T1, T2, C1, C2, C3.
Post-test 1.2	11	11	Writing to dictation all sets of treated and control items	T1, T2, C1, C2, C3.
Baseline 2.1	20	9 [a]	Writing to dictation all sets of treated and control items	T1, T2, C1, C2, C3.
Baseline 2.2	22	11 [a]	Writing to dictation all sets of treated and control items	T1, T2, C1, C2, C3.
Treatment Phase Two	23-26	12-15	Copy and recall treatment	T2
Post-test 2.1	27	16	Writing to dictation all sets of treated and control items	T1, T2, C1, C2, C3.
Post-test 2.2	29	18	Writing to dictation all sets of treated and control items	T1, T2, C1, C2, C3.

[a] For participant JOD the post-tests after treatment Phase One acted as baseline assessments for treatment Phase Two.

Assessment protocol. At each assessment point, all treated and untreated words were assessed on a writing to dictation task: The experimenter said the word aloud, and correct repetition was ensured before writing the word. No feedback was given regarding spelling accuracy. All 100 words were tested in randomised order. The order of administration was the same at each assessment.

Treatment. The treatment used a copy and recall method (based on work by e.g., Beeson, Rising, & Volk, 2003; Rapp & Kane, 2002; Schmalzl & Nickels, 2006). Both treatment phases consisted of eight sessions of around one hour, over the course of four weeks (two sessions a week).

In each session all 20 treated items were treated in a different random order. Items were individually printed on flashcards in the case opposite to the participant's preferred case in writing (i.e., printed in upper case for GEC and lower case for JOD). The flashcard was presented to the participant and they were asked to read the word aloud before (cross-case) copying the word, with the target in sight, four consecutive times. Any unnoticed errors that were made were corrected by the experimenter by pointing out an error was made and asking the participant to copy the word again until four correct responses were written.

The four consecutive copying attempts were followed immediately by a direct recall[5], where the target was removed from sight, in order to practice recall of the spelling. The direct recall in the copying task in the initial assessment showed that both participants performed poorly on this task (see Appendix C). Therefore, during treatment the number of letters to recall was gradually increased, aiming to increase delayed copying ability while maintaining accuracy. Hence, in the first session the participant was asked to recall the first two letters of the word, the next session the first three letters of the word, until by the final treatment sessions the entire word was to be recalled from memory. After the participant had recalled the required number of letters

[5] In the literature the conditions 'direct' and 'delayed' copying are used to describe the same task where the word is recalled immediately after exposure without introducing an extra delay. We choose to use the term 'direct' copying for this task (see Appendix C for the different copying conditions).

the word was uncovered and the remaining letters were copied in sight. Then the experimenter moved on to the first step with the next item.

Participants were allowed to correct their response in the recall tasks, but when they were unable to recall the required number of letters or an incorrect letter remained unnoticed, feedback was provided to indicate an error was made, and a second attempt was given: The word was uncovered, the participant looked at the word again before it was covered up again and a second attempt at recalling the word was given. If this second attempt also resulted in an error or no response, the word was uncovered and copied in sight, before moving on the next item. See Figure 2 for a flowchart of the treatment procedure.

When all 20 items had been copied and recalled, the items were shuffled and practiced in a second round where each item was copied in sight once, followed by a delayed recall from memory. In total, this resulted in at least seven spellings of each word per session. The treatment procedure was identical for Phase One and Phase Two of treatment.

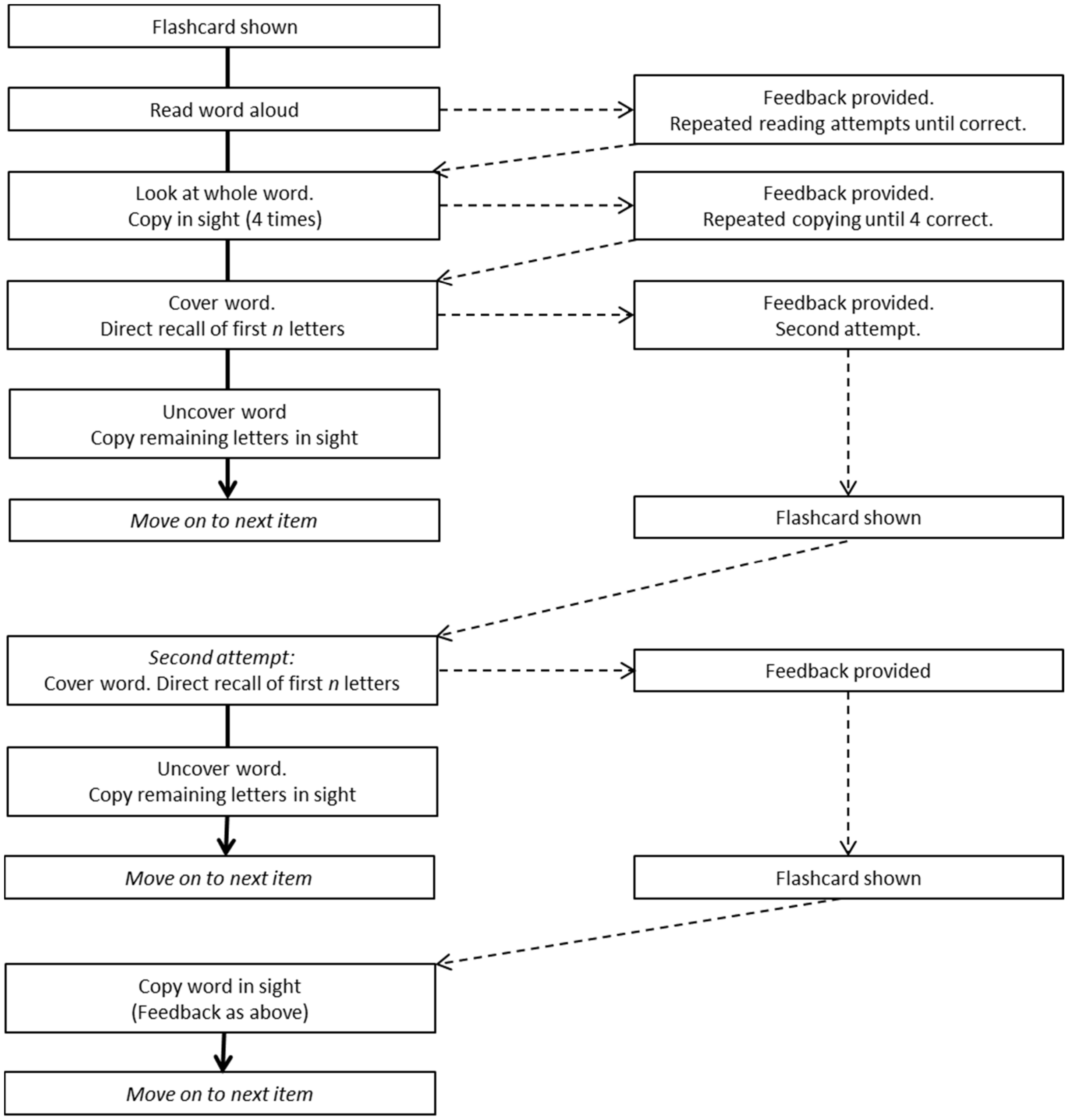


Figure 2. Flowchart of the treatment protocol. Solid arrows indicate next step following correct response, dotted arrows indicate next step following incorrect response.

Results

In order to show a positive effect of treatment we needed to demonstrate 1) that there was evidence of an overall trend for improvement over the course of the study; 2) that there was a greater rate of change during the treatment than the no-treatment phases. For both these analyses we used Weighted Statistics (Howard, Best, & Nickels, 2015), applying the requisite weightings for Trend (WEST-Trend) for Analysis 1 and for Rate Of Change (WEST-ROC) for Analysis 2. A significant treatment effect was present if both analyses showed a significant improvement for both item and letter accuracy.

We first measured accuracy as the number of words written correctly. However, word accuracy does not detect whether an incorrect response might be closer to correct after treatment. Therefore a letter accuracy score was calculated by counting the number and position of correct target letters present in the response. This measure follows a scoring system reported in Caramazza and Miceli (1990) and also adapted by, for example, Buchwald and Rapp (2009). Each target letter receives a score between 0 and 1. A correct letter in the correct relative position[6] receives a score of 1. If letters are in the incorrect position, points are subtracted. In the case of a transposition (e.g., *widonw* for *window*), half a point is subtracted from the transposed letter (in this example from the *n*). An addition is scored by subtracting half a point from each of the flanking letters. For example, the response *diarty* to the item *diary* receives a total score of four (five correct letters minus 0.5 for *r* and *y*) out of five target letters, which is a score of 0.8. Letters that are deleted or substituted receive a score of 0. For example, *aro* for the target *arena* receives a score of two letters correct (*a* and *r*) out of the five target letters, which results in a score of 0.4 for this response. (Note: *o* is assumed to be a substitution for *e*, and not an addition. Therefore, no scores are subtracted from the *r*.)

[6] Relative position indicates the position of a letter in relation to the other letters in the word, as opposed to absolute position in the target. For example, as the result of a deletion of the letter *r* from the target *dress* in the response *dess*, the *e* is no longer the third letter in the word, however it is still in the correct position in relation to the *d* and *s*. Therefore, the *e* is scored correct for position.

We discuss the results relating to our two research questions in turn (see Figures 3 and 4, and Tables 6 and 7 for a summary of the results of the statistical analyses).

