



Published in final edited form as:

Neuropsychologia. 2019 September ; 132: 107126. doi:10.1016/j.neuropsychologia.2019.107126.

Re-learning and remembering in the lesioned brain

Brenda Rapp^{1,2,3}, Robert W. Wiley^{1,4}

¹Department of Cognitive Science, Johns Hopkins University

²Department of Psychological and Brain Science, Johns Hopkins University

³Department of Neuroscience, Johns Hopkins University

⁴Department of Psychology, University of North Carolina Greensboro

Abstract

It is well known that re-learning language skills after a brain lesion can be very difficult. However, while learning and memory challenges have been extensively researched in amnesic individuals, very little research attention has been directed at understanding the characteristics of learning and memory that are relevant to recovery and rehabilitation of acquired language impairments. Even in the absence of damage to the medial temporal lobe regions classically associated with learning and memory, these individuals often suffer damage to frontal and other subcortical areas associated with learning and memory that may contribute to the learning challenges they face. Therefore, an understanding of the learning and memory profiles of poststroke language impairments is important for the development and optimization of rehabilitation approaches. In two studies, we examine the degree to which certain basic characteristics of learning and memory, identified in neurotypical individuals, are intact in individuals with poststroke language impairment. We specifically consider fundamental principles regarding the optimal spacing of learning trials that have been shown to reliably operate in neurotypical adults, across a wide range of language domains. We report on two studies that examine whether or not these principles also apply in language re-learning and retention for individuals with acquired deficits in written language production. Study 1 compared distributed vs. clustered training schedules, while Study 2 examined—for the first time in the context of re-learning—the relationship between the spacing of training trials and retention period. This investigation revealed that, despite significant cognitive deficits and brain lesions, remarkably similar principles govern re-learning and retention in the lesioned brain as have been found to apply in neurologically healthy individuals. These results allow us to begin to integrate our understanding of recovery with the broader literature on learning and memory and have implications for the optimal organization of rehabilitation. Specifically, the findings raise questions regarding the traditional compression of rehabilitation within relatively short time windows.

Brenda Rapp: Conceptualization, Methodology, Investigation, Resources, Writing – Original Draft, Supervision, Project Administration, Funding Acquisition. **Robert Wiley:** Methodology, Software, Formal Analysis, Resources, Data Curation, Writing – Original Draft, Visualization.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Re-learning cognitive skills after a brain lesion can be very challenging, even for individuals who are not amnesic and do not suffer from generalized memory impairments. However, remarkably little research has investigated the specific learning and memory challenges faced by these individuals in the course of their recovery and/or rehabilitation. In the research we report on here, we examine the integrity of learning mechanisms in individuals with post-stroke brain lesions suffering from acquired impairments in written language processing. Specifically, we consider one of the most well-researched aspects of learning and memory in neurotypical individuals—the optimal spacing of learning trials. In that research literature, the consistency of findings regarding the optimal spacing of learning trials from numerous studies of adults and children (covering a range of cognitive domains) indicates that specific and fundamental learning principles govern the processes of learning and retention. Here, we consider the extent to which these principles apply in re-learning in individuals who underwent rehabilitation for their language impairments. Understanding the degree to which fundamental principles of learning and memory are preserved/disrupted in this context will advance our understanding of the nature of the underlying challenges faced in the recovery of function subsequent to brain lesion. Additionally, the results of this investigation have translational implications for rehabilitation practice, given that the optimal spacing of learning trials is critical to developing rehabilitation experiences that maximize long-lasting recovery

Neural substrates of learning and memory

A large amount of research has been directed at understanding and characterizing the learning and memory challenges faced by different groups of individuals who have suffered neurological damage. These groups primarily include amnesic individuals who have suffered damage to medial temporal lobe structures well-known to be critical for learning and memory, as well as individuals with neurodegenerative diseases such as Alzheimer's Disease, which are also clearly associated with memory impairments. Furthermore, as it has become increasingly clear that cortical areas (e.g., frontal cortex) and other subcortical regions (e.g., the striatum) also play a key role in effective learning and memory, neuropsychological and neuroimaging studies have begun to investigate the specific roles that these brain regions play in successful learning and memory.

For example, with regard to frontal cortex, it had been noted for quite some time that individuals with prefrontal lesions often suffered memory deficits (e.g., Milner 1962; Shimamura et al., 1990). This was then followed by neuroimaging studies showing that, in healthy individuals, prefrontal activation during encoding of to-be-learned materials predicts later performance on tests of recognition and recall (e.g., Wagner et al., 1998). This body of research has led to a clearer understanding of the critical role that frontal cortex and its specific subregions (e.g., ventral and dorsal lateral prefrontal cortex: BA 44, 45, 47/12 and BA 9, 46 respectively) play during both the encoding and retrieval stages of learning and memory.

With regard to subcortical areas beyond the classical medial temporal lobe structures such as the hippocampus, there is considerable research investigating the roles of the basal ganglia

and components of the striatum (McDonald & White, 1993; Packard & Knowlton, 2002) in learning and memory. For example, dorsal striatal structures such as the caudate nucleus have been especially linked to non-declarative aspects of learning and memory (Poldrack et al., 2001) while the ventral striatum has been more consistently associated with the reward and reinforcement aspects of learning. Research shows a close coordination between the striatum and the medial temporal lobes as well as between subcortical and cortical components in long-term memory.

Understanding the precise contributions of these areas, the stages and time-frames within which they operate, and how they are integrated with one another is critical for understanding the complex system that supports learning and memory processing (e.g., McClelland et al., 1995). Given the extent of research on these issues, it is somewhat surprising that findings and paradigms on the behavioral and neural aspects of learning and memory have been scarcely brought to bear on the learning challenges faced by individuals suffering post-stroke language and other cognitive deficits. Specifically regarding language deficits, the relevant prefrontal regions and striatal regions are often lesioned in individuals with acquired language impairments. This makes it especially important to develop an in-depth understanding of the nature of the learning and memory challenges to be overcome for the recovery of language functions. This study focuses on behavioral aspects of re-learning, specifically examining the issue of the optimal spacing of learning trials in a group of 23 individuals with post-stroke dysgraphia, the majority of whom had lesions affecting frontal and/or striatal structures associated with learning and memory, while sparing the medial temporal lobes. It is worth noting that the focus of the research reported here is on understanding behavioral aspects regarding principles of optimal spacing and that the many important issues regarding the neural substrates of specific language re-learning challenges fall outside the scope of the current paper.

The optimal spacing of learning trials

Learning typically involves the repetition of to-be-learned material over multiple trials. Quite naturally, this has led learning and memory researchers to the following fundamental question: How can we organize learning trials to optimize learning and retention? Much of the extensive research on the spacing of learning trials with neurotypical adults has examined either: (a) the effectiveness of spacing schedules that differ in the time between learning trials (i.e., the interstudy interval—ISI) or (b) the relationship between spacing schedules and the time that learned materials need to be remembered (i.e., the retention interval—RI). Regarding spacing schedules, dozens of studies with neurotypical adults have consistently shown that distributing learning trials across a training period (or across a training session) produces superior learning compared to massing them within a training period (or a training session) (for a comprehensive review see Cepeda et al., 2006). Regarding the relationship between the optimal spacing of learning trials and retention intervals, it has been determined that optimal ISI's increase with RI (Cepeda et al., 2008, 2006). In other words, the longer some material needs to be remembered, the longer the optimal intervals between learning trials. For example, as seen in Figure 1, for a retention interval of a week, optimal spacing between learning trials may be 1 day, whereas for a retention interval of a year, optimal spacing may be several weeks. These findings are clearly

contrary to most educational (and rehabilitation) practices, where materials are repeatedly studied with short inter-study intervals over a short period of time, despite the goal of long-term retention (Cepeda et al., 2008).

Here we were specifically concerned with acquired language disorders where issues concerning the optimal spacing of learning have received little attention. However, in language rehabilitation research, an issue that *has* received attention is the possible benefit of high vs. low “intensity” training. When intensity is defined as the amount of training per unit of time (e.g., sessions per week), intensity is generally comparable to spacing. Unfortunately, as Dignam et al., (2016) noted, language rehabilitation studies examining intensity have largely confounded intensity with dosage (total amount of therapy). Thus, studies contrasting different intensity schedules have not usually been equated for total dosage, with the result that high intensity conditions have typically involved more training time. Largely on the basis of these studies, high-intensity treatment schedules are proposed to be superior. However, according to Dignam et al., (2016), the existing dosage-controlled studies of language therapy do not support this conclusion. For example, Sage et al., (2011) compared language therapy delivered for 5 sessions per week for a duration of 2 weeks to therapy delivered 2 sessions per week for a duration of 5 weeks. While their focus was on intensity, the intensity manipulation was implemented via spacing differences. Neither Sage et al., (2011) nor the few other dosage-controlled language studies found an advantage in the short or long term for shorter-interval (more intense, massed) spacing (Martins et al., 2013; Ramsberger & Marie, 2007; Raymer et al., 2006). Instead, they found that more distributed spacing produced superior long-term retention (Sage et al., 2011; Dignam, et al., 2015). Middleton et al., (2016) similarly found superiority of distributed versus massed spacing when training schedules were manipulated *within* a single training session.

