Remediation of deficits affecting different components of the spelling process

Brenda Rapp and Amy Kane

*Johns Hopkins University, Baltimore, MD, USA*

**Background:** There have been relatively few studies concerned with the treatment of spelling deficits. Among these, there have been a small number that have targeted specific components of the spelling process. Although most of these studies report success using treatments that involve repeated spelling and/or copy, the results have been mixed, especially as concerns the generalisation of treatment benefits to untreated items.

**Aims:** This investigation was designed to examine the responsiveness to the same treatment protocol of deficits affecting different cognitive mechanisms of the spelling process.

**Methods & Procedures:** We applied the same delayed-copy treatment protocol to two individuals with selective deficits of the orthographic output lexicon and the graphemic buffer. The two individuals were otherwise matched in terms of the severity of their deficits and their general cognitive profiles.

**Outcomes & Results:** Both individuals exhibited long-lasting word-specific benefits from the treatment. However, they differed in that the graphemic buffer deficit exhibited generalisation to untreated words, whereas the orthographic output lexicon did not.

**Conclusions:** The absence of presence of generalisation effects in response to the successful treatment of target items is determined by the specific cognitive component/s that constitute the source of the deficit.

In contrast to the numerous treatment studies of spoken language deficits, there have been little more than two dozen papers concerned with the remediation of deficits affecting the spelling process (see Beeson and Rapcsak, 2002 for a review).

Nonetheless, the treatment of spelling impairments is especially important both because of the pervasive use of spelling in everyday life (e.g., writing cheques, making shopping lists, taking phone messages, writing Christmas cards, address-book entries), and because the remediation of written language deficits may sometimes be more successful than that of spoken impairments (Beeson, 1999). For these reasons, the remediation of written language may often constitute a successful means of reinstating functional communication skills.

In this paper we report the results of the application of the same remediation protocol with individuals with dysgraphia suffering from deficits affecting two different components of the spelling process. The differential responsiveness of the two

---

1 We refer here to treatments for deficits affecting the ‘central’ aspects of the spelling process, as opposed to treatments for deficits of motor execution.

---

Address correspondence to: Brenda Rapp, Dept. of Cognitive Science, Johns Hopkins University, Baltimore, MD 21218, USA. Email: rapp@cogsci.jhu.edu

We are very grateful to MMD and RSB for their cheerful participation in this project and to Donna Gotsch for referring MMD to our laboratory. Delia Kong’s assistance in preparing the manuscript is very much appreciated. This work was carried out with the support of NIMH grant R29MH55758.
individuals to the same protocol contributes to our understanding of the underlying cognitive mechanisms themselves as well as of the rehabilitation of deficits involving these mechanisms.

**THEORY OF SPELLING**

We will define and discuss the spelling process and dysgraphic deficits within a theory that specifies the cognitive components of the spelling process (Caramazza, 1988; Ellis, 1989). According to the theory depicted in Figure 1, spelling can be achieved by means of two sets of processes—lexical and sublexical. The lexical “route” makes use of information about words that has been stored in long-term memory. The sublexical route makes use of stored knowledge of the regular relationships between the sounds and letters of English.

In the task of spelling to dictation, within the lexical route an auditorily presented word activates a phonological representation of the word in the long-term memory store referred to as the phonological input lexicon. Activation of this representation allows the

![Diagram of the spelling process](image)

**Figure 1.** Schematic depiction of the cognitive architecture of the spelling system (from Rapp, Epstein, & Tainturier, 2002).
listener to gain access to a representation of the word’s meaning in the semantic system. This representation, in turn, can be used to activate a representation of the word’s spelling in the orthographic output lexicon (OOL)—the long term memory repository of the spellings of familiar words.

The sublexical system accepts all phonological strings as input and does not require that the stimulus be a word that is familiar to the listener (nonwords are processed in the same manner as words). This sublexical process applies knowledge of sound-to-spelling correspondences to the input phonological string and yields a phonologically plausible spelling (e.g., ‘‘yacht’’ → YOT).

The spelling representations generated by either route are stored in a short-term working memory component called the graphemic buffer that maintains the orthographic representation active while each letter is assigned either a shape—upper case, lower case, cursive—for written spelling or a letter name (for oral spelling). Finally, the letter shape or letter name representations are produced by means of the appropriate motor systems.

Research has shown that individual components of the spelling process can be selectively affected by neurological damage (Caramazza, Miceli, Villa, & Romani, 1987; Miceli, Silveri, & Caramazza, 1985). Importantly, damage to different components of the process will manifest itself differently. Because the two individuals we describe in this paper suffered damage to the orthographic output lexicon and the graphemic buffer we will restrict our review of expected behavioural manifestations of damage to these two components (for a more detailed discussion that includes the full set of components see Rapp & Gotsch, 2001).