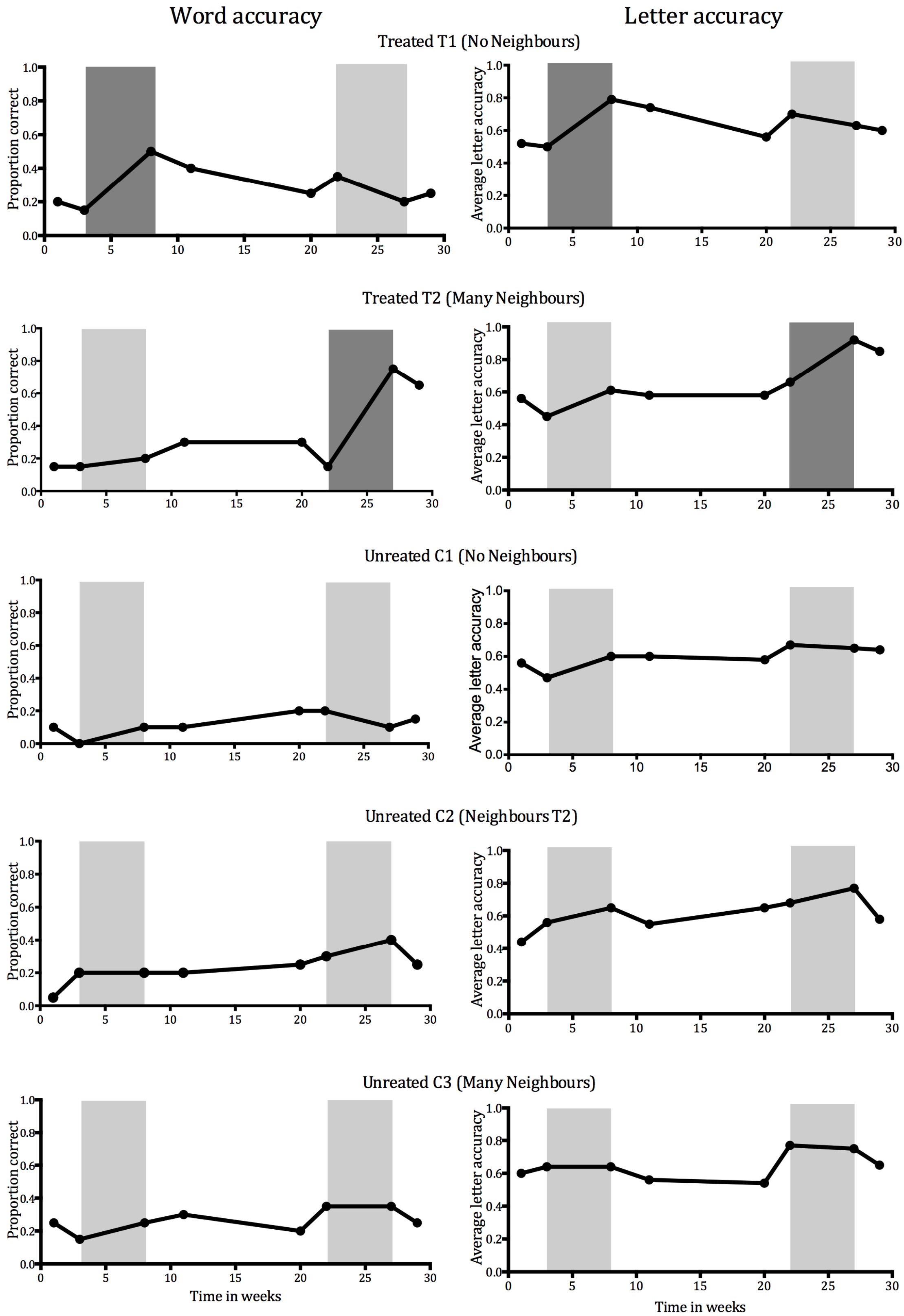


Figure 3. GEC: Word accuracy (left side panels) and letter accuracy (right side panels) across the study phases. Shaded areas indicate the periods when treatment occurred: Dark shading indicates when treatment was applied to those items (for T1 and T2), light shading indicates treatment applied to another set of items.

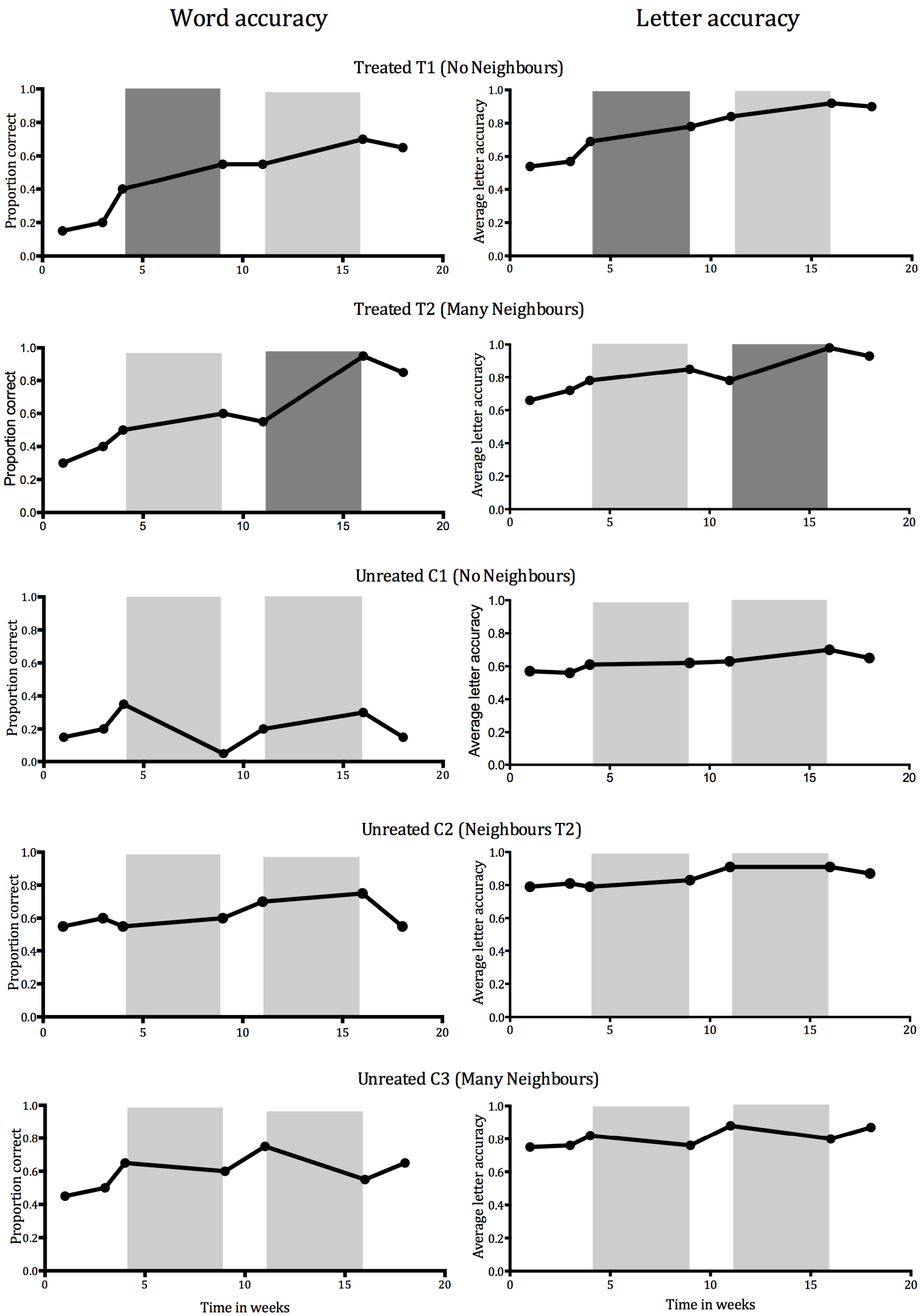


Figure 4. JOD: Word accuracy (left side panels) and letter accuracy (right side panels) across the study phases. Shaded areas indicate the periods when treatment occurred: Dark shading indicates when treatment was applied to those items (for T1 and T2), light shading indicates treatment applied to another set of items.

Table 6

Results of Weighted Statistics examining the Trend across the study (WEST-Trend) and comparison of Rate Of Change between treatment and no treatment periods (WEST-ROC): GEC

	Phase One								**Phase Two**							
	WEST-Trend				WEST-ROC: No Treatment vs. Treatment periods				WEST-Trend				WEST-ROC: No Treatment vs. Treatment periods			
	word accuracy		letter accuracy		word accuracy		letter accuracy		word accuracy		letter accuracy		word accuracy		letter accuracy	
	t	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
All words (n=100)	3.26	.001 **	2.87	.002 **	1.69	.047 *	3.01	.002**	1.70	.046 *	2.70	.004 **	1.12	.134	0.59	.278
T1 (n=20)	1.84	.040 *	3.30	.002 **	1.89	.037 *	2.73	.007**	-0.59	.719	0.25	.402	-0.97	.829	-1.16	.869
T2 (n=20)	1.49	.077	0.82	.211	0.0	.500	1.99	.030 *	3.42	.001**	3.95	<.001***	3.51	.001**	2.45	.012 *
All untreated (n=60)	2.32	.012 *	1.26	.107	0.80	.214	1.30	.099	0.00	.500	0.88	.190	0.00	.500	0.02	.491
C1 (n=20)	1.37	.093	1.23	.118	1.37	.093	1.85	.040 *	-0.82	.787	1.10	.142	-0.68	.747	-0.90	.810
C2 (n=20)	1.76	.048 *	1.86	.039 *	-0.65	.737	0.73	.236	0.26	.398	-0.40	.654	1.24	.115	1.65	.058
C3 (n=20)	0.89	.191	-0.39	.651	0.64	.264	0.16	.436	0.39	.351	1.40	.089	-0.13	.553	-0.91	.812

Note. The shaded cells indicate significant results.

* $p < .05$. ** $p < .01$. *** $p < .001$. All *p*-values are one-tailed

Table 7

Results of Weighted Statistics examining the Trend across the study (WEST-Trend) and comparison of Rate Of Change between treatment and no treatment periods (WEST-ROC): JOD

	Phase one								**Phase Two**							
	WEST-Trend				WEST-ROC: No Treatment vs. Treatment periods				WEST-Trend				WEST-ROC: No Treatment vs. Treatment periods			
	word accuracy		letter accuracy		word accuracy		letter accuracy		word accuracy		letter accuracy		word accuracy		letter accuracy	
	t	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
All words (n=100)	4.77	<.001***	5.94	<.001***	-1.43	.921	-0.11	.543	2.14	.017 **	3.59	<.001***	1.26	.106	1.11	.135
T1 (n=20)	3.93	<.001***	5.27	<.001***	0.38	.353	0.52	.303	0.96	.174	1.85	.040 *	1.51	.074	0.87	.198
T2 (n=20)	2.57	.009 **	2.49	.011 **	0.00	.500	0.08	.469	3.93	<.001***	3.83	.001 **	2.54	.010*	2.53	.010*
All untreated (n=60)	2.34	.011 *	3.34	.001 **	-2.09	.980	-0.44	.670	0.39	.349	1.84	.035 *	-0.36	.641	-0.86	.803
C1 (n=20)	-0.44	.667	1.35	.096	-3.04	.997	-0.05	.521	1.36	.095	0.93	.183	0.49	.315	1.04	.155
C2 (n=20)	1.06	.150	1.95	.033 *	0.28	.391	0.63	.268	-0.24	.594	0.97	.172	0.46	.327	-0.37	.641
C3 (n=20)	2.77	.006 **	2.76	.006 **	-0.93	.817	-0.88	.806	-0.12	.549	1.24	.114	-1.99	.969	-2.49	.989

Note. The shaded cells indicate significant results.

* $p < .05$. ** $p < .01$. *** $p < .001$. All *p*-values are one-tailed.

1) Did neighbourhood size play a role in the effect of treatment?

To investigate the role of neighbourhood size we first examined whether treatment was effective in each of Phase One (treated items with no orthographic neighbours) and Phase Two (treated items with many orthographic neighbours) and compared the size of any treatment effect across these two phases. If neighbourhood size had an effect of treatment we would predict a larger improvement after Phase Two compared to Phase One.