The current investigation

This small set of studies, showing superiority for distributed over massed spacing schedules, suggests that similar principles of learning and retention may operate in lesioned and healthy brains. Here, we report on two studies that further examine this issue in individuals with acquired impairments in written word production (spelling) who underwent spelling rehabilitation. Study 1 compared distributed vs. clustered training schedules, while Study 2 examined—for the first time in the context of re-learning—the relationship between the spacing of training trials (ISI) and retention period (RI). These studies allow us to determine if the learning principles observed in neurotypical individuals are also operational in post-stroke recovery. Each participant trained on an individualized word set with each word trained on multiple sessions, with variable time delays between sessions. For each training word, a learning trial involved both testing and training, following a spell-study-spell approach used in previous research (Beeson, 1999; Rapp & Kane, 2002). The training approach included features shown to promote learning, namely, retrieval practice and errorful learning with feedback (Middleton et al., 2016; Pashler et al., 1991). The experimental protocol allowed for both assessment of learning rate across sessions as well as a comparison of accuracy changes from pre to post treatment and at 3-month follow-up.

Study 1 compared two dosage-matched training schedules—Distributed and Clustered. The latter corresponded to a hybrid schedule with massed practice within training sessions (Figure 2A). This allowed for the comparison of “bursts” of intense, massed within-session training with more regular, but less intense, repetition. Unlike most previous studies that employed across-subject designs, we used a within subject-design and, also, rather than a crossover design we simultaneously evaluated both training schedules, assigning different word sets to the different schedules. Study 2 examined the relationship between ISI and RI for words trained in a distributed schedule (for an average of 28 sessions across an average of 19 weeks). In the training protocol, each training trial included both test and study of a training word. In other words, on each training trial, participants were asked first to try to spell the word (test) and then to practice (study) the word. Because each trial provided test data and served as a learning trial, the approach allowed the time period between learning trials to serve both as an ISI (the number of days between learning trials) and also as an RI (the number of days between learning and test). For example, as illustrated in Figure 2B, the test results on the 3rd learning trial (day 7) for PEN would provide data regarding the effect of an ISI of 2 days for a RI of 4 days. In addition, the test results for PEN on day 10 would provide data regarding the effect of an ISI of 4 days for an RI of 3 days. Laboratory-based studies evaluating the relationship between ISI and RI intervals typically explicitly manipulate these variables across trial triads (two learning trials and one test trial). Instead, in this study, treatment was delivered in a manner similar to a clinical setting and the natural variability in the number of days between training trials (due to weekends, illness, etc.) allowed for the evaluation of many different ISI/RI combinations.

In addition to the manipulation of spacing and retention intervals, we evaluated all participants on language, cognitive and working memory and learning tasks before and after treatment. This allowed us to evaluate their memory and learning skills and also to assess if training effects were selective to the trained materials or generalized to untrained words or other cognitive domains. Overall, this investigation allowed us to determine if the spacing principles that govern re-learning of verbal materials in a brain-injured population are comparable to those that govern neurotypical learning and memory.

2. Material and methods

2.1 Participants

Twenty-three individuals with chronic (> 1 year) impairments in written language production (spelling) subsequent to a single left-hemisphere stroke were enrolled. Eleven individuals participated in Study 1; these participants were also included in Study 2 along with an additional 12 participants (n=23). Participants had no other neurological disease or history of developmental dyslexia/dysgraphia. Each individual participated in bi-weekly behavioral treatment sessions over an average of 19 weeks. See Table 1 for relevant information regarding demographics, lesion size and spelling deficit severity¹. Consent was

¹Note that although RFZ and RHH scored 100% on JHU Dysgraphia Battery, these two individuals complained that their post-stroke spelling was significantly inferior to their very high pre-stroke spelling abilities. Their diminished post-stroke spelling performance was confirmed by errors committed on low frequency words that had been spelled correctly in their pre-stroke writing samples.

obtained using procedures consistent with the Declaration of Helsinki and the Johns Hopkins University Institutional Review Board.

2.2 Lesion distribution

Structural MPRAGE scanning was carried out with the following characteristics: 176 sagittal slices, multishot, turbo field echo pulse sequence, slice thickness = 1mm, and in-plane resolution of 1 x 1mm². Each structural scan was aligned to the AC-PC plane. Lesion masks were manually drawn using MRIcron (Rorden & Brett, 2000) in voxels with hypointense T1 signal; the T1 images were then normalized to standard Montreal Neurological Institute (MNI) template space using the enantiomorphic method (Nachev et al., 2008). Normalization parameters were then applied to the lesion masks in order to warp them to standard MNI space. Figure 3 depicts the overlap of the lesions of the 23 individuals, where it can be seen that most commonly lesioned areas include left posterior frontal, anterior parietal and superior temporal regions. In fact, 15/23 of the individuals suffered damage that affected the dorsal and/or ventral frontal areas that have been associated with learning and memory (as reviewed in the Introduction). With regard to subcortical regions that have been associated with learning and memory, the structural scans were closely examined to determine if the lesions affected voxels in these regions². This examination revealed that the bilateral hippocampus, amygdala and ventral striatum were intact in all of the participants. With regard to the dorsal striatum, the caudate nucleus, putamen and globus pallidus were examined. It was found that in 10 of 23 of the participants all three structures appeared to be entirely intact bilaterally, while in the remaining individuals there was at least some degree of damage. Note that we did not specifically quantify the extent of damage to these structures in each participant, rather simply recorded if damage was present or absent.

2.3 Assessments: Spelling, language, learning/memory

An individualized training word set (n=40) was developed for each participant, consisting of words with 25 - 80% letter accuracy on two baseline tests, ensuring that all participants were training within a generally comparable difficulty range. Learning effects were evaluated both in terms of rate of improvement and overall improvement. Therefore, there were two assessments of the training items: (1) to evaluate the rate of learning during the treatment period (see "Treatment methods" for more details), the spelling accuracy of each training word was evaluated on each of its training trials; (2) to evaluate the overall effects of spelling treatment, accuracy on the training word set was assessed at three time-points: before and after treatment and at 3 months follow-up. To quantify spelling performance, letter accuracy (rather than word accuracy) was used such that the accuracy of each letter in target spellings was scored (e.g., the misspelling of BOAT as BOET has a 75% letter accuracy while a misspelling such as BUAD has a 25% letter accuracy).

Additionally, multiple language assessments as well as tests of learning and memory were administered to: (a) evaluate working memory and short-term learning skills and (b)

²These structures were specifically examined for presence/absence of lesion and we did not compare volumes with control participants. Therefore, we cannot report whether any of these regions suffered atrophy/volume reduction.

determine the selectivity of the treatment by evaluating if treatment affected only the spelling of trained words, or if it also generalized to the spelling of untrained words or to other language/cognitive functions. Five language assessments consisted of: oral reading (PALPA 35, Kay et al., 1996), reading comprehension (PALPA 51, Kay et al., 1996), single word auditory comprehension (Northwestern Naming Battery, Thompson et al., 2012), semantics/conceptual processing (Pyramids and Palm Trees; Howard & Patterson 1992) and spoken word production (Northwestern Naming Battery-global naming score; Thompson et al., 2012) (see Table 2 for pre and post-treatment scores). With regarding to learning and memory testing, working memory was assessed with the non-verbal Corsi blocks task (Kessels et al., 2000) rather than verbal working memory (digit span), since the spoken language deficits suffered by many of the participants made it difficult to obtain an estimate of verbal working memory capacity uncontaminated by word production difficulties. Non-verbal learning was assessed with two tasks—Cambridge Memory Test for Faces (CMTF; Duchaine & Nakayama 2006) and Doors and People (Baddeley et al., 1994). The CMTF evaluates face learning by asking participants to memorize 6 faces and then to immediately identify the faces in a 3-alternative forced choice task. Doors and People evaluates visual object learning by presenting 12 pictures of doors for 3 seconds each, followed by a 4 alternative forced choice test. This procedure is then repeated for a 2nd set. (See Table 1 for scores on these learning and memory tests.)

2.4 Treatment methods

A spell-study-spell technique (Beeson, 1999; Rapp & Kane, 2002) was administered during sessions lasting approximately 90 min, typically 2x/week for an average of 28 sessions per individual. Not all words were presented on every training session; the number of trained items on any given session was influenced by the participant's working speed, accuracy, and fatigue, and therefore changed over the course of the treatment period. Treatment ended with greater than 90% accuracy on training items for two consecutive sessions or failure to improve after six sessions. Each learning trial was structured as follows: (1) the individual heard a target word, repeated it, and attempted to write the spelling (test). (2) Regardless of accuracy, the individual was shown the correct spelling while the experimenter said aloud the word's letters. The individual copied the word once (study). (3) If the word was spelled correctly at Step 1, then Step 3 was omitted and the experimenter continued to the next item; otherwise, the word was removed from view and the individual was asked to spell it. Steps 2 and 3 were repeated until the word was spelled correctly or for a maximum of 3 times before moving to the next item. This simple training protocol is designed to include key elements that have been shown to promote learning: Step 1 provided an opportunity for retrieval practice, and Steps 1 and 2 allowed for errorful learning with feedback (Kornell et al., 2009; Pashler et al., 2007; Roediger & Karpicke, 2006; Middleton et al., 2016).