According to the theory depicted in Figure 1, damage to the orthographic output lexicon (OOL) should give rise to the following pattern of performance: (1) intact comprehension; due to the fact that the damage is post-semantic; (2) an effect of lexical frequency: the OOL is sensitive to the frequency with which a word has been encountered and thus damage to the OOL results in high-frequency words being spelled more accurately than low-frequency words; (3) the absence of word length effect: the OOL is assumed to be insensitive to the number of letters in a word; (4) phonologically plausible errors (PPEs) for irregularly spelled words: damage to the OOL forces reliance on the sublexical route to assemble phonologically plausible spellings for the words that have been affected by OOL damage (e.g., ‘‘rough’’ → ruff).

Damage to the graphemic buffer should yield the following: (1) intact comprehension; as in the case of damage to the OOL, the damaged component is post-semantic; (2) the absence of an effect of lexical frequency: the graphemic buffer is thought to be insensitive to word frequency; (2) an effect of word length: the graphemic buffer acts as a working memory component and is therefore sensitive to the amount of material that must be held in temporary storage; (3) letter substitutions, deletions, and transpositions: these are the types of errors expected to arise from disruption that occurs subsequent to intact lexical and sublexical processing; no semantic errors and very few PPE’s would be expected.

TREATMENT LITERATURE

A number of treatment studies have targeted specific components of the spelling process: the semantic component (Hillis, 1991, 1992; Hillis & Caramazza, 1991), the sublexical conversion system (Carломagno, Iavarone, & Colomba, 1994; DePartz, Seron, & Van der Linden, 1992; Hatfield, 1983; Luzzatti, Colomba, Frustaci, & Vitolo, 2000); the

---

2PPEs should be observed only to the extent that they ‘‘accidentally’’ result from simple letter substitutions, deletions, etc. (e.g., ‘‘science’’ → SIENCE).
graphemic buffer (Hillis, 1989); the orthographic representations within the OOL (Aliminosa, McCloskey, Goodman-Schulman, & Sokol, 1993; Beeson, 1999; Beeson & Hirsch, 1998; Behrman, 1987; Behrmann & Byng, 1992; Carlomagno et al., 1994; DePartz et al., 1992; Hillis & Caramazza, 1987; Seron, Deloche, Moulard, & Rousselle, 1980; Weekes & Coultheart, 1996); and treatments designed to facilitate the interaction between lexical and sublexical processes (Beeson, Rewega, Vail, & Rapcsak, 2000). Given the focus of this investigation, we will briefly summarise findings relevant to the OOL and the graphemic buffer.

The majority of treatments directed at strengthening orthographic representations of word spellings in the OOL, although differing in certain details, share the fact that the correct spellings are presented repeatedly for study and spelling (and/or delayed copy). The rationale for such an approach is that it is assumed that repeated exposure to, and activation of, the target words will strengthen their representations in long-term memory. For example, Beeson and Hirsch (1998) and Beeson (1999) examined the efficacy of an anagram and copy treatment (ACT) and a delayed copy treatment (CART) and found that the delayed copy treatment alone was efficacious (see also Hatfield & Weddel, 1976). This prompted Beeson and Rapcsak (2002) to write “we advocate combined stimulation of repeated reading and writing of target words in order to maximize activation of the orthographic lexicon” (p. 20).

Although studies using this general approach for strengthening word representations in the OOL all report success in the treatment of targeted items (unsuccessful outcomes are, of course, unlikely to be published), they vary regarding the generalisation of the treatment benefits to untreated items. It is generally assumed that strengthening the representations of specific words in the OOL should not necessarily yield a benefit for other untreated items. Consistent with this, Aliminosa et al. (1993), DePartz et al. (1992), Weekes and Coltheart (1996), and Beeson (1999) all reported successful treatment of target items with no generalisation to untrained items. In fact, Aliminosa and colleagues (1993) and Beeson (1999) use the absence of generalisation as evidence that the deficit was indeed localised to the OOL. In contrast, Behrman (1987, see also Behrmann & Byng, 1992) and Seron et al. (1980) reported at least partial generalisation effects.