GEC. Table 6 summarises the statistical analysis for GEC. For the items with no neighbours treated in the first phase of treatment (T1) GEC showed a significant trend for improvement and also showed significantly greater improvement during the treatment phase compared to the baseline period for both word accuracy and letter accuracy analyses.

After the second phase of treatment GEC also showed a significant trend for improvement for the treated items with many neighbours (T2) and significantly greater improvement on this set of words when they were treated compared to the untreated period for both item and letter accuracy analyses.

The rate of change during therapy of the two treatment sets was comparable (two sample t-test, word accuracy: $t(38) = 0.98$, $p = .166$; letter accuracy: $t(38) = 0.44$, $p = .332$, one-tailed). Hence, GEC showed a significant treatment effect for treated items when they were treated, but the neighbourhood size of the treated items did not influence the size of the treatment effect.

JOD. Table 7 summarises the statistical analyses for JOD. JOD showed a significant trend for improvement after the first phase of treatment (both item and letter accuracy) however improvement over the treated period was not significantly greater compared to the untreated period (for word accuracy or for letter accuracy). After the second phase of treatment with stimuli that had many neighbours JOD did show both a significant trend for improvement and significantly greater rate of change during treatment compared to no treatment periods, for both item and letter accuracy. However, the difference between the treatment effects for the two sets failed to reach significance (two sample t test, word accuracy: $t(38) = 1.48$, $p = .074$; letter accuracy: $t(38) = 1.39$, $p = .087$, one-tailed). We will return to this result in the Discussion.

2) Did generalisation occur? Did neighbourhood play a role? To test for generalisation after treatment we first looked at all 60 untreated items together, and then analysed the different control sets separately. Word accuracy as well as letter accuracy is reported. If generalisation occurred we would expect to see an improvement of untreated words, and if neighbourhood plays a role we would predict neighbours of treated items from Phase Two (C2), or the set of words with many neighbours (C3) to show the largest improvement.

GEC. After Phase One of treatment (treated words without neighbours), GEC showed a significant trend for improvement for all untreated items only for word accuracy, but there was no difference in the rate of change during therapy compared to the untreated period. After Phase Two (treated words with many neighbours) there was no significant improvement for either word accuracy or letter accuracy. Hence there was no generalisation overall for untrained words after either Phase One or Phase Two.

When the individual sets of untreated items were analysed separately after Phase One, no set of control words showed both a trend for improvement and a significant rate of change (as required to indicate an effect of treatment on performance). After Phase Two no control set showed significant improvement. Hence, for GEC treatment effects did not seem to generalise to untreated words.

JOD. JOD showed a significant trend for improvement in the spelling of all untreated words combined after Phase One of treatment, both for item and letter accuracy, but the rate of change was not significantly greater during treatment compared to no-treatment periods. After Phase Two, JOD only showed a significant trend for improvement based on letter accuracy, and no significant difference in rate of change between treated and untreated periods. Hence, there was no generalisation for untrained words after either Phase One or Two.

When the untreated sets were analysed separately there were no sets that showed evidence of treatment related improvement (i.e., no set showed both significant trend and rate of change). Hence, JOD also showed no generalisation after treatment, whether measured by word accuracy or by, the potentially more sensitive, letter accuracy.

Summary of results.

GEC. Practicing spelling of a set of words appeared to be a successful method of treatment for GEC. Both the first and second phase of treatment resulted in significant treatment effects. The size of the effect was similar for the two phases, indicating that the neighbourhood size of treated items did not influence the effect of treatment. After treatment, accuracy declined in both phases.

JOD. JOD seemed to show a different pattern to GEC. After the first phase of treatment he did not show a significant improvement: treatment did not provide additional benefit over and above the improvements from testing during the pre-treatment baseline. After the second phase of treatment that focused on words with many neighbours, JOD did show a significant treatment effect. While it seemed that the neighbourhood size of treated items had a significant effect on treatment, there was no significant difference between the two treated sets in the amount of benefit from treatment.

In order to gauge the effect that treatment had over and above the improvement JOD showed in periods without treatment, we performed WEST-Trend and WEST-ROC analyses including all assessment points across the study. This analysis showed that improvement in the treated items with many neighbours (T2) was only close to significantly greater when the items were treated, compared to untreated periods and only when measured as word accuracy (word accuracy: $t(19) = 1.49$, $p = .076$; letter accuracy: $t(19) = 1.08$, $p = .147$, one-tailed). Furthermore, in the design we used, potential effects of neighbourhood and effects of order of treatment were confounded. Consequently, the role of neighbours in the effectiveness of treatment for JOD remained unclear. Because of the potential importance of the result, we therefore carried out a follow up experiment in an attempt to clarify the pattern. Treating these two types of words at the same time would allow us to investigate the role of neighbourhood size in more detail. If neighbourhood size of treated items played a role we would expect a larger treatment effect for treated words with many neighbours. If treatment was not influenced by neighbourhood size we would predict no difference in the size of the treatment effect between the two types of targets.

JOD: Treatment Phase Three

Method. The stimuli were those used in treatment Phases 1 and 2, with treatment of two of the previously untreated, control sets. Set C1, which previously was the control set without neighbours, was now the treated set without neighbours. Set C3, previously the control set of words with many neighbours, was now the treated set with many neighbours. The two sets were treated simultaneously. All other sets of words acted as experimental control. A similar copy and recall treatment was administered, however in order to be able to treat all 40 words in one session, all words were practiced only once each session, rather than twice as in Phases 1 and 2. For the same reason, words were copied in sight only twice (rather than four times) before the delayed recall.

Treatment consisted of two baseline assessments (two weeks apart), six treatment sessions, followed by two post-tests (two weeks apart). During the assessment all words were administered in a writing to dictation task in a random order.

The focus of this study was to examine whether neighbourhood size plays a role in the size of treatment effects for JOD, if so the treatment effect for the treated set with many neighbours (C3) would be larger compared to that of the set without neighbours (C1).

Results. Results were analysed as for Phases 1 and 2 (see Figure 5 and Table 8 for a summary of the statistical analyses). Both sets of treated items showed significant improvement (Trend and Rate Of Change), for both word accuracy and letter accuracy. However, the extent of improvement did not differ significantly between treated words with neighbours and the treated words without neighbours (two sample *t*-test comparing the rate of change, word accuracy: $t(38) = 0.94$, $p = .176$; letter accuracy: $t(38) = 0.48$, $p = .316$, one-tailed). Hence, for JOD the neighbourhood size of treated items did not influence the size of the treatment effect.

JOD showed a significant trend for improvement over Phase Three for all untreated items (n=60), for both item and letter accuracy. However, the improvement only just reached significance for greater improvement during treated compared to untreated periods for word accuracy ($p = .047$), and did not for letter accuracy. We therefore believe it would be unwise to take these results as evidence for generalisation to untreated items.

In addition, the untreated items showed significantly less improvement than the treated items (Rate Of Change: treated (n=40) vs. untreated (n=60): two sample *t*-test, word accuracy: $t(98) = 2.40$, $p = .009$; letter accuracy: $t(64.60) = 3.78$, $p < .001$, one-tailed). This result confirms that treatment was effective for treated items.

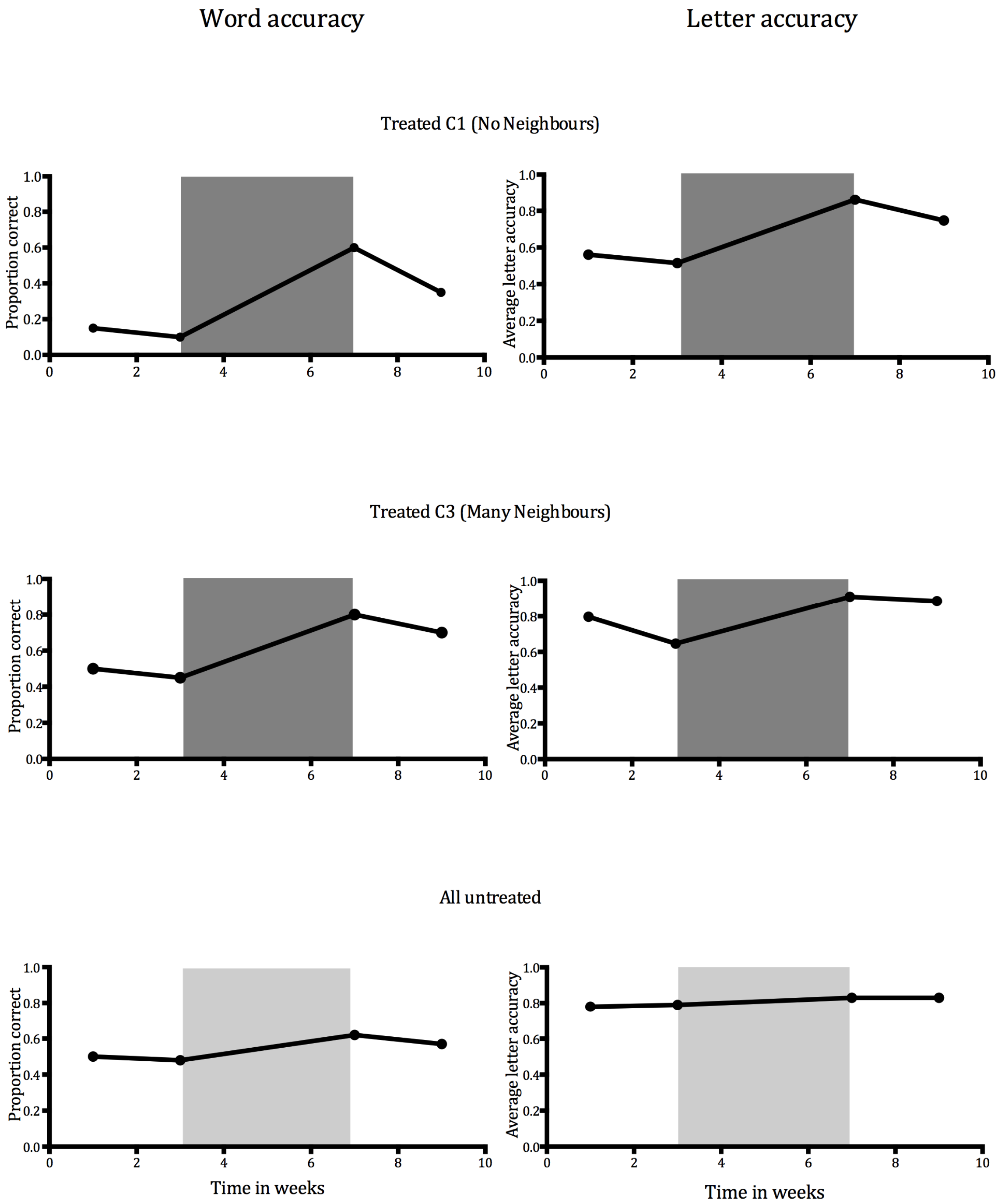


Figure 5. JOD: Word accuracy (left side panels) and letter accuracy (right side panels) across Phase 3. Shaded areas indicate the periods when treatment occurred: Dark shading indicates when treatment was applied to those items (for C1 and C3), light shading indicates treatment applied to another set of items.