2.5 General behavioral results

Table 2 presents an overview of the results of the pre and post training assessments of spelling, language and other cognitive skills. It shows impaired performance on spelling, reading and naming. Critically, the results show that performance on all tasks was stable across the training period, with the exception of significant improvement in spelling performance on both trained and untrained words, even with correction for multiple

comparisons. These findings clearly indicate that the treatment selectively affected spelling skills.

With regard to performance on learning and memory tasks, the key interest was in understanding how performance compared to the neurotypical range. In terms of working memory, the group as a whole performed at the 25th percentile on the Corsi Blocks task, with only three participants performing differently from control participants ($p < 0.05$; controls from Kessels et al., 2000; $n = 70$, age 18-72 years). In terms of the two learning tasks, on the CMFT the performance of the group as a whole did not differ from controls ($p > 0.05$; controls from Tirta Susilo (personal communication); $n = 41$, mean age = 69 years), with only two individuals exhibiting below normal performance. Similarly, on the Doors Test, the group overall performed at the 25th percentile (norms from the Doors and People manual, Baddeley et al., 1994), with half of the participants obtaining scores at the 50th percentile or above, and only 3 scoring lower than the 10th percentile.

2.6 Data analysis: Study 1³

The goal of the study was to evaluate the effectiveness of Distributed compared to Clustered (a form of massed) spacing of training trials, defined across sessions. Training schedule was manipulated within participants ($n=11$), with each receiving training on two subsets of items, with the sets matched on pre-treatment accuracy, letter length and word frequency (English Lexicon Project, Balota et al., 2007) to be trained on either Clustered or Distributed Schedules (Figure 2A). In the Clustered Schedule, words were trained for 3-4 trials within a session, every 8 sessions (on average). In the Distributed Schedule, words were trained only once within a session, every 2 sessions (on average). The total number of training trials per word was the same regardless of schedule, ranging from 12–27 depending on the individual. Because each word was tested (Step 1) and trained (Steps 2 and 3) on every trial, each trial constituted both a training trial and a test trial, allowing for an evaluation of the rate of improvement across training sessions.

2.6.1 Analysis 1: The Effect of schedule on rate of improvement—For the analysis of the effect of training schedule on improvement rate, the letter accuracy data from all training trials were evaluated with a generalized linear mixed-effects model (LMEM), fit using R (R Core Team, 2018), package *lme4* (Bates et al., 2015) function *glmer* (binomial family), with p-values obtained by package *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017) and R^2 measures of variance explained obtained by package *MuMin* (Bartón, 2018). Specifically, individual letter accuracy was the dependent variable, with the following independent variables: fixed effects included Word Length, (log) Word Frequency, Session, Schedule (Clustered vs. Distributed), RI (Retention Interval), and Session X Schedule Interaction. The two-way interaction of Session X Schedule was the term of primary interest, as it indexes whether the amount of improvement per session of training differed significantly between items trained in the Clustered versus the Distributed schedules. Additional fixed effects were included for Age, Education (years), and Lesion Volume (mm^3), with each of these entered into three-way interactions with Session X

³A subset of these data were previously reported in Wiley & Rapp (2018)

Schedule as well (e.g., Age X Session X Schedule), to account for the fact that the participants varied along these three dimensions (Table 1), any of which could plausibly affect response to treatment. The Schedule variable was sum-coded (Clustered – 1, Distributed + 1). All continuous variables were standardized (centered on their mean values and divided by their standard deviations).

The initially-specified LMEM included a maximal random effects structure by-participants (Barr et al., 2013) crossed with random intercepts by-items. Multicollinearity was then assessed via Variance Inflation Factors (VIFs), and the random effects structure was reduced by removing random slopes for terms associated with $VIF > 5$ (Sheather 2009, page 203). The resulting LMEM had all $VIF < 5$, and comparison of the model with the full random effects structure to that with the reduced structure via a likelihood ratio test (Kuznetsova, Brockhoff, & Christensen, 2017) indicated no significant change in model fit. Additionally, the reduced random effects structure did not result in any changes in the significance of the fixed effects estimates. The full model specification and parameter estimates are available in Appendix A (Table A1).

2.6.2 Analysis 2: The Effect of schedule on accuracy at pre, post and follow-up time-points—Data from the Pre, Post, and Follow-up assessments were evaluated with a generalized LMEM analysis, similar to that used in Analysis 1. The only difference was that there was no RI variable, and in place of the Session variable there was the categorical variable Time-point, with three levels (Pre, Post, and Follow-up, simple-coded with the Pre Time-point as the reference level). The interaction of Time-point X Schedule allowed for evaluating (1) the difference in the amount of improvement for Clustered versus Distributed items from Pre to Post, and (2) the difference in retention between the two Schedules from Pre to Follow-up (i.e., whether a benefit to the Distributed schedule remained 3 months after completing training).

As with Analysis 1, the random effects structure was reduced from the maximal by removing terms with $VIF > 5$, which did not impact the significance of any of the fixed effects estimates; the full model specification is presented in Appendix A (Table A2).

2.7 Data Analysis: Study 2

The goal of this study was to understand if, in the context of a distributed schedule, the optimal interval for repeated study changes as a function of the duration of the retention interval. To do so, Study 2 examined only data from Distributed spacing schedules and included both the data from the Distributed Schedule items for Study 1 participants (20 items each, average of 11.6 trials each) as well as from an additional set of 12 participants (40 items each, average of 11.0 trials each) who were trained only on a Distributed Schedule. As described in the Introduction and depicted in Figure 2B, the intervals between successive trials could serve both as an ISI or RI, providing considerable variability in the combinations of ISI x RI. See Figure 4 for the frequency distribution of spacing intervals between training trials for all the training words across the set of 23 participants.

2.7.1 Analysis 3: The Relationship between study and retention intervals—The same LMEM analysis approach as used for the Study 1 (both Analyses 1 and 2) was

applied to Study 2, except only the data from Distributed training schedules was included ($n = 23$ participants). Both the fixed and random effects were the same but for the following exceptions: there was no Schedule variable or Session X Schedule interaction. Instead, ISI and the full 3-way interaction Session X ISI X RI were included. In addition, the fixed effects for Age, Education, and Lesion Volume were included with 4-way interactions (e.g., Age X Session X ISI X RI), however they were not entered into simultaneous regression because multicollinearity was found to be too high ($VIFs > 20$). Instead, three models were computed, one with each of the three covariates (e.g., one model included Age X Session X ISI X RI, another included Education X Session X ISI X RI, etc.). The results of those three models revealed that none of the three covariates had any impact on the ISI X RI interaction (i.e., there was no change in significance and only minimal changes in the magnitude of the effect when these covariates were included), thus the results reported are based on a model *without* any of these effects. As with Analyses 1 and 2, the random effects structure was determined by removing random slopes from the maximal structure until all $VIFs < 5$, which did not impact the significance of the fixed effects. The full model specifications including results for the models that did include the covariates Age, Education, and Lesion Volume are reported in Appendix A (Tables A3–A6).

3. Results

3.1 Analysis 1: The Effect of schedule on rate of improvement

The results of the LMEM analysis are presented in Figure 5 (Figure B1 in Appendix B depicts raw data that is plotted by binning the sessions, along with standard errors of the mean). The critical term of interest is the interaction of Session X Schedule, which reveals a significant difference in the rate of improvement across training sessions between the two schedules, Clustered and Distributed ($\beta = 0.205$, $p < 0.001$). This beta is an odds ratio effect size, interpretable as 8.2% improvement in the odds of correctly spelling a letter per training session for the Distributed schedule, versus just 5.0% improvement per training session for the Clustered schedule⁴.

In terms of the covariates (Age, Education, and Lesion Volume), there were significant interactions of Session X Education and Session X Age, revealing that individuals with more years of education improved more rapidly ($\beta = 0.645$, $p < 0.001$) whereas older participants improved less rapidly ($\beta = -0.178$, $p = 0.001$). There was also a significant Lesion Volume X Session X Schedule interaction ($\beta = 0.138$, $p = 0.001$), indicating that participants with larger lesions in fact showed a greater benefit for the Distributed over the Clustered schedule. This 3way interaction was driven by a slower rate of improvement for Clustered items among patients with larger lesions (e.g., rate of improvement with a lesion one standard deviation larger than average $\beta = 0.227$ versus 1.085 for a lesion one standard deviation smaller than average), not by a faster rate of improvement for the Distributed items (with a larger lesion $\beta = 0.915$ versus 1.220 with a smaller lesion). Overall the model had a conditional R^2 of 60% (fixed + random effects) and a marginal $R^2 = 23%$ (fixed effects

⁴After back-transforming the betas from their standardized form; see Methods section 2.6.1.

alone); the beta estimates and p-values for all parameters are reported in Appendix A (Table A1).