With regard to the remediation of graphemic buffer deficits, it is important to note that there have been very few published treatment studies involving individuals with clearly diagnosed graphemic buffer deficits. Given this, it is not surprising that there is not yet any consensus regarding the rationale for a particular approach to the treatment of these deficits. For example, Hillis and Caramazza (1987) investigated the effectiveness of teaching error-detection and correction strategies to individuals with graphemic buffer deficits who have intact word recognition skills. In another study Hillis (1989) applied the same cueing hierarchy treatment to individuals with different underlying spoken and written production deficits (only patient 2 in this study suffered from a graphemic buffer deficit). Hillis suggested “that identical treatment can improve the performance of different patients for different reasons” (p. 635). In both investigations treatment of target items was successful.

With regard to expectations regarding generalisation of treatment of graphemic buffer deficits, it is generally assumed that as the graphemic buffer is employed in the spelling of all words (not simply those targeted by the treatment), the successful treatment of the graphemic buffer should generalise to untreated items. However, here again we see mixed results—Hillis and Caramazza (1987) reported generalisation, whereas Hillis (1989) did not.
In sum, although the published studies report considerable success in treating targeted items in cases of OOL and graphemic buffer deficits, the studies differ considerably as to whether or not transfer to untreated words was obtained. Furthermore, although the expectation was that the successful treatment of graphemic buffer deficits should lead to generalisation while the successful treatment of OOL deficits should not, the generalisation results do not pattern in this way. What might account for these discrepancies? There are a number of possibilities. Although limitations in space preclude a thorough review of this question, the following factors are likely to be relevant. First, there may have been an incorrect diagnosis of the deficits, or the presence of mixed spelling deficits may have complicated the picture. Second, although the remediation procedures for OOL deficits largely shared a ‘repeated study–copy–spell’ component, they may have differed in other crucial respects. Finally, there is the possibility that the differences among the patients with regard to other cognitive skills may have influenced their responsiveness to treatment and/or generalisation.

The research we report here was designed to address some of these concerns and to provide a means for examining the underlying cognitive mechanisms of the spelling process and their responsiveness to treatment. In order to do so we applied the same treatment protocol to two individuals with fairly selective deficits to the OOL and the graphemic buffer. Furthermore, the severity of the subjects’ deficits is comparable and many of their other cognitive skill levels are also similar (e.g., memory and spoken language comprehension and production). It was our expectation that, given these conditions, we should be able to more readily attribute any differences observed in responsiveness to treatment to differences in the underlying deficits and the character of the cognitive mechanisms themselves.

**CASE STUDIES**

MMD was a right-handed woman who suffered a CVA in November of 1998 at the age of 65, two and a half years before the onset of this investigation. MMD worked in a clerical position until retirement. She graduated from high school and reports having always been a good speller and avid reader. Premorbid writing samples reveal less than 0.2% errors in spelling. CT scans indicate left posterior parietal and temporal lesions. The CVA produced mild spoken language difficulties and a significant spelling impairment. MMD was able to drive and live completely independently.

RSB was a 58-year-old, right-handed man who suffered a CVA subsequent to aortic valve replacement at the age of 54, four years before the onset of this investigation. RSB held a PhD and worked as a toxicology researcher, reporting no premorbid spelling difficulties. He was a bilingual speaker of English and Spanish, although he had spoken Spanish only infrequently after the age of 15. MRI scans reveal a lesion in the left anterior parietal region. The CVA produced moderate difficulties in spoken language production and significant difficulties in number processing and spelling. Although the stroke forced his early retirement, he was able to drive, read for pleasure and manage most of the family’s affairs.

**Pre-treatment assessment**

*Memory.* The Wechsler Memory Scale–Revised (WMS–R; Wechsler, 1987) was administered to both MMD and RSB. RSB was administered this test once, before the onset of the remediation investigation. His general memory score was 101 (mean = 100, SD = 15) and his scores on the various memory indexes ranged from 63 to 117 (with a
standard score of 117 on the index of visual memory). MMD was administered the WMS–R on two occasions, once prior to the onset of the remediation investigation and again after the final evaluation. On the first administration her general memory score was 88 (mean = 100, SD = 15) and she scored between 77 and 105 on the various memory indexes (with a score of 105 on the visual memory index). On the second administration, her general memory score was 79 and the various index scores ranged from 73 and 96, with a score of 94 on the visual memory index. The general memory scores of both individuals are within one standard deviation of the normal mean on testing prior to the evaluation, although MMD’s score on the second administration fell just outside this range; their visual memory scores were always within normal range.\(^3\)

**Spoken language comprehension & production.** Single word comprehension was evaluated by means of an auditory word/picture verification task. In this task, the examiner said a word and presented a picture and then asked the individual to respond yes/no if the picture and word matched (line drawings from the Snodgrass & Vanderwart, 1980, picture set were used). Foils consisted of semantically or phonologically similar words. MMD scored 95\% correct (247/260) and RSB scored 100\% correct (260/260). This excellent single word comprehension performance is consistent with their good comprehension skills in conversational settings.