Table 8

Results of Weighted Statistics examining the Trend across the study (WEST-Trend) and comparison of Rate Of Change between treatment and no treatment periods (WEST-ROC): JOD Phase Three.

	WEST-Trend				WEST-ROC: No Treatment vs. Treatment periods			
	word accuracy		letter accuracy		word accuracy		letter accuracy	
	t	*p*	*t*	*p*	*t*	*p*	*t*	*p*
All words (n=100)	3.79	<.001***	4.33	<.001***	4.07	<.001***	4.13	<.001***
All treated (n=40)	3.51	.001**	4.34	<.001***	4.51	<.001***	4.97	<.001***
Treated no N (C1; n=20)	2.57	.009**	3.51	.001**	4.47	<.001***	4.50	<.001***
Treated many N (C3; n=20)	2.33	.015*	2.59	.009**	2.24	.018*	2.78	.006**
All untreated (n=60)	1.88	.033*.	1.87	.033*	1.70	.047*	1.04	.152

Note. The shaded cells indicate significant results.
* $p < .05$. ** $p < .01$. *** $p < .001$. All *p*-values are one-tailed.

Discussion

This study with two individuals with graphemic output buffer impairment investigated the role of neighbourhood size in the effects of treatment on spelling to dictation and in the generalisation of these effects. For one participant, GEC, treated items improved after treatment, and there was no influence of the neighbourhood size of the treated items on the extent of improvement. Effects of treatment were item specific: there was no improvement in the spelling of untreated words.

In the first treatment study, the second participant, JOD, showed improved spelling of words with many neighbours, and no significant improvement for those with no neighbours. However, there was no significant difference between the two sets in the extent of improvement. Because of this result, and that the effects of neighbourhood were confounded with effects of order, and that there was a marked improvement in the baseline period before Phase One (treatment for items with no neighbours), we carried out a further treatment phase, where words without neighbours and with many neighbours were treated simultaneously. After this second treatment study, all treated items improved significantly, and the size of the treatment was equal for items with and without neighbours.

In sum, there was no clear evidence that neighbourhood size of treated items influenced the effect of treatment for either participant. We also found no evidence for generalisation of the effects of treatment to untreated stimuli in any phase. We will now discuss these results as they relate to our research questions.

The Effect of Treatment

The assessment results showed that both participants performed very poorly on spelling tasks. A copy and recall protocol was chosen to improve word spellings, as the repeated exposure to and copy of target spellings had previously been found to be effective for a number of individuals with acquired dysgraphia in general (e.g., Beeson, 1999; Beeson, Hirsch, & Rewega, 2002; Schmalzl & Nickels, 2006), and for graphemic output buffer impairment specifically (e.g., Rapp & Kane, 2002; Rapp, 2005; Raymer et al., 2003). For both GEC and JOD, item specific improvements in spelling were observed due to the treatment.

After both phases of treatment, for both participants, accuracy on the set of treated items declined between the initial post-test and the follow-up assessment two weeks later. It therefore seems likely that, for these participants, regular practice is required to maintain high levels of accuracy. Therefore, this method of practice may be most useful for a set of functional words that are rehearsed regularly.

The question arises what the mechanism behind this treatment effect is. In their delayed copying treatment, Rapp and Kane (2002) treated a participant with a graphemic output buffer impairment (RSB), and an individual with a deficit to the orthographic lexicon (MMD): Improvement of treated items was found for both individuals, but generalisation only occurred for RSB. Copy and recall treatment has been argued to benefit the representations in the orthographic output lexicon, which explained the improvement for participant MMD (Rapp & Kane, 2002). To explain the generalisation of improvement for RSB, the individual with buffer impairment, Rapp and Kane argued that a strengthening of target representations in the orthographic output lexicon combined with some general benefit to the graphemic output buffer could explain RSB's improvement of treated items and generalisation to untreated items (Rapp & Kane, 2002). Considering the lack of generalisation in the current study, the strengthening of lexical representations leading to greater resistance to graphemic output buffer impairment seems to account best for GEC and JOD's results.

The question remains, however, why JOD did not improve on the treated items in Phase One, especially considering that he showed improvement over the baseline testing period. Given that he was more than 15 years post onset, it seems unlikely that this baseline improvement reflected spontaneous recovery: it is more likely to reflect a benefit for subsequent spelling of a previous attempt at spelling - a practice effect. It has been argued that such 'practice' effects might be the result of a similar mechanism to that which occurs in treatment: priming of the orthographic form (Sage & Ellis, 2006; see Nickels (2002) for a similar explanation in the context of spoken word retrieval). In their study, Rapp and Kane (2002) also included a set of words that were administered for spelling before and throughout treatment, however no feedback was given, and the correct spelling of these words was never visually presented. After treatment, participant MMD (who had an orthographic lexicon impairment) showed not only improvement on the treated items, but also a significant effect of repeated attempts at spelling these untreated repeated words, whereas no improvement was reported for a different set of untreated words that were only administered once prior to treatment and once at the end of treatment.

However, considering that JOD showed improved spelling for untreated words with repeated testing, one would expect that an intense period of treatment, i.e., writing all words at least seven times during one session, twice a week for four weeks, should result in additional benefits from priming. However, this did not occur: there was no improvement in spelling of the treated words, after Phase One. Alternatively, there might be different mechanisms behind the improvement resulting from repeated practice and from treatment, although even if this were the case one might expect the effects of repeated practice still to be evident especially as it was not the case that performance was influenced by ceiling effects (40% accuracy for T1 items immediately prior to treatment).

Why might JOD have benefited more (if not significantly so) from subsequent phases of treatment? Although it is possible that he changed his strategy for learning from the second phase of treatment onwards, we have no evidence to support this. It is also possible that he was more familiar with the type of treatment administered, or reached some 'threshold' to enable him to benefit (although once again this seems inconsistent with the baseline improvement). It did appear to be the case that throughout treatment Phase One he made a few more errors in the first round of recall compared to the second and third phases of treatment.

The Effect of Neighbourhood Size on Treatment

The design of this study also enabled us to investigate the effect of the neighbourhood size of treated items on the results of treatment. We hypothesised that as a result of a mechanism of feedback between the lexicon and graphemic output buffer, co-activated letters from neighbours could provide additional support for the correct spelling of a target in treatment. Consequently, treated words with many neighbours might respond better to treatment compared to words that have no neighbours. The results of our treatment did not confirm this hypothesis: effects of treatment were equal for words with and without neighbours, for both participants. We will discuss three possible accounts for these results.

One potentially critical factor in our design relates to the definition of neighbourhood size. We chose to use the relatively standard 'Coltheart's N' (neighbours are words with one letter substitution; Coltheart et al., 1977), as this has been widely used in previous studies investigating neighbourhood size (e.g., Roux & Bonin, 2009; Sage & Ellis, 2006). However, other studies have investigated the nature of neighbourhood in more detail (e.g., Goldrick, Folk, & Rapp, 2010), suggesting that in addition to orthographic overlap, other variables like frequency, grammatical category, target length, can contribute to the activation of non-target words in spoken and written production (Goldrick et al., 2010). Furthermore, broader concepts of 'orthographic similarity' have been suggested to play a role in visual word *recognition* (e.g., Yarkoni, Balota, & Yap, 2008). Yarkoni et al. put forward another measure of orthographic similarity, orthographic Levenshtein distance 20, which is a more graded measured compared to the binary definition of neighbours using 'Coltheart's N'. It is possible that our definition of neighbours was not a sensitive enough measure of orthographic similarity, and therefore an effect of neighbourhood size.. However, when we calculated the orthographic Levenshtein distance (OLD) for the two sets of treated words from Phase One and Phase Two, (Balota, et al., 2007) the two sets also differed significantly (mean OLD T1 = 2.07, T2 = 1.60; two sample t test: $t(28.37) = 8.77$, $p < .001$).

Alternatively, and perhaps most plausibly, the individual's underlying graphemic output buffer impairment could have decreased the amount of feedback that is typically available in the unimpaired spelling system. During treatment any possible co-activation of neighbours of the treated items may not have been sufficient to support the target spelling, and consequently there was no additional benefit for treated words with neighbours compared to words without neighbours.

Normal amounts of feedback in the spelling system predict facilitation from having just spelled a word's neighbour (also see Sage and Ellis, 2006). We tested this prediction post-hoc using the manipulation of the order of the word and its neighbour across the writing to dictation task at pre-test: Some words from set T2 (treated words with many neighbours), for example *grade*, were administered in the first half of the assessment, while their neighbours from control set C2, *grave,* were administered in the second half of the list. Words that were not preceded by a neighbour, might be predicted to be spelled less accurately than words written after having correctly written a neighbour. This was not found to be the case for either participant, as both participants were less accurate on the subset of words that had been preceded by a neighbour compared to the initial administration. At first baseline, GEC only spelled four words correctly, all in the first half of the assessment, and therefore none of these had been preceded by a neighbour. JOD wrote ten words correctly in the first half (i.e., none preceded by a neighbour), and seven words correctly after being preceded by a neighbour. Hence, there was no facilitation from neighbours, which is consistent with decreased feedback in the spelling system for GEC and JOD.

A third potential explanation is that neighbours do provide feedback to support target spellings, however this may be negligible compared to the effect of treatment. It may be the case that both a word without neighbours and a word with many neighbours benefit equally from specific training of the target spelling. It is this explicit training that results in improvement of the target word, and any additional support from neighbours is not sufficient to make a difference in the size of the treatment effect for words with and without neighbours.

Generalisation of Treatment Effects

This study also aimed to investigate generalisation of treatment effects, and the role of orthographic neighbourhoods and neighbourhood size might play. However, treatment did not generalise to untreated words for GEC or JOD, which contrasts with several other studies (Harris et al., 2012; Raymer et al., 2003; Panton & Marshall, 2008; Rapp and Kane, 2002; Sage & Ellis, 2006). Rapp and Kane (2002; Rapp, 2005) reported generalisation to untreated words for an individual with graphemic output buffer impairment (RSB). Panton and Marshall (2008) also reported generalisation to a set of untreated words and errors occurring at a later position in the word compared to baseline. Other studies have found generalisation to untreated words that share orthography with the target words: Sage and Ellis (2006) showed generalisation to untreated neighbours, Harris et al. (2012) found generalisation to neighbours with shared middle letters, and Raymer et al. (2003) reported some generalisation to untreated words, mainly words that shared orthography with the target.

What factors might have caused the difference between these studies and our results?