3.2 Analysis 2: The effect of schedule on spelling accuracy at pre, post and follow-up time-points

The results of the LMEM analysis are presented in Figure 6 (Figure B2 in Appendix B depicts raw data plotted with standard errors of the mean). First, comparing the improvement from Pre to Post for each Schedule, there was a significant benefit for Distributed versus Clustered items ($\beta = 0.261$, $p < 0.001$). As depicted in Figure 6, this equates to a predicted improvement in accuracy (i.e., when controlling for the other variables included in the LMEM) from 66.8% to 95.3% for Distributed items, versus 66.7% to 92.3% for Clustered items. Second, there was a significant difference between the two Schedules comparing Pre to Follow-up ($\beta = 0.171$, $p = 0.007$): as depicted in Figure 6, the predicted improvement from Pre to *Follow-Up* (3-months post training) was 25.8% for the Distributed Schedule versus 23.7% for the Clustered schedule. A post-hoc LMEM analysis, identical to the Pre-Post-Follow-up model except including only the Follow-up data, was conducted to test the simple effect of Schedule at the Follow-up time point. This analysis revealed that the higher retention at 3 months after the end of the treatment for Distributed items versus Clustered items (raw average: 86.3% versus 82.7%) was significant ($p = 0.04$).

In terms of the covariates, there was both a significant Education X Schedules X Time-point interaction ($\beta = 0.176$, $p = 0.023$) and Lesion Volume X Schedules X Time-point interaction ($\beta = -0.249$, $p = 0.013$). Specifically, participants with more education showed a larger difference in improvement from Pre- to Follow-up for the Distributed versus Clustered schedule: $\beta = 2.590$ for Distributed versus 1.895 for Clustered for an individual with one standard deviation more education than the mean, compared to $\beta = 0.999$ versus 1.009 for an individual with one standard deviation less education than the mean—whereas from Pre- to Follow-up those with larger lesions showed *less* difference between improvement for Distributed versus Clustered items: $\beta = 1.766$ for Distributed versus 1.921 for Clustered for an individual with a lesion one standard deviation larger than the mean, compared to $\beta = 1.823$ versus 0.983 for an individual with a lesion one standard deviation smaller than the mean. Overall the model had a conditional R^2 of 44% and a marginal R^2 of 23%; the beta estimates and p-values for all parameters are reported in Appendix A (Table A2).

4. Study 2: Spacing effects on learning and retention—the relationship between ISI and RI

4.1 Analysis 3: The Relationship between study and retention intervals

The results of the LMEM analysis are presented in Figure 7 (raw data are plotted with standard errors of the mean, binning the ISI's and RI's into short, medium, and long, in Appendix B, Figure B3). Overall, and as expected, both longer ISI's ($\beta = -0.486$, $p < 0.001$) and longer RI's ($\beta = -0.381$, $p = 0.004$) were associated with lower spelling accuracy. Importantly, the interaction of ISI X RI was significant, revealing that the effect of RI on spelling accuracy differed significantly depending on the ISI ($\beta = 0.135$, $p = 0.028$). Specifically, the negative effect of increasing ISI on accuracy diminishes with longer RI,

and, in fact, reverses to a positive effect at very long RIs. This interaction is depicted in Figure 7, where it can be seen that, for example, for a short 2-day RI accuracy is predicted to be very high with a short 1-day ISI ($\approx 98\%$) but lower with a long 40-day ISI ($\approx 92\%$). This contrasts with a long 100-day RI—while, in general, a longer RI is predicted to result in lower accuracy than a shorter RI, performance at long RIs is predicted to actually benefit from a relatively longer ISI—82% accuracy when paired with a long 40-day ISI compared to just 71% when paired with a short 1-day ISI. Intermediate RIs exhibit a transition between these patterns, for example, medium 50-day RIs show little effect of ISI. This pattern of shorter and longer RIs showing contrasting ISI benefits is strikingly similar to that reported by the meta-analysis of Cepeda et al., (2006) reviewed in the Introduction (compare Figures 1 and 7).

As explained in the Methods section (2.7.1), none of the covariates Age, Education, or Lesion Volume interacted significant with the spacing effect (i.e., the ISI x RI interaction), and indeed had negligible impact on the estimated magnitude of the effect (comparing β in model without any of the three covariates = 0.135 to β 's = 0.143, 0.149, and 0.135 when controlling for Age, Education, and Lesion Volume respectively). Overall the model (without covariates) had a condition R^2 of 78% and a marginal R^2 of 22%; the beta estimates and p-values for all parameters are reported in Appendix A (Tables A3–A6 for the model without covariates and with each one of the covariates respectively; see section 2.7.1 for details of the methods).

5. General Discussion

This investigation of the optimal spacing of learning trials in the rehabilitation of post-stroke written language impairments revealed that, despite significant deficits and brain lesions, remarkably similar principles apply to re-learning and retention in the lesioned brain as have been found to apply in neurologically healthy individuals. Specifically, we found: (1) Significant benefits for a distributed compared to more massed (clustered) spacing of learning trials, including both faster learning rates and greater accuracy immediately after training as well as three months later; (2) As the length of time between study and test increases (the retention interval or RI), increasingly longer gaps between prior learning trials (the interstudy interval or ISI) produced better retention; and (3) normal performance on tests of learning and non-verbal working memory for the group as a whole ($n = 23$) and for the vast majority of individual participants, indicating generally preserved learning and memory skills despite significant brain injury. Overall, for rehabilitation focused on re-learning individual items, individuals benefit more from spacing learning trials over the treatment period rather than clustering them together. Further, while spacing of 1 day or 1 week between repeated learning trials may be optimal if the anticipated retention period is a few days or a month, longer retention periods—such as 6 months—will benefit from spacing of as much as one or even two months between learning trials.

Mechanisms underlying spacing effects

While the effects of spacing in learning (and now in *re-learning*) are well-documented, less well understood are the underlying mechanisms of learning and memory that give rise to

these effects. The fundamental facts are that memory traces can grow in strength with learning and diminish in strength due to forgetting, and that forgetting typically follows a negatively accelerating function in which the decay rate is high soon after the learning and then slows down over time (e.g., Wixted, 2004). Considered in this light, the specific learning/re-learning challenge is to determine the timepoint(s) during the forgetting trajectory when it is best to “refresh” the memory trace with additional learning trials. However, making this determination requires understanding, among other things, the manner in which memory traces from multiple learning attempts combine to create long-lasting, accessible representations. On this there is no clear consensus, although a number of computational models point in the same direction (e.g., Staddon et al., 2002; Mozer et al., 2009; Pavlik & Anderson 2005). Each of these models exhibits an interaction between the strength of a previous trace and a new memory trace such that the decay rate of a new memory trace increases with the strength of the previous trace, as long as the previous trace is still accessible. Accordingly, if a second learning trial is presented while the previous memory trace is still strong, then the new memory trace can be weaker than if the second learning trial was presented later in the forgetting trajectory. At a later point, when the initial trace itself is weaker, the new memory trace will decay more slowly. This is generally consistent with other proposals such as “desirable difficulty” (Bjork et al., 2013), according to which optimal learning occurs when retrieval is most appropriately difficult. In this context, the general advantage of distributed versus massed training arises because, in a distributed schedule, repeated learning trials are more likely to occur when the memory trace of the previous learning trial is weaker, resulting in a stronger and longer-lasting representation. A deeper understanding of these principals will ultimately be required to fully optimize the spacing of learning trials in education and rehabilitation.

While the same general principles apply across individuals, there will likely be individual differences in the specific characteristics of the learning and forgetting curves. As Lindsey et al., (2014) show, understanding these differences will be critical for developing personalized, precision teaching that optimizes learning and retention. This is especially important with regard to re-learning, where individual variability is likely to be far greater than in the neurotypical population. Furthermore, in the context of post-lesion learning, it will be critical to understand the potential impact of additional factors, such as the fact that the learning involves previously known targets, possible neurophysiological responses to brain lesions, and the role of concomitant cognitive deficits. Furthermore, it is important to underscore that the results reported here were obtained with individuals in the chronic stage (more than 1 year post lesion) and the possible relationship of spacing and recovery stage needs to be investigated. In terms of the implications of these findings for rehabilitation, given that the primary rehabilitation goal is the retention of information for a long period of time (years), we should space learning trials accordingly. The compression of rehabilitation into a short time period may produce results that are unlikely to achieve the long-term retention that is desired and therapy may be better distributed over a longer period with to-be-learned materials retrained in a more cyclic manner.

Neural bases

The results from this investigation show that at least certain fundamental aspects of learning and memory—such as those that govern the effects of optimal spacing—can be preserved, *even within the skill domain affected by the lesion*. It is important and encouraging to know that these basic mechanisms can be preserved in the face of significant damage to the spelling network. However, this raises two important questions: What neural substrates support these basic mechanisms? and, Why is relearning still so difficult? Neither of these was directly addressed in this work, and so answers are not possible. However, the work may provide some relevant clues to be pursued in future work.