Both individuals suffered from word-finding difficulties in spontaneous speech, with RSB exhibiting greater difficulties than MMD. Confrontation naming was assessed using the Snodgrass and Vanderwort (1980) picture set. MMD named 233 out of 253 (92\%) pictures correctly. Her errors consisted of: 1\% visual errors (cigar → “crayon”), 1\% don’t know responses, 2\% phonological errors (raccoon → /r E k u n/), 2\% semantic errors (caterpillar → “bird”), 2\% morphological errors (carrot → “carrots”), 1\% circumlocutions, and 1\% other errors. RSB correctly named 75\% (192/255) of the pictures correctly. His errors consisted of: 1\% semantic errors, .4\% morphological errors, 4\% phonologically similar word errors (fox → “fat”), 2\% don’t know responses, and 18\% phonologically similar nonword errors (goblet → /g a b l E k/).

In sum, MMD and RSB exhibited excellent auditory single word comprehension but both suffered mild to moderate naming difficulties, with RSB’s anomia being somewhat more severe than MMD’s.

**Spelling.** Both subjects were administered the Length List of the JHU Dysgraphia Battery (Goodman & Caramazza, 1985). MMD wrote to dictation 51\% (36/70) of the words correctly. Of her errors, 44\% were PPEs (e.g., “igloo” → EGLU), 3\% were other words (“province” → PROVEN), and the remaining errors were phonologically implausible nonwords (e.g., “pirate” → PRIDAT). RSB’s accuracy in written spelling to dictation was 47\% (33/70) on this same list. Of his errors, 81\% were phonologically implausible nonwords involving letter deletions, substitutions, and additions (“poem” → PO_M) and 19\% were PPEs (“colour” → COLER). Of the PPEs, all but two resulted from a single letter substitution or deletion.

The Direct Copy Transcoding Task of the JHU Dysgraphia Battery (Goodman & Caramazza, 1985) was administered to evaluate the integrity of the peripheral components involved in written spelling. In this task subjects are presented with words

---

\(^3\)Visual memory might be considered a more appropriate index of general memory abilities as performance on the subtests included in this measure are unlikely to be affected by MMD and RSB’s spoken production difficulties.
and pseudowords printed in lower case letters and asked to ‘transcode’ (copy) them using upper case letters. Both subjects scored 61 out of 62 (98%) on this task, indicating intact abilities for retrieving and producing letter shapes.

In sum, both individuals suffered moderate difficulties in the written spelling of words to dictation that could not be attributed either to comprehension difficulties or to a peripheral breakdown in the retrieval and production of letter shapes.

Localization of the spelling deficits

MMD’s pattern of spelling performance was clearly consistent with the characteristics of an OOL deficit, as described in the Introduction: (1) intact phonological input lexicon and semantic system: indicated by her excellent single word comprehension; (2) a lexical frequency effect: high-frequency words were spelled more accurately (71% correct, 25/35) than low-frequency words (41% correct, 14/35) ($\chi^2 = 5.8, p < .02$); (3) an absence of word length effect: four-letter words were spelled as accurately (57% correct, 8/14) as seven-letter words (57% correct, 8/14); (4) a considerable number of PPEs (e.g., ‘copy’ → COPPIE).

RSB’s spelling performance was consistent with the characteristics of a graphemic buffer deficit. This conclusion is supported by: (1) word length effects: four-letter words spelled more accurately (79% correct, 11/14) than seven-letter words (14% correct, 2/14) ($\chi^2 = 9.2, p < .002$); (2) error types consistent with a graphemic buffer deficit: largely letter deletions and substitutions (‘future’ → FUTRE) and few PPEs (see footnote 2); (3) no significant effects of word frequency on a number of lists developed to contrast high- and low-frequency words, although a significant difference was noted for one list that contrasted words from the very extremes of the frequency range.

In summary, MMD and RSB have very similar profiles in terms of memory and spoken language comprehension and production abilities. They are also similar with regard to the overall severity of their spelling impairments, although the details of their spelling performance reveal that they suffer from damage to different components of the spelling process. The similarity in their overall profiles makes it particularly useful to examine their reactions to an identical treatment protocol. In this situation, any differences in treatment effectiveness can be more readily attributed to the characteristics and differential responsiveness of the affected cognitive components, rather than to other cognitive factors.