1. Severity of impairment. It is possible that GEC and JOD failed to show generalisation due to the severity of their impairment. Rapp and Kane (2002) suggested that their delayed copying treatment could have strengthened processes that are beneficial to overcome buffer damage (e.g., scanning speed). As the buffer is used in the spelling of all words, generalisation to untreated words could then occur. It is possible that perhaps severe buffer impairment is less amenable to this strengthening.

Indeed, while it is hard to compare across studies due to the different stimuli that were used, it does appear that GEC and JOD are more severely impaired than patients in other treatment studies. For example, Ray (Panton & Marshall, 2008) was accurate in spelling 59% of words, BH (Sage & Ellis, 2006) 52%, RSB (Rapp & Kane, 2002) 47%, JF (Harris et al., 2012) 36%, and NM (Raymer et al., 2003) 35% of words, compared to 17% and 25% for GEC and JOD, respectively. As discussed above, perhaps in severe buffer impairment, the extent of feedback to neighbours in the lexicon is reduced. Consequently untreated neighbours do not inherit the effects of treatment.

Furthermore, it may be the case that GEC and JOD had additional impairments that may have reduced the benefit from treatment. First, JOD was unable to write in lower case. This impairment of allographic conversion may have limited treatment effects on higher level processes. Second, both GEC and JOD were at floor when spelling non-words, which may indicate an additional impairment to sub-lexical processing. Nevertheless, the difference between matched sets of words and non-words was not significant for either participant. Furthermore, the error pattern for both words and non-words showed characteristics of buffer impairment, and it is therefore unclear if non-word spelling abilities affected the treatment results. However, post-treatment data investigating the effect of treatment on JOD's conversion deficit and on the possible sub-lexical impairment in both GEC and JOD may have provided additional insight regarding the mechanisms of treatment.

2. Type of treatment. The copy and recall treatment used here was similar to treatments reported by Rapp and Kane (2002) and Raymer et al. (2003) resulting in generalisation. However Sage and Ellis reported a different treatment that was based on the principle of 'errorless learning'. Throughout therapy BH was prevented from producing errors, allowing correct representations to be formed (Sage & Ellis, 2006). It was argued that these tasks were more likely to have boosted lexical representations rather than facilitated spelling at the level of the graphemic output buffer (Sage & Ellis, 2006). It is possible that using errorless learning tasks could have boosted treatment effects and ensured more accurate feedback between lexicon and buffer (Sage & Ellis, 2006), perhaps resulting in generalisation of treatment effects for GEC and JOD.

3. Type of untreated neighbours. Another possibility for why we did not find generalisation to neighbours could be the nature of the untreated neighbours we used and specifically the position of the letter change. As reported above, Raymer et al. (2003) reported improvement for words that shared letters (initial and final position) with the treated items, and the largest improvement was shown for words sharing initial position (*rac*ket – *rac*oon). In contrast, Harris et al. (2012) found that generalisation was more likely between neighbours that shared middle positions (*pouch* and *couch*) compared to neighbours with changed medial letters (*couch* and *coach*). This seems to suggest that the position of the shared letters is crucial, with generalisation being more likely for neighbours which share middle letters. In our study the position of the letter change in the neighbour pairs varied. For example, of the neighbour pairs in sets T2 – C2, the majority of pairs (16/20) had changes affecting the first or second letter (*match*-

catch, spell-shell). It was only a minority of items (4/ 20) that shared initial letters, with the substitution in the fourth or fifth position (*grade-grave*).

However, it seems unlikely that this could be the sole reason for the lack of generalisation in our study: Sage and Ellis (2006) found generalisation with neighbours selected across different letter positions (if the word had more than one neighbour). It therefore seems unlikely that position of overlap is the critical variable affecting generalisation to neighbours.

4. Measurement of improvement. Could it be possible that our measurement of improvement was not sensitive enough to show generalisation? Panton and Marshall (2008) argued that their treatment resulted in improved buffer functioning in an individual with graphemic output buffer dysgraphia. Before treatment, errors appeared on average in letter position 2.8. After treatment, errors shifted to position 3.5. It was argued that after treatment a larger window in the graphemic output buffer was available to keep letter information active (Panton & Marshall, 2008). While we did not examine position of error, we did examine the number of letters that were correct in the response: If treatment had resulted in a larger window in the buffer, this would have been reflected in higher letter accuracy scores, which were not found for untreated items for either GEC or JOD.

Conclusion

This study was set out to investigate the mechanism of feedback in the spelling system by testing two predictions of this mechanism. We investigated 1) if the orthographic neighbourhood size of treated items has an effect on treatment, and 2) whether generalisation occurs and whether neighbourhood plays a role. We found that a copy and delayed recall treatment resulted in item specific effects only, and the neighbourhood size of treated items did not influence the size of the treatment effect. No strong evidence for generalisation was reported, and therefore the influence of neighbourhood size on generalisation could not be investigated.

The results are hypothesised to be the result of GEC and JOD's underlying impairment affecting the mechanism of feedback between the lexicon and the buffer. First, as a consequence of decreased feedback, neighbours of treated items could not provide additional support that could have resulted in a larger treatment effect for these items compared to treated words without neighbours. Alternatively, the effect of feedback from neighbours of treated items could be insignificant compared to the effects of treatment. Hence, no difference was found between treated words without neighbours and with many neighbours.

Second, generalisation to neighbours of treated items has been argued to be evidence for interactivity in processing. GEC and JOD did not show generalisation to untreated neighbours. Once again, this could be the consequence of an impairment affecting the mechanism of feedback between the lexicon and the buffer, with the result that no sufficient activation of orthographic neighbours is provided: the target letters did not provide sufficient activation back up to orthographic neighbours in the lexicon, any activation was too weak to provide extra support down to the level of the buffer to result in improved spelling.

In sum, this study has provided further insight into the mechanisms of spelling and hypothesised that reduced feedback from buffer to lexicon can result in diminished effects of orthographic neighbours in spelling. These hypotheses can be tested with individuals with different levels of severity of impairment to further understand the role of neighbourhood size of treated items on the effects of treatment and generalisation.

Acknowledgements

We would like to thank David Howard for assisting with the analyses.

References

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4*(4), 439-461.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX Lexical Database. Release 2* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

Beeson, P.M. (1999). Treating acquired writing impairment: Strengthening graphemic representations. *Aphasiology*, *13*, 767–785.

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

Beeson, P. M., Rising, K., & Volk, J. (2003). Writing treatment for severe aphasia: who benefits? *Journal of Speech, Language, and Hearing Research, 46*, 1038-1060.

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

Buchwald, A., & Falconer, C. (2014). Cascading activation from lexical processing to letter-level processing in written word production. *Cognitive Neuropsychology, 31*(7-8), DOI: 10.1080/02643294.2014.951609

Buchwald, A., & Rapp, B. (2009). Distinctions between orthographic long-term memory and working memory. *Cognitive Neuropsychology, 26*(8), 724-751.

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: Evidence from a case of acquired dysgraphia. *Cognition, 26,* 59-85.

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

Colenbrander, D., Kohnen, S., & Nickels, L. (2011). *Diagnostic Reading Test for Nonwords (DiRT).* Retrieved from www.motif.org.au.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). New York: Academic Press.

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology, 79*, 1-12.

Davis, C. (2005). N-Watch: A program for deriving neighbourhood size and other psycholinguistic statistics. *Behavior Research Methods, 37*(1), 65-70.

Falconer, C., & Buchwald, A. (2013). Do activated letters influence lexical selection in written word production? *Aphasiology, 27*(7), 849-866.

Fieder, N., Nickels, L., & Biedermann, B. (2015) How 'some garlic' turns into 'a garlic' or 'capsicum': A follow-up study on mass and count processing in aphasia. Manuscript submitted for publication.

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.

Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop's neighbourhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language, 62*, 113-134.

Goodman, R. A., & Caramaza, A. (1985). *The Johns Hopkins University Dysgraphia Battery.* Baltimore, MD: Johns Hopkins University.

Goodman, R. A., & Caramazza, A. (1986). Aspects of the spelling process: Evidence from a case of acquired dysgraphia. *Language and Cognitive Processes, 1*(4), 263-296.

Grainger, J., O'Regan, J. K., Jacobs, A. M., Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics, 45*, 189-195.

Grainger, J. (1990) Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29*, 228-244.

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

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

Howard, D., & Patterson, K.E. (1992). *Pyramids and Palm Trees*. Thames Valley Test Company, Bury, St. Edmunds, UK.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic Assessments of Language Processing in Aphasia (PALPA)*. Hove: Lawrence Erlbaum Associates.

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.

Kohnen, S., Nickels, L., & Castles, A. (2009). Assessing spelling skills and strategies: A critique of available resources. *Australian Journal of Learning Difficulties, 14*(1), 113-150. DOI:10.1080/19404150902783450.

Krajenbrink, T., Nickels, L., & Kohnen, S. (2015). Generalisation after treatment of acquired spelling impairments: A review. *Neuropsychological Rehabilitation, 25*(4), 503-554. DOI: 10.1080/09602011.2014.983135.

Krajenbrink, T., Nickels, L., & Kohnen, S. (2015). *Graphemic output buffer impairment with fragment errors: a case of reduced activation in the buffer.* Manuscript submitted for publication.

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

Larsen, L., Kohnen, S., Nickels, L., & McArthur, G. (In Press). The Letter-Sound Test (LeST): A reliable and valid comprehensive measure of grapheme-phoneme knowledge. *Australian Journal of Learning Difficulties*. DOI:10.1080/19404158.2015.1037323

Laxon, V. J., Coltheart, V., & Keating, C. (1988). Children find friendly words friendly too: words with many orthographic neighbours are easier to read and spell. *British Journal of Educational Psychology, 58*, 103-119.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

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.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review 76*(2), 165-178.

Nickels, L. A. (1995). Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word production. *Language and Cognitive Processes*, 10, 13-45.

Nickels, L. (2002). Improving word finding: Practice makes (closer to) perfect? *Aphasiology, 16*(10-11), 1047-1060.

Nickels, L., & Cole-Virtue, J. (2004). Reading tasks from PALPA: How do controls perform on visual lexical decision, homophony, rhyme and synonym judgments? *Aphasiology, 18*(2), 103-126.

Panton, A., & Marshall, J. (2008). Improving spelling and everyday writing after a CVA: A single-case therapy study. *Aphasiology, 22*(2), 164-183.

Rapp, B. (2002). Uncovering the cognitive architecture of spelling. In A. E. Hillis (Ed.), *The handbook of adult language disorders: Integrating cognitive neuropsychology, neurology and rehabilitation* (pp. 47-69). New York: Psychology Press.

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. doi:10.1080/02687030544000209

Rapp, B., & Caramazza, A. (1997). From graphemes to abstract letter shapes: Levels of representation in written spelling. *Journal of Experimental Psychology: Human Perception and Performance, 23*(4), 1130-1152.