The neural bases of spacing effects in learning have been scarcely investigated (Maddox, 2016), although Callan & Schweighofer (2009) did report evidence associating spacing effects with the left frontal operculum. In a lesioned population, the default assumption regarding “normal learning/memory functions” is that they would be supported by intact substrates that have been traditionally associated with learning/memory. In this group, the most likely candidates are the areas unlesioned in all of the participants: the hippocampus (bilaterally), right hemisphere frontal areas or the right dorsal striatum. One caveat, however, is that although we have shown that basic principles of optimal spacing are at work, we don’t know if the magnitudes of the effects of scheduling and ISI/RI interaction are actually comparable to the magnitudes we would expect in the intact brain. That determination would likely require an evaluation of the effects of these variables in a direct comparison of learning in intact and brain-lesioned populations.

Why is relearning so difficult? Certainly one possibility is that this is due to the damage to learning and memory mechanisms supported by the damaged left ventral and dorsal lateral prefrontal cortex and components of the dorsal striatum suffered by most of the study participants. While these areas may indeed be responsible for the ongoing challenges faced in recovery, it will not be enough to identify the relevant substrates, rather it will be critical to characterize the specific nature of the cognitive difficulties that lesions to these areas produce. For example, it would be important to specify which specific cognitive processes are affected and how they contribute to the effective re-learning and retention of language knowledge. There are, of course, many other possible candidates for explaining the difficulties faced in re-learning. For example, future work can examine the role played by the reduction in representational neural space that has been documented in sensory domains (Medina & Rapp, 2014; Jenkins & Merzenich, 1987) and is a likely consequence of the damage to language processing areas. While representational space is not itself a mechanism of learning and memory, it may determine the difficulty/ease with which knowledge can be neurally re-represented and/or integrated into remaining knowledge structures. In sum, investigating the cognitive and neural bases of the challenges faced in recovery and rehabilitation is certainly a complex undertaking, but it is critical to pursue these issues if we are to advance significantly in our ability to develop optimal, individualized rehabilitation protocols.

Acknowledgements.

We gratefully acknowledge NIH support (DC012283) support for the multi-site project examining the neurobiology of language recovery in aphasia that this work forms a part of. We thank Jennifer Shea and Donna Gotsch for their invaluable contributions to data collection, scoring, and analysis, Jeremy Purcell for his analysis of the integrity of cortical subcortical regions and Michael Mozer for his helpful input regarding computational models of learning.

Appendix A

Table A1.
LMEM of schedule effect on the rate of improvement during training.

Formula: accuracy ~ Length + Frequency + RI + Session*Schedule*(Age + Education + Lesion Volume) + (1 + Length + Frequency + Session| participant) + (1 | item)

	Predictor:	Estimate	Std. Error	z-value	Pr(> z)	
	(Intercept)	2.960	0.404	7.336	<0.001	***
	Word Length	-0.279	0.163	-1.709	0.088	.
	Word Frequency	0.153	0.089	1.713	0.087	.
Main Effects	Session	0.862	0.168	5.143	<0.001	***
	Schedule	0.330	0.043	7.749	<0.001	***
	Age	0.145	0.340	0.426	0.670	
	Education	1.175	0.294	4.000	<0.001	***
	Lesion Volume	-0.220	0.283	-0.777	0.437	
	RI	0.052	0.028	1.895	0.058	.
	<hr/>					
2-Way Interactions	Session:Schedule	0.205	0.036	5.719	<0.001	***
	Session:Age	-0.139	0.171	-0.809	0.418	
	Session:Education	0.645	0.166	3.884	<0.001	***
	Session:Lesion Volume	-0.291	0.161	-1.800	0.072	.
	Schedule:Age	-0.178	0.054	-3.285	0.001	**
	Schedule:Education	-0.093	0.050	-1.863	0.063	.
<hr/>						
3-Way Interactions	Session:Schedule:Age	0.020	0.048	0.416	0.678	
	Session:Schedule:Education	-0.023	0.050	-0.460	0.646	
	Session:Schedule:Lesion Volume	0.138	0.042	3.277	0.001	**

Signif. codes: 0.001

0.01

**
0.05

*
.01

Table A2.
LMEM of schedule effect across Pre, Post, and Follow-Up time points.

Formula: accuracy ~ Length + Frequency + Time-point*Schedule*(Age + Education + Lesion Volume) + (1 + Length + Frequency + Time-point| participant) + (1 | item)

	Predictor:	Estimate	Std. Error	z-value	Pr(> z)	
Main Effects	(Intercept)	1.960	0.282	6.964	<0.001	***
	Word Length	-0.155	0.097	-1.602	0.109	
	Word Frequency	0.069	0.080	0.868	0.385	
	Time Point (Pre vs. Post)	2.024	0.202	10.016	<0.001	***
	Time Point (Pre vs. Follow Up)	1.623	0.245	6.634	<0.001	***
	Schedule	0.125	0.036	3.500	<0.001	***
	Age	0.352	0.241	1.459	0.144	
	Education	0.355	0.139	2.546	0.011	*
	Lesion Volume	0.114	0.203	0.561	0.575	
2-Way Interactions	Time Point (Pre vs. Post):Schedule	0.261	0.066	3.934	<0.001	***
	Time Point (Pre vs. Follow Up):Schedule	0.171	0.063	2.710	0.007	**
	Time Point (Pre vs. Post):Age	-0.214	0.201	1.066	0.286	
	Time Point (Pre vs. Follow Up):Age	-0.175	0.317	-0.553	0.580	
	Time Point (Pre vs. Post):Education	0.435	0.147	2.970	0.003	**
	Time Point (Pre vs. Follow Up):Education	0.619	0.180	3.442	0.001	***
	Time Point (Pre vs. Post):Lesion Volume	0.258	0.190	1.353	0.176	
	Time Point (Pre vs. Follow Up):Lesion Volume	0.220	0.369	0.597	0.551	
	Schedule:Age	-0.009	0.053	-0.162	0.871	
	Schedule:Education	-0.086	0.042	-2.029	0.043	*
	Schedule:Lesion Volume	-0.001	0.053	-0.022	0.982	
3-Way Interactions	Time Point (Pre vs. Post):Schedule:Age	-0.047	0.097	-0.487	0.626	
	Time Point (Pre vs. Follow Up):Schedule:Age	-0.163	0.095	-1.721	0.085	
	Time Point (Pre vs. Post):Schedule:Education	-0.088	0.082	-1.075	0.282	
	Time Point (Pre vs. Follow Up):Schedule:Education	0.176	0.077	2.278	0.023	*
	Time Point (Pre vs. Post):Schedule:Lesion Volume	0.080	0.106	0.751	0.452	
	Time Point (Pre vs. Follow Up):Schedule:Lesion Volume	-0.249	0.100	2.491	0.013	*

Signif. codes: 0.001

0.01

**
0.05

*
.01

Table A3.
LMEM of spacing effect, without additional covariates
for Age, Education, or Lesion Volume.

Formula: accuracy ~ Length + Session*ISI*RI + (1 + Length + Frequency + Session + ISI*RI + Session:ISI + Session:RI | participant) + (1 | item)

Predictor:	Estimate	Std. Error	z-value	Pr(> z)	
(Intercept)	3.236	0.452	7.156	< 0.001	***
Word Length	-0.330	0.121	-2.718	0.007	**
Word Frequency	0.237	0.074	3.187	0.001	**
Session	1.524	0.325	4.685	< 0.001	***
ISI	-0.486	0.124	-3.922	< 0.001	***
RI	-0.381	0.133	-2.876	0.004	**
Session:ISI	-0.240	0.103	-2.333	0.020	*
Session:RI	-0.367	0.109	-3.369	0.001	***
ISI:RI	0.135	0.061	2.198	0.028	*
Session:ISI:RI	-0.011	0.024	-0.479	0.632	

Signif. codes: 0.001

0.01

**

0.05

*

.0.1

Table A4.
LMEM of spacing effect, including covariate of Age.

Formula: accuracy ~ Length + Session*ISI*RI*Age + (1 + Length + Frequency + Session + ISI*RI + Session:ISI + Session:RI | participant) + (1 | item)

Predictor:	Estimate	Std. Error	z-value	Pr(> z)	value
(Intercept)	3.190	0.450	7.086	< 0.001	***
Word Length	-0.338	0.120	-2.811	0.005	**
Word Frequency	0.242	0.074	3.259	0.001	**
Session	1.514	0.324	4.676	< 0.001	***
ISI	-0.496	0.126	-3.921	< 0.001	***
RI	-0.390	0.137	-2.847	0.004	**
Age	-0.416	0.192	-2.165	0.030	*
Session:ISI	-0.258	0.108	-2.396	0.017	*
Session:RI	-0.370	0.114	-3.240	0.001	**
ISI:RI	0.143	0.066	2.157	0.031	*
Session:Age	-0.142	0.093	-1.523	0.128	
ISI:Age	0.047	0.062	0.761	0.447	
RI:Age	0.062	0.067	0.925	0.355	
Session:ISI:RI	-0.012	0.024	-0.523	0.601	

Predictor:	Estimate	Std. Error	z-value	Pr(> z)	value
ISI:RI:Age	-0.020	0.027	-0.742	0.458	

Signif. codes: 0.001

0.01
**
0.05
*
.0.1

Table A5.
LMEM of spacing effect, including covariate of Education.