TREATMENT INVESTIGATION

The treatment study was designed to evaluate the following: (1) word-specific treatment effects, (2) generalised learning effects, and (3) word-specific repetition effects. In order to do so, three sets of matched words were developed: Treated, Repeated, and Control words. Performance was evaluated during three phases: pre-treatment baseline, treatment, and follow-up. During the pre-treatment baseline phase all three word sets were

---

4 As indicated earlier, MMD also produced errors that were not phonologically plausible. We assume that these originate from an additional deficit to the PG conversion system. A PG system deficit is independently supported by her difficulty in nonword spelling, where accuracy was 22%. Nonetheless, the PG system was sufficiently intact to yield PPEs on a considerable proportion of trials.

5 A pure graphemic buffer deficit is assumed to lack any effect of frequency. In RSB’s case the small and usually nonsignificant effects of frequency may be an indication of an additional and very mild lexical deficit or they may reflect our, as yet, less than full understanding of the graphemic buffer and the ways in which it may be damaged.
presented; during the treatment phase both Repeated and Treated words were presented; and at the end of the treatment phase and during the post-treatment follow-up phase all three sets were presented again.

During the treatment phase the Treated words underwent the remediation procedure (described later) while the Repeated words were simply tested at each remediation session. In this way the Repeated words serve the very important function of indexing the degree of improvement that arises simply from the repeated testing of items as well as the passage of time. For these reasons, the repeated words provide an ideal baseline from which to measure the effectiveness of the treatment.6

Given these stimulus types and the evaluation structure, establishing the effects of interest requires the following: (1) word-specific treatment effects: better spelling performance on Treated words at the end of treatment than at the beginning, as well as better spelling performance on Treated versus Repeated words at the end of treatment; (2) generalised learning effects: better spelling performance on the Control words at the end of treatment than at the beginning; (3) word-specific repetition effects: better spelling performance on the Repeated words at the end of treatment than at the beginning and either (a) no generalised learning effect, or (b) an effect of repetition beyond that of generalised learning (spelling performance with the Repeated words better than with the Control words at the end of treatment). This is because a repetition effect can only be distinguished from a generalised learning effect if the benefit to repeated words is greater than any benefits that apply to all words.

Stimuli

Three sets of words (n = 30 for each set) matched for word frequency and length were developed for each subject. Lists ranged in mean length from 5.9-6.1 letters and in mean frequency from 30.0 to 34.5 (frequency from Francis & Kucera, 1982). Stimuli were uninflected, with each list containing only 0–2 derived forms (e.g., ‘florist’ or ‘absence’) and consisting of 60–90% nouns (the remaining items were verbs or adjectives).

Treatment plan and evaluation procedures

Phase 1: Pre-treatment baseline. Spelling performance on the Treated and Repeated lists was evaluated in two pre-treatment baseline sessions. Performance on the Control list was evaluated in one pre-treatment session. During this phase, each word was simply read aloud by the experimenter and the subject repeated the word and then attempted to spell it, without receiving any feedback.

Throughout the experiment, and for all conditions, spelling accuracy was evaluated by the number of target letters spelled correctly. Each target letter was assigned a score of 0, .5, or 1. A score of 1 was assigned if the target letter was present and in the correct position; .5 if it was present but in the incorrect position, and 0 if the target letter was absent. Using this scoring method, percent correct equals the total score/total number of letters in all of the stimulus words. Responses were also scored according to whether the word was correct or incorrect. The results obtained by considering letter or word accuracy

---

6In fact, the use of a repeated condition allows one to study the effectiveness of treatment even in individuals who are not stable in pre-treatment baseline testing. It does so because effectiveness is measured as the difference between treated and repeated conditions so that even if error rates are dropping in the repeated condition, the effectiveness of treatment can still be determined.
revealed essentially the same pattern, although we prefer letter accuracy as it is a more sensitive treatment measure. Using this measure, we report the results in terms of percent letter errors: 100%–percent correct letters.

**Phase 2: Treatment.** The Treated words and the Repeated words were administered in a blocked fashion at each session, with the order of blocks alternating from session to session (session 1: treated block then repeated block; session 2: repeated block then treated block; etc.). The Control list was administered during the final treatment evaluation sessions. Both MMD and RSB typically underwent two treatment sessions per week. They were both instructed not to practise the words outside the treatment sessions.

Items from the Repeated and Control lists were administered for written spelling to dictation as in the pre-treatment baseline stage. Treated items were presented in the following manner: (1) The subject heard a word, repeated it, and attempted to spell it correctly. (2) After the initial response, regardless of its accuracy, the subject was shown the correct spelling of the word on a note card while the experimenter said aloud each of the letters of the word. The subject was instructed to study the correct spelling. There was no time limit on viewing the note card. (3) If the subject’s initial attempt at spelling the word had been correct this step was omitted; otherwise, after the note card was removed, he/she was given another chance to spell the word correctly. This procedure was repeated until the subject correctly spelled the word.