Rapp, B., Epstein, C., & Tainturier, M. J. (2002). The integration of information across lexical and sublexical processes in spelling. *Cognitive Neuropsychology, 19*(1), 1-29.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107* (3), 460-499.

Rapp, B., & Kane, A. (2002). Remediation of deficits affecting different components of the spelling process. *Aphasiology, 16*(4-6), 439-454. DOI:10.1080/02687030244000301.

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

Roux, S., & Bonin, P. (2009). Neighborhood effects in spelling in adults. *Psychonomic Bulletin & Review, 16*(2), 369-373.

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

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

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

Shallice, T. (1981). Phonological agraphia and the lexical route in writing. *Brain, 104*, 413-429.

Swinburn, K., Porter, G., & Howard, D. (2004). *The Comprehensive Aphasia Test*, Hove, UK: Psychology Press.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15* (5), 971-979.

Appendix A

Results background assessment GEC and JOD

Task	Maximum score	Cut-off score %	GEC %	JOD %
Comprehension				
Spoken word-picture matching (PALPA 47) [a]	40	95	98	100
Written word picture matching (PALPA 48) [a]	40	95	98	100
Semantic processing				
Auditory synonym judgement (PALPA 49) [a]	60	n/a	80	100
- High imageability	30	n/a	90	100
- Low imageability	30	n/a	70	100
Written synonym judgement (PALPA 50) [a]	60	87	82 *	93
- High imageability	30	91	90 *	93
- Low imageability	30	82	73 *	93
Conceptual semantics				
PPT [b] (3 pictures)	52	94	85 *	98
Production				
Word fluency (CAT) [c]	n/a	13 items	0 *	12 items *
Spoken picture description task (CAT)	n/a	score: 33	score: 19 *	score: 31.5 *
Written picture description task (CAT)	n/a	score: 19	score: 0 *	score: 4 *
Auditory discrimination				
Non-word minimal pairs (PALPA 1 – subset) [a]	36	n/a	97	86
Lexical processing				
Auditory lexical decision (PALPA 27 – subset) [a]	60	n/a	95	92
- Regular	15		100	87
- Exception	15		87	93
- Non-homophonic non-words	30		97	93
Visual lexical decision (PALPA 27) [a]	60		93	88
- Regular	15	94	87 *	100
- Exception	15	94	87 *	87 *
- Pseudo-homophones	15	87	100	80 *
- Non-homophonic non-words	15	95	100	87 *
Reading				
Cross case matching letters	29		100	100
Letter naming (PALPA 22) [a]	52		54	98
- Upper case	26	100	53 *	100
- Lower case	26	100	53 *	96 *
Letter sounding (LeST) [d]	51	n/a	27	19
Reading words (CAT)	48 [2]	score: 45	score: 43 *	score: 42 *

Reading non-words (DiRT) [e]	105	n/a	28	10

[1]Cut-off scores are scores > 2 standard deviations below the mean score of healthy controls. Cut-off scores from the CAT represent the score that at least 95% of normal subjects exceed. Cut-off scores are taken from the tests manuals or from Nickels & Cole-Virtue (2004) norms. * score indicates an impairment (at or below cut-off).

[2] This CAT subtest allows for a score of 2 (immediate correct response) or 1 (>5 seconds delayed correct response or a self-correction). Therefore '*n*' reflects maximum score here, which is two times the actual number of items.

[a] Psycholinguistic Assessments of Language Processing in Aphasia: PALPA; Kay, et al, 1992.

[b] Pyramids and Palm Trees Test, Howard & Patterson, 1992

[c] Comprehensive Aphasia Test: CAT; Swinburn, Porter, & Howard, 2004

[d] Letter Sound Test (LeST): Larsen, Kohnen, Nickels, & McArthur (In Press)

[e] Diagnostic Reading Test – Non-words: DiRT; Colenbrander, Kohnen, & Nickels (www.motif.org.au).

Appendix B

Picture description task (CAT, Swinburn et al., 2004)

1. GEC

a) Spoken output

I see a boy and he is (…) going to have a motor car.
And I see coffee (…) and an album, and socks on the table (…) and a coffee table (...) and a tie.
(experimenter: "So, what is happening, in the picture?")
Asleep. *(GEC laughs)* And (...) it is (…) it is a cat, and he is (...) counting the fish and gripping the fish and (…) (*GEC points to falling book in the picture)* one down and scones the fellow. (*GEC laughs)* And I don't know what the other one is (*GEC points to the radio in the picture*). It's (…) rrr (…) speaker, and turntable and recorder, but (…)

b) Written output

Kid. Car/Truck.
Coffee. Track (*target: table)*. Cou (*target: cup)* of tea. Pho… (*target: photograph)*. C.
Books. Haed (*target: head).*
Cat wa (*target: was).* Fishing on the water, C …
Plate.

2. JOD

a) Spoken output

There's a man sitting on the couch having a sleep. His feet are on the table, cup of, cup of coffee. A book. And /er/ a child is (…) crying for help or something like that. The car. There's a radio there, a radio (…) /er/ plays music. And cat trying to catch some fish, knocks over the books, and /er/ a flower pot.

b) Written output

man

cat goldfish

cat … the book

a st… (*target: stereo*)

child wi (*target: with*) a toy

A table wi (*target: with*) a

Appendix C

Copying tasks

Task		*N* of items	GEC (%) [1]	JOD (%) [1]
Words				
In sight	Short (4 letters)	20	100	100
	Long (8 letters	20	100	95
Direct (no delay)	Short (4 letters)	20	80	90
	Long (8 letters	20	5	0
Delayed (5 sec)	Short (4 letters)	20	60	75
	Long (8 letters)	20	0	0
Non-words				
In sight	Short (4 letters)	10	100	100
	Long (8 letters)	10	95	100
Direct (no delay)	Short (4 letters)	10	20	90
	Long (8 letters	10	20	0
Delayed (5 sec)	Short (4 letters)	10	0	60
	Long (8 letters	10	0	0

[1] Words were presented in lower case. GEC copied the words in the same case, however JOD preferred to write all responses in upper case.

Appendix D

Overview of stimuli for writing to dictation

Source of word list	Factors of interest
Krajenbrink, Nickels, Kohnen (unpublished)	Frequency, Imageability, Regularity
Buchwald & Rapp (2009, revised version from Lavidor & Ellis, 2002)	Length
Coltheart, Laxon & Keating (1988)	Age of Acquisition, Imageability
Laxon, Coltheart & Keating (1988)	Neighbourhood size
Johns Hopkins Dysgraphia battery (Goodman & Caramazza, 1985): Length [1]	Length, Frequency
Johns Hopkins Dysgraphia battery (Goodman & Caramazza 1985): PGC	Phoneme-Grapheme probability, Frequency
Johns Hopkins Dysgraphia battery (Goodman & Caramazza 1985): Part of Speech [1]	Word class, Frequency
PALPA 44 Regularity and spelling (Kay et al., 1992)	Regularity
PALPA 40 Imageability and Frequency and Spelling (Kay, et al., 1992) [2]	Imageability, Frequency

[1] Two Johns Hopkins Dysgraphia Battery lists (Length and Part of Speech) were only administered to participant GEC.
[2] The PALPA 40 list was only administered to participant JOD.

Appendix E

Item characteristics treatment stimuli: Mean and standard deviation (in brackets) for lexical variables (see Notes for explanation of variables)

	Treated T1	Treated T2	Untreated C1	Untreated C2	Untreated C3
CELEX total	53.12 (56.00)	37.50 (40.71)	52.70 (57.41)	38.61 (63.76)	70.17 (107.07)
CELEX_W Log	1.54 (0.45)	1.42 (0.39)	1.54 (0.42)	1.36 (0.44)	1.45 (0.61)
Length (L)	5.55 (0.51)	5.55 (0.60)	5.60 (0.50)	5.55 (0.60)	5.35 (0.49)
Length (S)	1.75 (0.92)	1.5 (0.94)	1.95 (0.64)	1.5 (0.51)	1.55 (0.51)
N size	0	5.55 (1.23)	0	5.10 (2.38)	5.55 (0.94)
HFN	0	1.70 (0.92)	0	1.45 (1.20)	1.25 (1.25)
LFN	0	3.85 (1.50)	0	3.65 (2.71)	4.30 (1.49)
NF_Max	0	144.65 (129.21)	0	111.85 (104.23)	177.23 (231.49)
NF_Min	0	1.36 (1.21)	0	11.67 (24.35)	0.95 (0.64)
NF_Mu	0	33.07 (19.79)	0	40.97 (43.02)	43.42 (53.35)
NF_Sig	0	67.66 (50.77)	0	60.18 (66.15)	89.32 (115.29)
PN	4.10 (4.36)	10.00 (4.27)	2.95 (2.48)	9.40 (5.05)	11.10 (5.99)
AoA	7.19 (1.96)	6.69 (1.87)	7.78 (2.86)	6.94 (1.83)	7.51 (2.44)
BF_TK	653.66 (319.92)	1458.68 (723.70)	881.69 (597.24)	1619.16 (771.81)	1629.41 (920.47)

Note. **T1** = Treated items Phase One (no neighbours); **T2** = Treated items Phase Two (many neighbours); **C1** = Control set 1 (no neighbours); **C2** = Control set 2 (direct neighbours T2); **C3** = Control set 3 (many neighbours).

CELEX total = total CELEX frequency per million; **CELEX W Log** = Logarithmic written frequency; **Length (L)** = Length in letters; **Length (S)** = Length in syllables; **N** = Orthographic neighbourhood size (one letter substitution neighbours, Coltheart et al., 1977); **HFN** = Number of orthographic neighbours that have higher frequency than input item; **LFN** = Number of orthographic neighbours that have lower frequency than input item; **NF_Max** = Frequency of the highest frequency neighbour; **NF_Min** = Frequency of lowest frequency neighbour; **NF_Mu** = Average frequency of neighbours; **NF_Sig** = Summed frequency of neighbours; **PN** = Number of phonological neighbours; **AoA** = Mean rating (in years) for age of acquisition (Kuperman et al., 2012); **BF_TK** = Average bigram token frequency across the entire letter string.

All treated and control sets consisted of 20 items. In all sets nouns were the most frequent word class. Across all sets the items did not differ on frequency or length in letters. The sets with neighbours were also matched on number of neighbours, summed and average frequency of the neighbours, number of neighbours higher and lower in frequency, and frequency of the highest and lowest frequency neighbour.

We were unable to apply the same criteria to phonological neighbourhood size, i.e., it was not possible to find words with no orthographic neighbours and no phonological neighbours. We did however ensure that the treated sets that differed in orthographic neighbourhood size (T1 and T2) also contrasted in phonological neighbourhood size ($t(38) = 4.32$, $p < .001$). Also, the sets without orthographic neighbours (T1 and C1) did not differ on number of phonological neighbours ($t(38) = 1.02$, $p > .100$).