Formula: accuracy ~ Length + Session*ISI*RI*Education + (1 + Length + Frequency + Session + ISI*RI + Session:ISI + Session:RI | participant) + (1 | item)

Predictor:	Estimate	Std. Error	z-value	Pr(> z)	value
(Intercept)	3.347	0.417	8.020	<0.001	***
Word Length	-0.318	0.121	-2.636	0.008	**
Word Frequency	0.238	0.074	3.189	0.001	**
Session	1.595	0.301	5.294	<0.001	***
ISI	-0.492	0.118	-4.160	<0.001	***
RI	-0.378	0.123	-3.064	0.002	**
Education	0.435	0.225	1.932	0.053	.
Session:ISI	-0.229	0.102	-2.248	0.025	*
Session:RI	-0.374	0.106	-3.532	<0.001	***
ISI:RI	0.149	0.069	2.161	0.031	*
Session:Education	0.100	0.162	0.619	0.536	
ISI:Education	0.020	0.080	0.253	0.800	
RI:Education	0.042	0.075	0.557	0.578	
Session:ISI:RI	-0.008	0.025	-0.314	0.754	
ISI:RI:Education	-0.007	0.028	-0.249	0.804	

Signif. codes: 0.001

0.01
**
0.05
*
.0.1

Table A6.
LMEM of spacing effect, including covariate of Lesion Volume.

Formula: accuracy ~ Length + Session*ISI*RI*Lesion Volume + (1 + Length + Frequency + Session + ISI*RI + Session:ISI + Session:RI | participant) + (1 | item)

Predictor:	Estimate	Std. Error	z-value	Pr(> z)	value
(Intercept)	3.203	0.448	7.152	< 0.001	***
Word Length	-0.322	0.126	-2.556	0.011	*

Predictor:	Estimate	Std. Error	z-value	Pr(> z)	value
Word Frequency	0.237	0.075	3.145	0.002	**
Session	1.482	0.326	4.542	< 0.001	***
ISI	-0.476	0.127	-3.739	< 0.001	***
RI	-0.402	0.146	-2.757	0.006	**
Lesion Volume	0.126	0.162	0.773	0.439	
Session:ISI	-0.233	0.111	-2.092	0.036	*
Session:RI	-0.386	0.122	-3.171	0.002	**
ISI:RI	0.135	0.066	2.045	0.041	*
Session:Lesion Volume	0.055	0.087	0.639	0.523	
ISI:Lesion Volume	0.001	0.058	0.024	0.981	
RI:Lesion Volume	0.025	0.069	0.356	0.722	
Session:ISI:RI	-0.009	0.025	-0.344	0.731	
ISI:RI:Lesion Volume	-0.005	0.019	-0.271	0.786	

Signif. codes: 0.001

0.01

**
0.05

*
.01

Appendix B

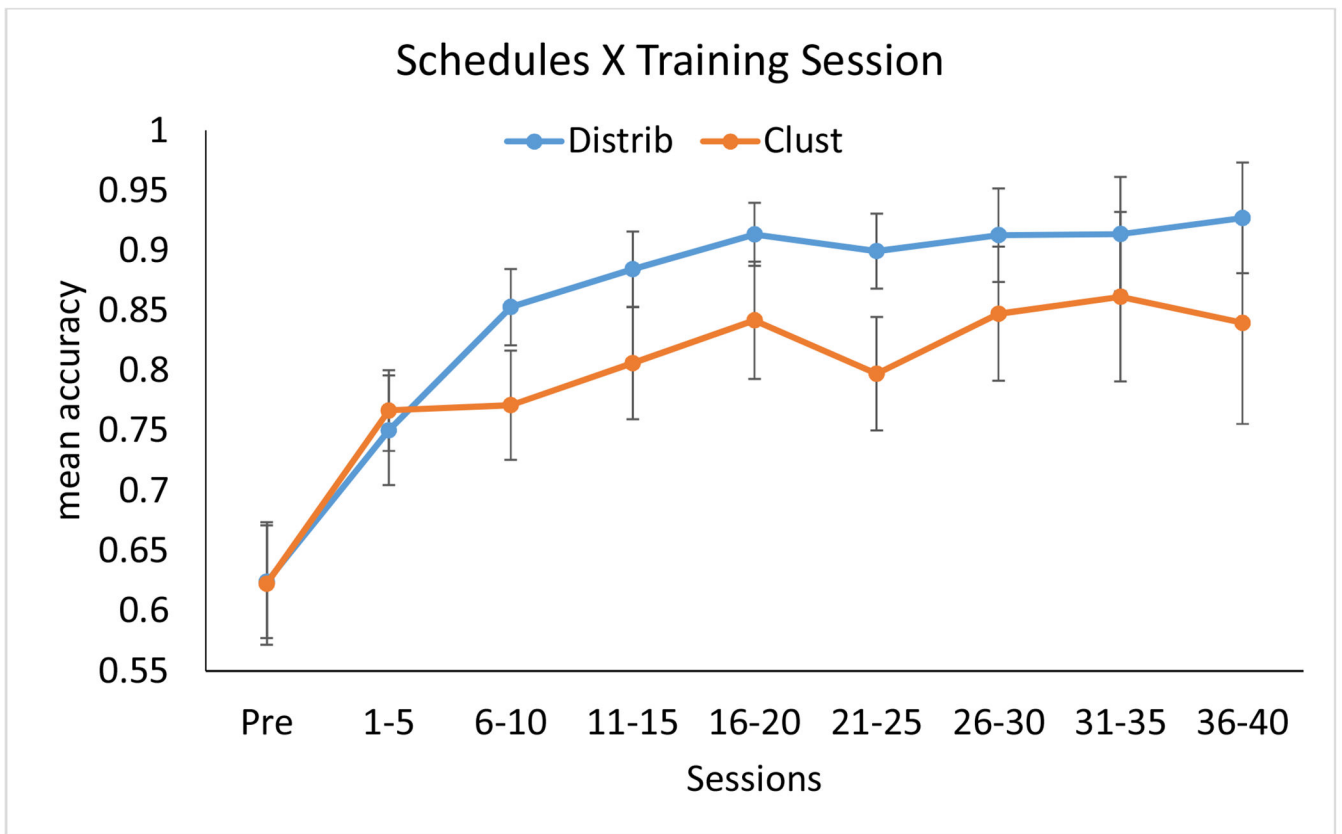


Figure B1. Effect of schedules on rate of improvement, plotting raw data (mean accuracy) in bins of 5 sessions. Error bars reflect standard error of the mean ($N = 11$). For a depiction of the LMEM analysis results, which control for covariates, refer to Figure 5 in the text.