Treatment sessions were discontinued when subjects reached the treatment goal of stable performance with fewer than 5% letter errors on the treated words.

**Phase 3: Follow-up.** Follow-up evaluations involved the administration of all items for spelling to dictation. MMD was evaluated at 12 and 20 weeks subsequent to the end of treatment, RSB at +8, +12, and +20 weeks.

**Results**

The results of the pre-treatment, treatment, and follow-up phases are reported in Figures 2 and 3. Results are reported in the figures in terms of percent errors for each treatment/testing session. MMD underwent a total of 25 treatment sessions, with a gap of 13 weeks between sessions 11 and 12. RSB underwent a total of 16 treatment sessions.

**Pre-treatment baseline results.** Pre-treatment baseline testing revealed no differences in error rates among the three list types for either MMD or RSB (MMD: $\chi^2 = 1.5, p > .5$; RSB: $\chi^2 = 0.33, p > .8$). In addition, there were no differences across the two pre-treatment testing sessions for either the Treated or the Repeated lists (MMD: treated $\chi^2 = 0.85, p > .3$, repeated $\chi^2 = 0.25, p > .6$; RSB: treated $\chi^2 = 0.32, p > .5$, repeated $\chi^2 = 0$, ns).

Thus, although the pre-treatment baseline testing period was somewhat reduced, it revealed no evidence of a general amelioration of their spelling abilities. This is unsurprising given the amount of time elapsed since the CVAs (MMD 2½ years, RSB 4 years) and the fact that during the time period between their CVAs and the onset of this particular investigation, both MMD and RSB had undergone extensive testing of their spelling abilities in their role as participants in research projects which involved no

---

7 Due to space limitations we will not discuss the intermediate results of the remediation at Session 11 nor the changes that took place between Sessions 11 and 12.
Figure 2. MMD’s percentage of letter errors for Treated, Repeated, and Control words across testing/treatment sessions; b1 and b2 = baseline testing sessions 1 and 2; I-1 and I-2 = initial evaluation sessions 1 and 2; F-1 and F-2 = final evaluation sessions 1 and 2; +12 = follow-up testing 12 weeks after final evaluation.
RSB: Treatment Effects

Figure 3. RSB’s percentage of letter errors for Treated, Repeated, and Control words across testing/treatment sessions; b1 and b2 = baseline testing sessions 1 and 2; I-1 and I-2 = initial evaluation sessions 1 and 2; F-1 and F-2 = final evaluation sessions 1 and 2; +12 = follow-up testing 12 weeks after final evaluation.
remediation (these studies were directed at furthering our understanding of the cognitive processes involved in spelling, such as Folk, Rapp, & Kane, 2000).

**Treatment results.** For the purpose of comparing error rates from the initiation of treatment with error rates at the end of treatment and with subsequent follow-up evaluations, the data from pairs of sessions were combined into single evaluation scores. This was done for two reasons. First, because (as indicated earlier) the order of list presentation (Treated vs Repeated words) alternated from session to session, by combining the data from two sessions we were able to consider data collected with a balanced presentation order (AB–BA). Second, by combining data from two sessions we could compare more stable performance measures, thus avoiding the danger of capitalising on, or being misled by, session to session fluctuation. Thus, when we refer to the initial evaluation we are referring to the combined results of initial treatment sessions 1 and 2 (referred to as I1 and I2 in the figures) and when we refer to the final evaluation we are considering the results of the two final treatment sessions (F1 and F2 in the figures). For the sake of consistency across measures, the pairs of follow-up evaluations are also combined into single scores.

(1) **Word-specific treatment effects:** MMD’s percentage of letter errors on the Treated items dropped significantly from the initial to the final evaluation from 15% (54/366) to 1% (3/366). In other words, MMD went from producing 54 letter errors out of the 366 attempted letters to 3 letter errors out of 366 attempted letters. RSB’s error rate also dropped significantly from 14% (40/362) to 3% (9.5/362) (MMD: $\chi^2 = 45.0, p < .0001$; RSB: $\chi^2 = 25.9, p < .0001$). This improvement on treated items from the beginning to the end of treatment is also revealed by highly significant results of regression analyses (MMD: $R^2 = .51; F = 23.7, p < .0001$ and RSB: $R^2 = .61; F = 21.6, p < .0001$).