The words in C3 all had many orthographic neighbours, and we aimed to exclude words that were neighbours of any of the other words in the stimuli set. In order to keep other variables controlled for, we had to include one orthographic neighbour of C2 in this set. Stimuli are available from the first author upon request.

GENERAL DISCUSSION

The overall aim of this thesis was to explore the nature of impairment and rehabilitation in acquired dysgraphia, following a cognitive neuropsychological approach. The thesis had the following specific aims:

1. To further specify the theory of impairment for acquired dysgraphia, in particular impairment to the graphemic output buffer and impairment in sub-lexical processing.
2. To study the mechanisms of treatment generalisation, by reviewing the literature of treatment studies and investigating the relationship between the type of impairment, the method of treatment, and whether generalisation occurred.
3. To explore interactivity within the spelling process by investigating the role of neighbourhood size on treatment and generalisation.

This chapter will first provide a brief summary of the four experimental papers, including their main contributions to our understanding of the spelling process. Finally, difficulties interpreting data from the different methodologies used will be discussed.

The aim of **Study One** was to further specify theory of impairment in graphemic output buffer dysgraphia, through a single case study. Three characteristics of GEC's spelling were investigated in order to inform theory: 1) effects of lexical factors on spelling; 2) the serial position curve of errors; and 3) fragment errors.

We reviewed GEC's performance in the light of the current literature on lexical influences on graphemic output buffer impairment and found that his error pattern did not support an additional lexical impairment, despite the lexical influences he showed on spelling performance. Instead, GEC's error pattern is consistent with graphemic output buffer impairment where the effects of frequency on buffer functioning occur as a result of interactive processing within the spelling system (e.g., Buchwald & Rapp, 2009; Sage & Ellis, 2004). Differences in strength of activation of lexical items (e.g., stronger activation for more frequent items) cascade down to the graphemic output buffer and impact processing with higher frequency items being more resistant to buffer impairment.

Second, we investigated the nature and serial position functions of GEC's errors. GEC showed a large number of fragment errors, resulting in a clear linear increase in errors towards the end of words. Fragment errors have been suggested to result from impairment to the graphemic output buffer (e.g., Katz, 1991; Schiller, Greenhall, Shelton, & Caramazza, 2001) or from a lexical impairment (e.g., Ward & Romani, 1998). A number of experimental tasks (e.g., a letter probe task) were used to test the underlying impairment resulting in fragment errors. GEC's retrieval of final letters improved on tasks that reduced the workload of the graphemic output buffer. These results indicate that fragment errors were due to rapid decay at the level of the buffer rather than impaired lexical representations.

Thorough analyses of errors and additional experimental tasks used in previous case studies allowed us to show that GEC's pattern of performance was best explained as the result of rapid decay of activation from the graphemic output buffer.

Study Two aimed to inform our understanding of sub-lexical impairment in dysgraphia, and specify theories of impairment, using a case series approach. We used Houghton and Zorzi's (2003) computational model of spelling and data from previous studies on normal spellers (Perry, Ziegler, & Coltheart, 2002) to guide our investigations.

We investigated the impairment of sub-lexical processing by examining spelling of sounds in isolation and spelling of non-words in people with aphasia, focusing on three areas: 1) the performance of people with aphasia on spelling sounds in isolation and spelling sounds in the context of a non-word, and the relationship between the tasks; 2) the effects of consistency and frequency on spelling of PGCs; and 3) the use of phonological context when spelling a vowel in a non-word.

We found that people with aphasia showed more difficulty with spelling sounds in the context of a non-word, compared to spelling sounds in isolation. Interestingly, this difference in performance disappeared for most individuals when the comparison only included PGCs in initial position in the non-word, indicating that orthographic working memory processes involved in non-word spelling can explain at least part of the added difficulties when spelling non-words.

However, some individuals performed worse on spelling initial sounds in non-words than spelling sounds in isolation, suggesting that PGC knowledge was relatively retained compared to segmentation. The possible differences in performance on spelling individual sounds and spelling non-words and their implications for underlying impairments show the importance of assessing sub-lexical processing using both tasks.

This paper found large variability in individual sub-lexical spelling patterns. We showed that sub-lexical impairment can be due to poor PGC knowledge, poor segmentation, and/or decreased orthographic working memory capacity. In terms of the impairment to phoneme grapheme conversion, we demonstrated that PGCs with higher frequency and consistency were more resistant to damage. Our final analysis revealed no evidence for an impairment of context specific rules when spelling vowels.

The third and fourth research papers investigated treatment of acquired dysgraphia. **Study Three** reviewed 40 treatment studies of acquired dysgraphia, in order to better understand the process of generalisation in treatment. A better understanding of the mechanism underlying generalisation may have clinical benefits (as it can improve treatment efficacy) and it can also inform theories of spelling and spelling impairment.

The review discussed several general principles that underlie the mechanism of generalisation: Direct treatment effects on representations or processes; interactive processing and summation of activation, and strategies or compensatory skills. We suggested future research to further discern the contributions of type of impairment and type of treatment to the outcome of treatment.

Study Four was designed to test one of the predictions of a possible mechanism that may induce generalisation: feedback between the orthographic lexicon and the graphemic output buffer (e.g., Buchwald & Rapp, 2009; Kohnen, Nickels, Coltheart, & Brunsdon, 2008; Sage & Ellis, 2006). We hypothesised that during treatment, untreated neighbours of treated items may benefit from treatment due to co-activation from the treated items. In turn, these untreated neighbours may also support spelling of the treated words. As a result, treating words with many neighbours may lead to a larger treatment effect compared to words without neighbours. We tested this assumption by first treating words that had no neighbours and in then treating items with a large neighbourhood size. The results showed that for two individuals with dysgraphia, the number of orthographic neighbours did not affect the size of the treatment. We argued that the severity of impairment in the two individuals may have inhibited the amount of available feedback.

Difficulties interpreting data in relation to theory. We will now discuss several issues that were observed in the studies in this thesis that may make it difficult to draw conclusions and inform theory of impairment and rehabilitation of dysgraphia.

From individual cases to theory. Study One showed that a detailed error analysis can be a powerful methodology to inform theory, and that it can help to distinguish between contrasting hypotheses, in this case regarding the origin of fragment errors. However, it can be difficult to make a direct comparison between cases studied in different labs, as often different sets of data have been used. For example, different serial position effects have been found in graphemic output buffer impairment: a bow shaped error pattern has been linked to the process of 'interference', and a linear increase in errors has been linked to 'decay'. In order to be able to draw conclusions about the nature of the impairment, it is crucial that a representative sample of errors be included to plot serial position effects. Sage and Ellis (2004) noted that serial position analyses are often based on single letter errors. Interestingly, for their participant BH many errors affected the final part of the word, often involving deletions of multiple letters. Including these errors would have shifted the serial position of errors to the right, and hence may have led to different inferences to be made about theory of serial processing. The reason why traditionally only single letter errors were included in position analyses is that many multiple letter errors are ambiguous regarding the position of the error. However, Study One shows that fragment errors should also be included in these analyses. Further studies examining the impact of including responses with multiple letter errors of different types are recommended.

This study also highlighted how different cases of graphemic buffer impairment in the literature have shown varied patterns of impairment (e.g., different serial position curves). It seems that buffer processing consists of different components that can be distinctly impaired in graphemic buffer patients, however traditional models of spelling remain somewhat underspecified regarding these components and how the different processes interact (Miceli & Capasso, 2006). Understanding how working memory processes in spelling relate to the broader concept of the organisation of serial behaviour (e.g., Olson, Romani & Caramazza, 2010), can further inform theory of spelling.

The difficulty in comparing and interpreting data across cases is also pertinent to the review of treatment studies (Study Three). Establishing the role that item characteristics have on generalisation in treatment (and hence on the process of spelling) was difficult because different item have been controlled and/or manipulated across studies. For example, a number of studies have suggested frequency and neighbourhood size of items can predict generalisation, but yet the frequency and neighbourhood size of treated and untreated stimuli were not specified in all studies, nor were the stimuli provided. It could be the case that some of the studies that showed generalisation used items that were indeed high in frequency and neighbourhood size, but it is impossible to determine from the published article. Of course, it can be difficult to consider the entire literature when conducting a treatment study (or carry one out in clinical practice), and different characteristics come to the fore as being important over time. Hopefully, the review conducted as part of this thesis will help researchers and clinicians to be more aware of the relevant features and item characteristics and thus include them in future studies. Moreover, even when this is not possible, stimuli should be made available to allow post hoc examination of their characteristics.

Furthermore, in order to test theories, it is important to have a clear understanding of the particular task. This may sound trivial, but for several of the tasks central to this thesis the exact processes involved are underspecified. For example, in Study One we used a backward spelling task. This was first reported by Katz (1991) who asked individual HR was asked to spell a word by starting with the last letter. Results of this task have been used to make conclusions about possible underlying impairment (Katz, 1991; Ward & Romani, 1998). Schiller et al (2001) raised the important issue that without a clear understanding of how backward spelling is performed, it is not straightforward to link performance on this task to underlying impairment. For example, when spelling backwards, rather than the graphemes being accessed in reverse order, it is possible that the orthography is ‘scanned’ in a forward manner until the to-be-retrieved letter is reached, making the task less different from forward spelling. Therefore, impairment on this task could be due to difficulty performing such an unusual task without this result being directly informative about the functioning of graphemic output buffer. It is vital that careful thought is given to the possible processes that could be involved in a task rather than taking them at face value.

Computational modelling of tasks may prove a fruitful method in specifying what the relevant mechanisms that are required to perform this and other tasks. However, currently many computational models are limited in their ability to simulate more than straightforward input-output mappings and it is unclear how more complex tasks would be implemented. Nevertheless, computational modelling has a strength in that it specifies theoretical assumptions and can be used to simulate patterns observed in human data. For example, studies of adult spelling have shown that spellers are sensitive to phonological context when spelling an inconsistent grapheme, such as after a long vowel /k/ is more likely to be spelled as K (e.g., *leak*), rather than being spelled as the digraph CK after a short vowel (e.g., *lick*). Houghton and Zorzi (2003) implemented this rule by allowing connections of different weights between phonological input and graphemic output (e.g., strong links between 'long vowel' and K but inhibitory links to CK), which allowed the model to use information from surrounding context to spell inconsistent phoneme-grapheme correspondences.

To be a 'complete' theory of a cognitive process a computational model also needs to be able to simulate impairment. For example, the data in Study Two showed a difference in performance between spelling sounds in isolation and in the context of a non-word (except for the initial phoneme). The two tasks have been proposed to involve different components: spelling sounds involves mainly phoneme-grapheme conversion, whereas non-word spelling also involves a segmentation component (e.g., Roeltgen, Sevush, & Heilman, 1983). The Houghton and Zorzi (2003) computational model of spelling does not incorporate an equivalent of a segmentation process and hence it is unclear whether it could simulate the difference in performance on these tasks. Data from impairments remain a powerful way to test the adequacy of computational models and theories generally.