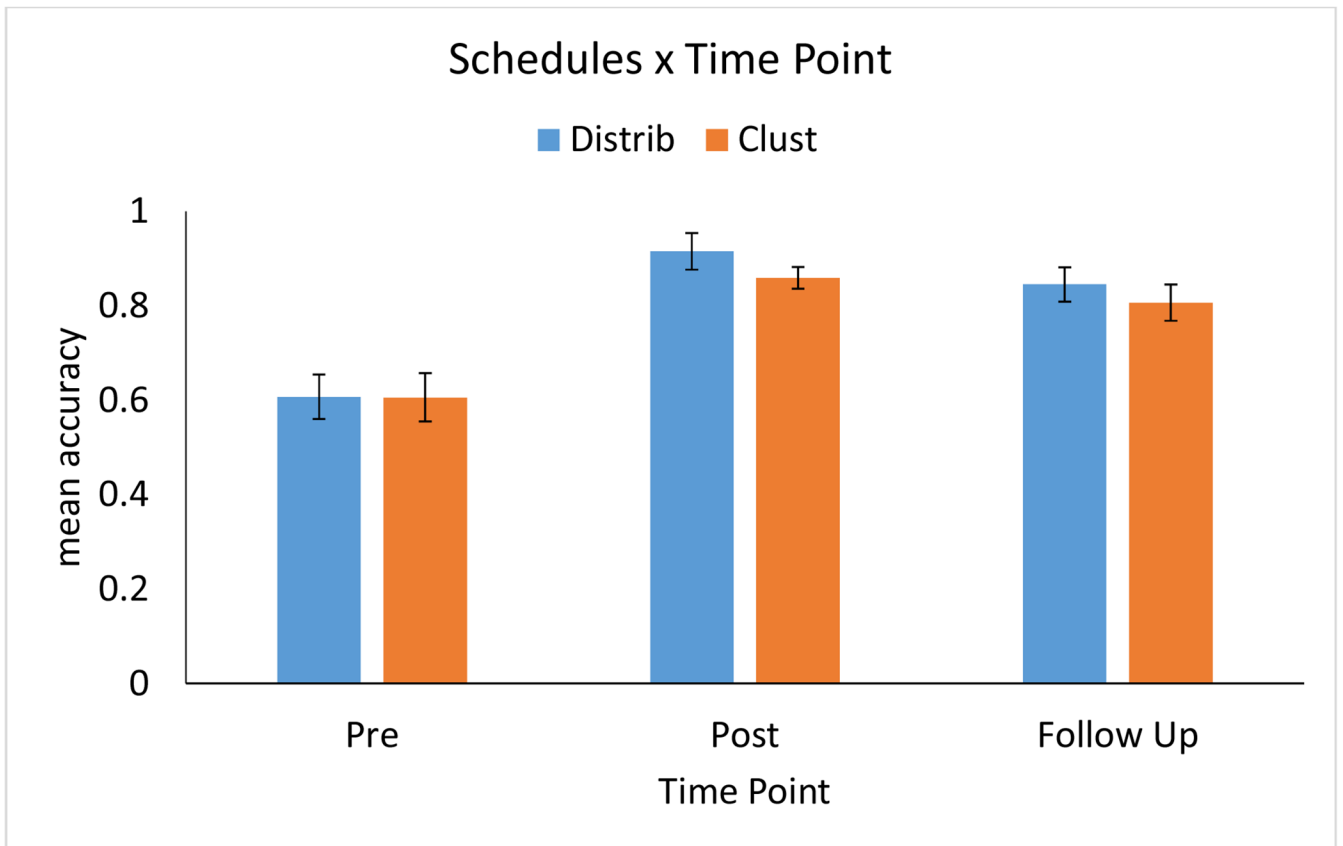


Figure B2. Effect of schedules across Pre, Post, and Follow-up, plotting raw data (mean accuracy). Error bars reflect standard error of the mean ($N = 11$). For a depiction of the LMEM analysis results, which control for covariates, refer to Figure 6 in the text.

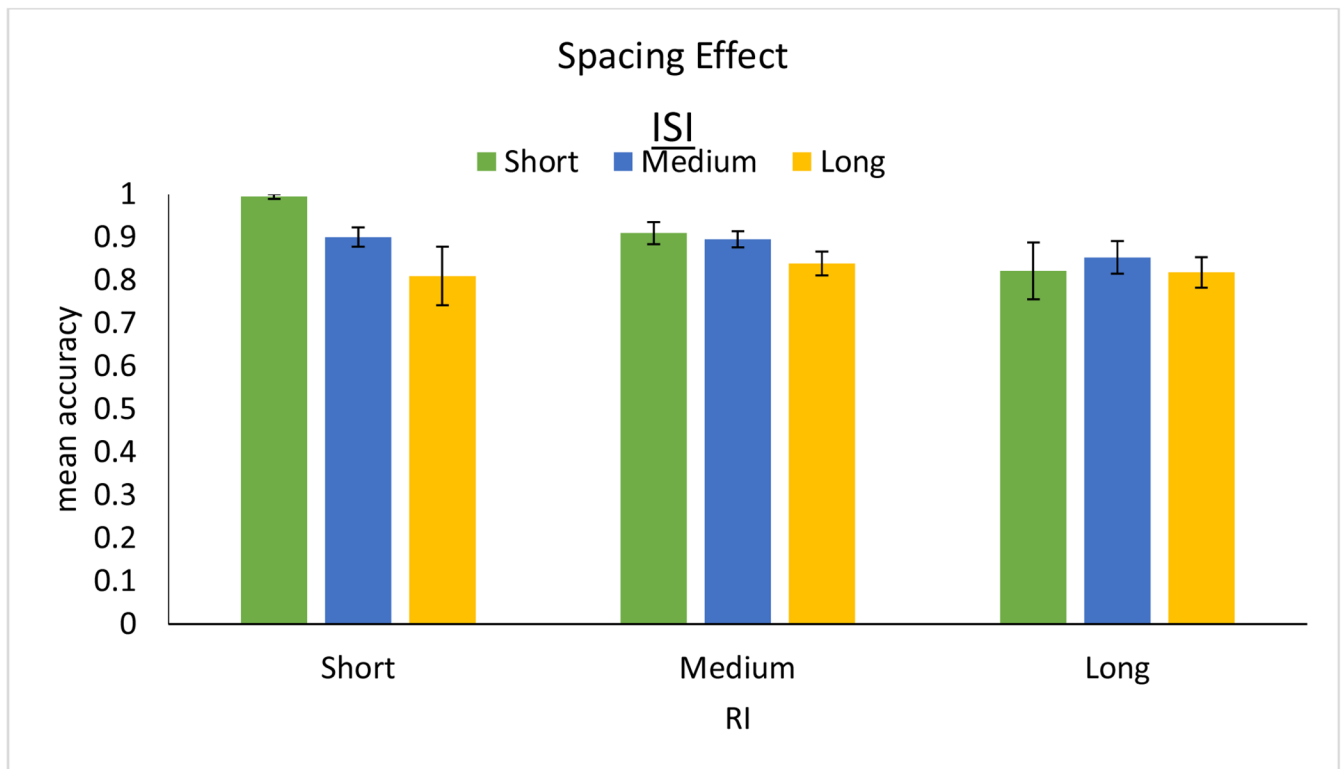


Figure B3.

Spacing effect, plotting raw data (mean accuracy) in bins of short (<4 days), medium (4-21 days), and long (>21 days) duration. Error bars reflect standard error of the mean (N = 23). For a depiction of the LMEM analysis results, which control for covariates, refer to Figure 7 in the text.

References

- Baddeley AD, Emslie H, & Nimmo-Smith I (1994). The doors and people test. Bury St. Edmunds, UK: Thames Valley Test Company.
- Barr DJ, Levy R, Scheepers C, & Tily HJ (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bartón K (2018). MuMIn: Multi-Model Inference. R package version 1.42.1. <https://CRAN.R-project.org/package=MuMIn>
- Bates D, Martin M, Bolker B, & Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Balota DA, Yap MJ, Hutchison KA, Cortese MJ, Kessler B, Loftis B, et al. The English lexicon project *Behav. Res. Methods*, 39 (3) (2007), pp. 445–459 [PubMed: 17958156]
- Beeson PM (1999). Treating acquired writing impairment: Strengthening graphemic representations. *Aphasiology*, 13(9-11), 767–785.
- Bjork RA, Dunlosky J, & Kornell N (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417–444.
- Callan DE, & Schweighofer N (2010). Neural correlates of the spacing effect in explicit verbal semantic encoding support the deficient-processing theory. *Human Brain Mapping*, 4(31), 645–659.
- Cepeda NJ, Pashler H, Vul E, Wixted JT, & Rohrer D (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354. [PubMed: 16719566]

- Cepeda NJ, Vul E, Rohrer D, Wixted JT, & Pashler H (2008). Spacing effects in learning: A temporal ridge of optimal retention. *Psychological Science*, 19(11), 1095–1102. [PubMed: 19076480]
- Crawford JR, & Garthwaite PH (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, 40(8), 1196–1208 [PubMed: 11931923]
- Dignam J, Copland D, McKinnon E, Burfein P, O'Brien K, Farrell A, & Rodriguez AD (2015). Intensive versus distributed aphasia therapy: A nonrandomized, parallel-group, dosage-controlled study. *Stroke*, 46(8), 2206–2211. [PubMed: 26106114]
- Dignam J, Rodriguez AD, & Copland DA (2016). Evidence for intensive aphasia therapy: consideration of theories from neuroscience and cognitive psychology. *PM&R*, 8(3), 254–267. [PubMed: 26107539]
- Duchaine B, & Nakayama K (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, 44(4), 576–585. [PubMed: 16169565]
- Fox J, & Weisberg S (2018). Visualizing Fit and Lack of Fit in Complex Regression Models with Predictor Effect Plots and Partial Residuals. *Journal of Statistical Software*, 87(9).
- Goodman RA, & Caramazza A (1985). *The Johns Hopkins University dysgraphia battery*. Baltimore, MD: Johns Hopkins University.
- Howard D, & Patterson KE (1992). *The Pyramids and Palm Trees Test: A Test of Semantic Access from Words and Pictures*. Thames Valley Test Company.
- Jenkins WM, & Merzenich MM (1987). Chapter 21 Reorganization of neocortical representations after brain injury: a neurophysiological model of the bases of recovery from stroke In *Progress in Brain Research* (Vol. 71, pp. 249–266). Elsevier. [PubMed: 3588947]
- Kay J, Lesser R, & Coltheart M (1996). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology*, 10(2), 159–180.
- Kessels RP, Van Zandvoort MJ, Postma A, Kappelle LJ, & De Haan EH (2000). The Corsi block-tapping task: standardization and normative data. *Applied Neuropsychology*, 7(4), 252–258. [PubMed: 11296689]
- Kornell N (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 23(9), 1297–1317.
- Kuznetsova A, Brockhoff PB, & Christensen RHB (2017). ImerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13).
- Lindsey RV, Shroyer JD, Pashler H, & Mozer MC (2014). Improving students' long-term knowledge retention through personalized review. *Psychological Science*, 25(3), 639–647. [PubMed: 24444515]
- Maddox GB (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: a case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology*, 28(6), 684–706.
- Martins IP, Leal G, Fonseca I, Farrajota L, Aguiar M, Fonseca J, et al. A randomized, rater-blinded, parallel trial of intensive speech therapy in sub-acute post-stroke aphasia: the SP-I-R-IT study Int. *J. Lang. Commun. Disord*, 48(4) (2013), pp. 421–431 [PubMed: 23889837]
- McClelland JL, McNaughton BL, & O'Reilly RC (1995). Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights From the Successes and Failures of Connectionist Models of Learning and Memory. *Psychological Review*, 102(3), 419–457. [PubMed: 7624455]
- McDonald RJ, & White NM (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience*, 107(1), 3–22. [PubMed: 8447956]
- Medina J, & Rapp B (2014). Rapid Experience-Dependent Plasticity following Somatosensory Damage. *Current Biology*, 24(6), 677–680. [PubMed: 24613304]
- Middleton EL, Schwartz MF, Rawson KA, Traut H, & Verkuilen J (2016). Towards a theory of learning for naming rehabilitation: Retrieval practice and spacing effects. *Journal of Speech, Language, and Hearing Research*, 59(5), 1111–1122.