Additionally and importantly, on the final evaluation their error rates for Treated items were lower than those for Repeated items—for MMD: 1% (4/366) vs 8% (30.5/364) and for RSB: 3% (9.5/362) vs 7% (24.5/362) (MMD: $\chi^2 = 19.9, p < .0001$; RSB: $\chi^2 = 6.1, p < .015$). This finding indicates that the significant improvement on the Treated items from the initial to the final evaluations cannot be reduced to the effects of repeated testing. This, therefore, constitutes the critical evidence of the effectiveness of the treatment. It is important to note that the possibility that there might also be significant effects of repetition would, in no way, invalidate these findings of treatment effectiveness. Given the pattern of results reported here, the finding of significant repetition effects (see later) would simply indicate that both treatment and repetition effects were present.

(2) **Generalised learning effects:** MMD showed no signs of generalized learning, whereas RSB did. Specifically, MMD’s performance on the Control items remained unchanged from pre-testing (21% [38/179] errors) to the final evaluation (19% [67/358] errors) ($\chi^2 = 1.8, p > .18$). In striking contrast, RSB’s accuracy with the Control items significantly improved from a rate of 22% (39/177) errors in pre-testing to 6% (20.5/354) on the final evaluation ($\chi^2 = 28.8, p < .0001$).

(3) **Word-specific repetition effects:** Both MMD and RSB exhibited a significant improvement on Repeated items from the first to the final evaluations. MMD’s rate of errors on Repeated items dropped from 18% (65.5/364) to 8% (30.5/364) ($\chi^2 = 13.9, p < .0001$) and RSB’s dropped from 14% (47.5/362) to 7% (24.5/363) ($\chi^2 = 7.5, p < .006$). This improvement was also confirmed by regression analyses (MMD: $R^2 = .77; F = 76.3, p < .0001$; RSB: $R^2 = .61; F = 21.6, p < .0001$).

However, because RSB exhibited a significant generalised learning effect and because, at the time of the final evaluation, there was no significant difference on his performance with Repeated vs Control items (7% [24.5/363] vs 6% [20.5/354] errors,
\( \chi^2 = 0.15, p > .7 \), his improvement on the Repeated items may have simply resulted from the generalised learning effect rather than from a specific benefit of repetition. In contrast, because MMD did not exhibit a significant generalised learning effect and error rates on Repeated (8%, 30.5/364) vs Control items (22%, 39/179) were significantly different at time of the final evaluation\( \chi^2 = 15.6, p < .0001 \), a word-specific repetition effect is clearly supported.

**Follow-up results.** MMD and RSB exhibited considerably different reactions to the 20-week absence of treatment and/or testing. We considered both the changes that took place from the beginning of treatment to +20 weeks of follow-up and also from the end of the treatment to the +20 week follow-up.

MMD experienced serious decrements in spelling performance from the final evaluation of the treatment period to the final follow-up evaluation 20 weeks later. Error rates increased from 1% (4/366) to 8% (29.5/366) for Treated items and from 8% (30.5/364) to 14% (50.5/364) for the Repeated items. These differences were statistically significant for both Treated items \( (\chi^2 = 17.5, p < .0001) \) and Repeated items \( (\chi^2 = 4.6, p < .03) \). To assess the long-term benefits of the treatment we compared error rates from the initial treatment evaluation to error rates on the +20 week follow-up evaluation. For MMD, the Repeated items showed no significant differences in error rates from the first evaluation (18%, 65.5/364) to the +20 week follow-up (13%, 50.5/364) \( (\chi^2 = 2.4, p > .12) \). In contrast, although we showed that the Treated items suffered significantly from the withdrawal of treatment, the long-term benefits of the treatment were still apparent in a comparison of error rates on the initial treatment evaluation (15%, 54/366) versus +20 weeks (8%, 29.5/366) \( (\chi^2 = 7.5, p < .006) \).

In contrast, RSB showed no significant loss from the end of treatment to the +20 week follow-up for any of the three stimulus types. Specifically error rates changed only ± 1% from the end of treatment to the +20 follow-up (Treated: \( \chi^2 = 0.3, p > .5 \); Repeated: \( \chi^2 = 0.05, p > .8 \); Control: \( \chi^2 = 0.75, p > .3 \)). Not surprisingly, therefore, the significant treatment benefits that he exhibited for the three list types at the end of treatment were still significant in a comparison of error rates from the first evaluation session with error rates from the +20 week follow-up (Treated: \( \chi^2 = 23.0, p < .0001 \); Repeated: \( \chi^2 = 9.6, p < .002 \); Control: \( \chi^2 = 28.8, p < .0001 \)).