From case series to theory. Case series aim to combine data from people with a similar impairment, to answer a question about a certain cognitive process (Schwartz & Dell, 2011). Investigations of the variability across cases can be especially informative in this context (e.g., Fischer-Baum & Rapp, 2012). However, as raised in the Introduction, if heterogeneity in the group impacts on performance of the factor of interest, this may limit the conclusions that can be drawn (Rapp, 2011). In Study Two we reported data from a series of cases, but the individuals in this study were not specifically selected on the basis of severity of impairment to a particular component in spelling. Considering the assumptions of a case series as set out by Schwartz and Dell, the heterogeneity in our sample may have limited the conclusions that could be drawn. However, in our study we did not aim to investigate, for example, how the severity of a graphemic output buffer impairment may impact on non-word spelling abilities, but rather provide a broad exploration of sub-lexical processing in a group of people, as this has not been studied before. Based on adult spelling skills and a computational model (Houghton & Zorzi, 2003) we were able to formulate specific hypotheses about the effects of impairment on processing. By using a case series rather than a single case we were able to provide stronger evidence of the generality of the findings.

Study Two can be seen as a first step in further specification of a theory of sub-lexical impairment. A follow up study could use a case series approach by selecting individuals on the basis of the severity and type of their sub-lexical impairment (e.g., segmentation difficulties, PGC knowledge, orthographic working memory). For example the impact of orthographic working memory impairment on sub-lexical spelling could be examined by selecting individuals with intact PGC spelling in isolation and examining the relationship between the severity of their orthographic working memory impairment and PGC accuracy in non-words. Case series analyses can always be supplemented by single case studies which enable a more detailed level of analysis.

From treatment to theory. Results of treatment can inform theory regarding underlying impairments or the cognitive process in general. For example, Study Four investigates whether feedback between the lexicon and buffer results in larger treatment effect for words with many neighbours. One difficulty interpreting treatment data is the interpretation of 'null results' (although of course other methodologies also suffer from this problem).

In Study Four we hypothesised that if orthographic neighbours play a role in treatment, the size of the treatment effect may be larger when words with many neighbours are treated compared to when words without neighbours were treated. No difference in the size of the treatment effects was found. How should this result be interpreted? A number of plausible explanations can be considered: 1) this result provides no evidence for a mechanism of feedback in the spelling system; 2) the critical manipulation (in this case: the number of one letter substitution neighbours for an item) is not sensitive enough to test the theoretical question or hypothesis, or 3) the individuals in this particular study show reduced feedback. In Study Four we argue for the third possibility - that severity of impairment affected the feedback available in spelling for GEC and JOD. Replication of the treatment in more studies of a similar design with individuals with different severity of impairment is important to further test this hypothesis.

Clinical implications. While the focus of this thesis has been on theory and theory of impairment, the papers in this thesis also have clinical implications.

Study One illustrated the range of characteristics of performance that can be associated with graphemic output buffer impairment. This is relevant for the diagnosis of spelling impairments: if fragment errors are produced by an individual, this may be an indication of graphemic output buffer impairment. A clear diagnosis of underlying impairment may then guide the choice of treatment.

As noted above, Study Two also has two clear implications for clinical practice: First, in order to provide the most accurate diagnosis of the functioning of sub-lexical spelling processes, it is important to include both non-word spelling and spelling of sounds in isolation. Secondly, knowing which sub-lexical processes (including which PGCs) are impaired can guide the decision of the type of treatment, and also which PGCs should be treated.

Finally, the treatment study (Study Four) showed that it is possible to improve spelling of words used in treatment, even for two individuals with relatively severe graphemic output buffer impairments resulting in poor overall performance. The results from this study also showed that for the individuals in this study, performance declined after the end of treatment. This suggests that in order to maintain improvement continued practice is required. The combination of item specific improvements and the need for regular spelling to maintain benefits also reinforces the importance that treated items should be functionally relevant.

Other methodologies and future directions. Another approach that has been used to inform theory is investigating neural substrates of spelling in a group of individuals with dysgraphia (e.g., Rapcsak et al., 2009). Rapcsak et al. (2009) reported a group study using information about lesions to further test the theoretical assumption of a general phonological impairment underlying phonological dyslexia and phonological dysgraphia. If both phonological dyslexia and dysgraphia are associated with an impairment of phonological processing, a correlation between the lesion and the deficit would be expected. For example, the perisylvian cortex has been suggested to be involved in phonological processing, and therefore impairment to this region may be related to phonological dyslexia and dysgraphia. In order to test this correlation, Rapcsak et al. (2009) selected 31 participants with damage to the perisylvian cortex, and investigated the relationship between reading and spelling performance and phonological processing skills. Results showed quantitative differences in spelling performance in relation to the degree of phonological impairment in individual patients. Rapcsak et al argue that this correlation suggests that phonological dyslexia and dysgraphia are both the result of a central phonological impairment.

In addition to informing theory, neuroimaging studies with healthy adults can also inform our understanding of the neural substrates of spelling (e.g., Beeson et al., 2003). Using functional magnetic resonance imaging, Beeson et al. (2003) examined writing in a group of 12 healthy adults, and investigated the role of different cortical regions in central and peripheral components of writing. It was found that the left posterior inferior temporal cortex was involved in central components of writing (i.e., a generative writing task), and a left-hemispheric fronto-parietal network was involved in more peripheral components of writing such as motor programming (while participants were drawing circles).

These studies show that in addition to cognitive neuropsychological single case or case studies, neuroimaging studies are a methodology to inform theory of the neural mechanisms of writing, using either data from healthy subjects or lesion data from impaired subjects. However, it has been debated exactly how neuroimaging studies can inform cognitive theories (Coltheart, 2013). The study by Beeson et al. aims to understand the localisation of specific cognitive subsystems, which is one aim of functional neuroimaging of cognition (Coltheart, 2013). However, Coltheart argues that localisation of cognitive processes does not always inform cognitive theories, but rather these studies are informed *by* cognitive theories, as the studies are based on a model of the cognitive system. Furthermore, when testing cognitive theories with functional imaging data it is important to establish clearly what the predictions are based on neuroimaging data, i.e., whether the data are consistent and inconsistent with the predictions of a particular theory (Coltheart, 2013).

Neuroimaging data can be informative about localisation of cognitive processes, but studies designed to test cognitive theories should establish clearly how the data inform a cognitive theory. Furthermore, some cognitive theories may not make any predictions about brain activity, and therefore neuroimaging experiments would not be able to inform these theories (Coltheart, 2013).

With regards to treatment it has been shown in Study Four that cognitive models of language may help to focus treatment, however they do not always predict exactly how individual patients will respond to treatment (Hillis & Heidler, 2005). Studies focusing on the nature of learning and mechanisms of recovery could help inform decisions regarding type and method of treatment (e.g., Sage, Snell, & Lambon Ralph, 2011). Furthermore, recent work on aphasia rehabilitation has explored neuro-stimulation techniques in addition to behavioural treatment (e.g., transcranial direct current stimulation (tDCS)). Understanding the mechanisms underlying these techniques combined with model-based treatment will provide useful in deciding on suitable treatment options for individual patients in the future.

General Conclusion

This thesis examined the nature of acquired dysgraphia, to inform theories of impairment and rehabilitation. The individual studies have contributed to our understanding of the components that make up the complex process of spelling. They have also shown the utility of a variety of methodologies used in cognitive neuropsychological research on spelling. They have demonstrated, for example, that detailed error analyses in single case studies, case series and treatment studies are all powerful methodologies to extend our understanding of the different features of dysgraphia and the spelling system. Furthermore, this thesis has highlighted how consideration of insights from computational modelling combined with case studies and case series can be a successful avenue for future research leading to a better understanding of the architecture of spelling.

More specifically, this thesis has investigated the nature of graphemic buffer impairment, sub-lexical processing and interactivity within the spelling system. It is hoped that all of these findings will encourage future research in order to better understand the spelling system and its impairment and this understanding will promote more effective treatments.

References

Beeson, P. M., Rapcsak, S. Z., Plante, E., Chargualaf, J., Chung, A., Johnson, S. C., Trouard, T. P. (2003). The neural substrates of writing: A functional magnetic resonance imaging study. *Aphasiology, 17* (6-7), 647-665.

Buchwald, A., & Rapp, B. (2009). Distinctions between orthographic long-term memory and working memory. *Cognitive Neuropsychology, 26*(8), 724-751.

Caramazza, A., & Hillis, A. (1993). For a theory of remediation of cognitive deficits. *Neuropsychological Rehabilitation, 3*(3), 217-234.

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: Evidence from a case of acquired dysgraphia. *Cognition, 26,* 59-85.

Coltheart, M. (2013). How can functional neuroimaging inform cognitive theories? *Perspectives on Psychological Science, 8*(1), 98-103.

Fischer-Baum, S., & Rapp, B. (2012). Underlying cause(s) of letter perseveration errors. *Neuropsychologia, 50*(2), 305-318.

Hillis, A. E., & Heidler, J. (2005). Contributions and limitations of the cognitive neuropsychological approach to treatment: Illustrations from studies of reading and spelling therapy. *Aphasiology, 19*(10-11), 985-993.

Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology, 20*(2), 115-162. DOI: 10.1080/02643290242000871.

Katz, R. B. (1991). Limited retention of information in the graphemic output buffer. *Cortex, 27*, 111-119.

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.

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

Olson, A., Romani, C., & Caramazza, A. (2010). Analysis and interpretation of serial position data. *Cognitive Neuropsychology, 27*(2), 134-151.

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.

Rapcsak, S. Z., Beeson, P. M., Henry, M. L., Leyden, A., Kim, E., Rising, K., Andersen, S., Cho, H. (2009). Phonological dyslexia and dysgraphia: Cognitive mechanisms and neural substrates. *Cortex, 45,* 575-591.

Rapp, B. (2011). Case series in cognitive neuropsychology: Promise, perils, and proper perspective. *Cognitive Neuropsychology, 28*(7), 435-444.

Roeltgen, D. P., Sevush, S., & Heilman, K. M. (1983). Phonological agraphia: Writing by the lexical-semantic route. *Neurology, 33*, 755-765.

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

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

Sage, K., Snell, C., & Lambon Ralph, M. A. (2011). How intensive does anomia therapy for people with aphasia need to be? *Neuropsychological Rehabilitation, 21* (1), 26-41.

Schiller, N. O., Greenhall, J. A., Shelton, J. R., & Caramazza, A. (2001). Serial order effects in spelling errors: Evidence from two dysgraphic patients. *Neurocase, 7*, 1-14.

Ward, J., & Romani, C. (1998). Serial position effects and lexical activation in spelling: Evidence from a single case study. *Neurocase, 4*, 189-206.