**Summary of results**

MMD and RSB’s responses to an identical treatment protocol exhibited certain similarities and a number of differences. Their responses were similar in that both MMD and RSB exhibited significant benefits of treatment and, furthermore, for both individuals the benefits of treatment were still evident 20 weeks after the termination of treatment. The two differed, however, in that: (1) MMD exhibited significant effects of repeated testing, whereas RSB did not; (2) RSB exhibited significant generalised learning effects, whereas MMD did not; (3) MMD showed a significant decrement in performance from the end of treatment to the +20 week follow-up, whereas RSB maintained his gains on all three list types.

**GENERAL DISCUSSION**

We found that subjects with deficits affecting either the orthographic output lexicon (OOL) or the graphemic buffer benefited significantly from a treatment involving the repeated study and delayed copy of target words. These findings are consistent with the treatment literature (discussed in the Introduction) indicating the effectiveness of this
general class of treatments for deficits affecting the OOL. The fact that MMD (with an OOL deficit) exhibited no generalised learning effect whereas RSB (with a graphemic buffer deficit) did, is also consistent with the general assumption that a successful treatment of OOL deficits is likely to yield only word-specific benefits, whereas a successful treatment of graphemic buffer deficits should generalise to all words.

Mechanisms of recovery

The type of treatment we administered is assumed to be beneficial to the OOL because the repeated exposure to and activation of the target words is thought to strengthen their representations in long-term memory. In this regard it is worthwhile to point out that with the treatment that we adopted, the subjects’ experiences with the Treated words differed from Repeated words in two ways: (1) Treated words were presented visually for study and (2) Treated words were allowed a greater number of spelling attempts because the treatment was repeated on each trial until the word was correctly spelled. Given these differences, we might ask: Is the mere visual exposure to the words sufficient for successful treatment? Are repeated attempts to spell also required? Are both exposure and attempts to spell necessary?

The fact that MMD exhibited significant effects of repeated testing (without exposure to the correct spellings) indicates that exposure to the correct spelling is not required and that repeated attempts to activate the representations are beneficial in and of themselves. However, we should also note that MMD’s improvement on Repeated items was not as long-lasting as that for Treated items. Again, this difference between Treated and Repeated items may be attributable either to the additional benefits of the exposure to and study of the correct forms or, alternatively, to the increased spelling attempts afforded the Treated items. On the basis of our findings we can, therefore, tentatively conclude that the repeated attempts to activate a word’s representation in the OOL are beneficial in strengthening a word’s representation and that the study of the visually presented form of the word may provide additional benefits.

With regard to the mechanisms supporting the successful treatment of graphemic buffer deficits, it is not immediately obvious why repeated exposure and delayed copy should provide a benefit to this type of cognitive component. Nonetheless, it is certainly possible that the capacity of the buffer to maintain the activation of representations was indeed improved by the treatment. Another possibility, however, is that other processes that are specifically beneficial for overcoming a graphemic buffer deficit were strengthened. For example, scanning speed (within the buffer) or speed of transfer to letter-shape conversion processes might have been increased so that letters were processed more quickly within the defective buffer. Finally, it is possible that training simply strengthened the orthographic representations making them more resistant to graphemic buffer damage. This seems unlikely to be the only explanation, however, as the benefits generalised to words that were not explicitly strengthened during treatment. More plausible is the notion that there was both a strengthening of individual word representations and some general benefit (either direct or indirect) to the buffering process. This would explain why RSB exhibited both a generalised learning effect and a significant advantage (albeit statistically marginal) for the Treated vs the Control words at the end of the treatment (RSB: $\chi^2 = 3.74, p < .054$). If only the buffering function had been ameliorated, there should not have been an item-specific effect above and beyond

---

8 We would like to thank Connie Tompkins for bringing this to our attention.

9 Beeson and Rapcsak (2002) discuss the possibility that mere visual study may be sufficient (see also Basso, Burgio, & Prandoni, 1999, for relevant findings from neurologically intact subjects).
the generalised learning effect. Interestingly Hillis and Caramazza (1987) also obtained both word-specific and generalised benefits with the graphemic buffer case that they successfully treated.

Clearly, therefore, a great many questions remain concerning, among other things: the specific aspects of treatment that are critical for recovery of function; the cognitive mechanisms underlying the successful treatment; the reasons for the observed differences between MMD and RSB with regard to the long-term stability of the outcomes, etc. However, we hope that by comparing treatment effects in individuals with well-defined deficits, and generally comparable cognitive profiles, we have made some contribution to our understanding of the specific components of the spelling process and the characteristics of their response to treatment.

REFERENCES