- Milner B (1962). Les troubles de la memoire accompagnant des le- sions hippocampiques bilaterales Physiologie de l'hippocampe [Memory impairment associated with bilateral hippocampal le- sions] (pp. 257–272). Paris: Centre National de la Recherche Scienti- fique.
- Mozer MC, Pashler H, Cepeda N, Lindsey R, & Vul E (2009). Predicting the optimal spacing of study: A multiscale context model of memory In Bengio Y, Schuurmans D, Lafferty J, Williams CKI, & Culotta A (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 1321–1329). La Jolla, CA: NIPS Foundation.
- Nakagawa S, & Schielzeth H (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142.
- Packard MG, & Knowlton BJ (2002). Learning and Memory Functions of the Basal Ganglia. *Annual Review of Neuroscience*, 25(1), 563–593.
- Pashler H, & Baylis GC (1991). Procedural learning: II. Intertrial repetition effects in speeded-choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1), 33.
- Pashler H, Rohrer D, Cepeda NJ, & Carpenter SK (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review*, 14(2), 187–193. [PubMed: 17694899]
- Pavlik PI, & Anderson JR (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, 29(4), 559–586. [PubMed: 21702785]
- Poldrack RA, Clark J, Pare-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, & Gluck MA (2001). Interactive memory systems in the human brain. *Nature*, 414(6863), 546–550. [PubMed: 11734855]
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL <https://www.R-project.org/>.
- Ramsberger G, & Marie B (2007). Self-administered cued naming therapy: A single-participant investigation of a computer-based therapy program replicated in four cases. *American Journal of Speech-Language Pathology*, 16(4), 343–358. [PubMed: 17971494]
- Rapp B, & Kane A (2002). Remediation of deficits affecting different components of the spelling process. *Aphasiology*, 16(4–6), 439–454.
- Raymer AM, Singletary F, Rodriguez A, Ciampitti M, Heilman KM, & Rothi LJJ (2006). Effects of gesture+ verbal treatment for noun and verb retrieval in aphasia. *Journal of the International Neuropsychological Society*, 12(6), 867–882. [PubMed: 17064449]
- Roediger HL III, & Karpicke JD (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. [PubMed: 16507066]
- Sage K, Snell C, & Lambon Ralph MA (2011). How intensive does anomia therapy for people with aphasia need to be?. *Neuropsychological Rehabilitation*, 21(1), 26–41. [PubMed: 21181603]
- Sheather SJ (2009). *A Modern Approach to Regression with R*. Springer Science & Business Media.
- Shimamura AP, Janowsky JS, & Squire LR (1990). Memory for the temporal order of events in patients with frontal lobe lesions and amnesic patients. *Neuropsychologia*, 28(8), 803–813. [PubMed: 2247207]
- Staddon JER, Chelaru IM, & Higa JJ (2002). Habituation, memory and the brain: The dynamics of interval timing. *Behavioural Processes*, 57(2–3), 71–88. [PubMed: 11947990]
- Thompson CK, Lukic S, King MC, Mesulam MM, & Weintraub S (2012). Verb and noun deficits in stroke-induced and primary progressive aphasia: The Northwestern Naming Battery. *Aphasiology*, 26(5), 632–655. [PubMed: 23188949]
- Wagner A, Desmond JE, Glover GH, & Gabrieli JDE (1998). Prefrontal cortex and recognition memory. Functional-MRI evidence for context-dependent retrieval processes. *Brain*, 121(10), 1985–2002. [PubMed: 9798751]
- Wixted JT (2004). On Common Ground: Jost's (1897) Law of Forgetting and Ribot's (1881) Law of Retrograde Amnesia. *Psychological Review*, 111(4), 864–879. [PubMed: 15482065]

- distributed practice is superior to massed practice in language re-learning
- longer retention is facilitated by longer interstudy intervals in acquired dysgraphia
- similar principles govern re-learning in the lesioned and intact brain

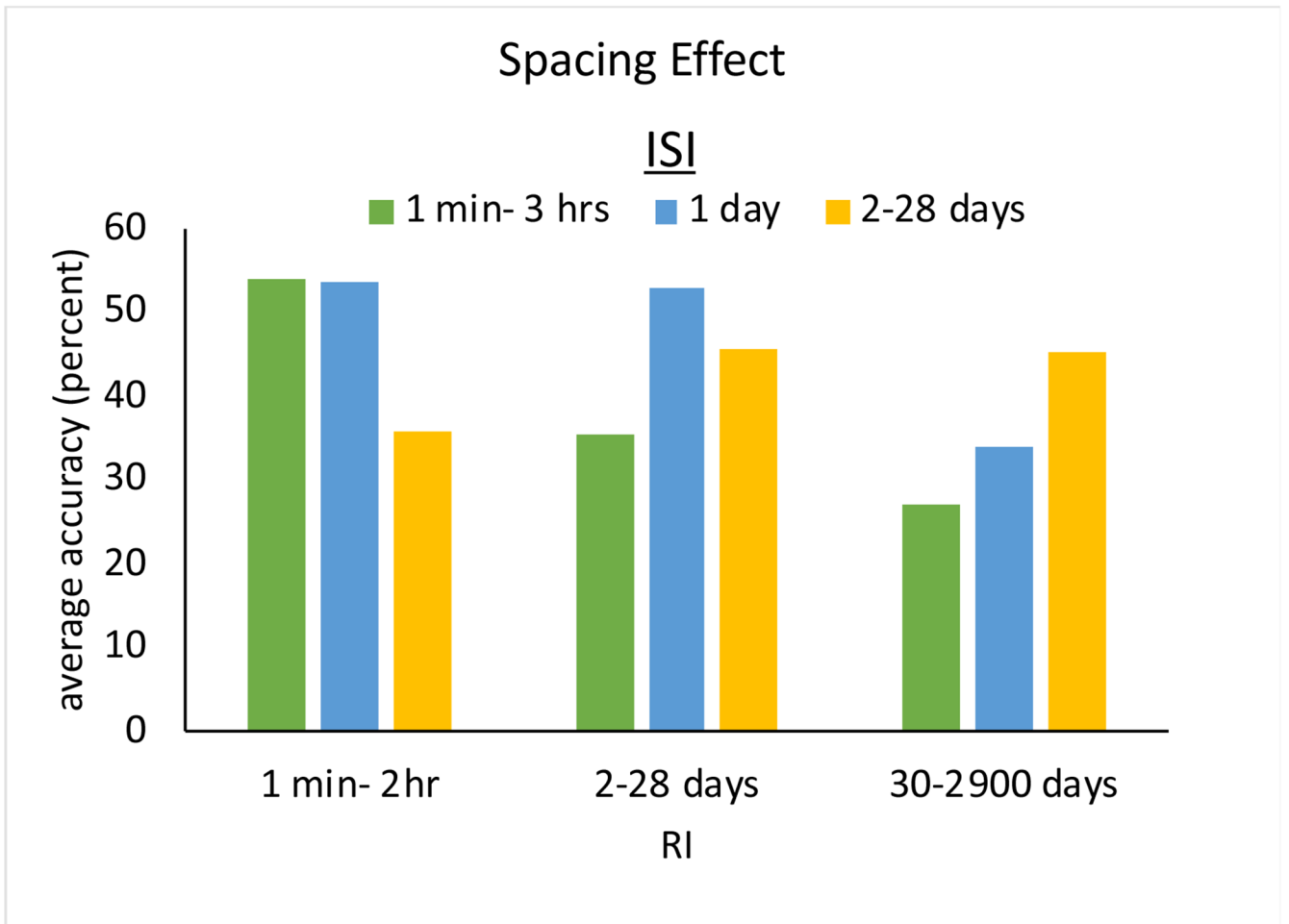


Figure 1.

The figure (adapted from Cepeda et al., 2006) depicts the finding that shorter interstudy intervals (ISI's) are more beneficial when material must be retained for shorter time periods, while the reverse is true for longer retention periods.

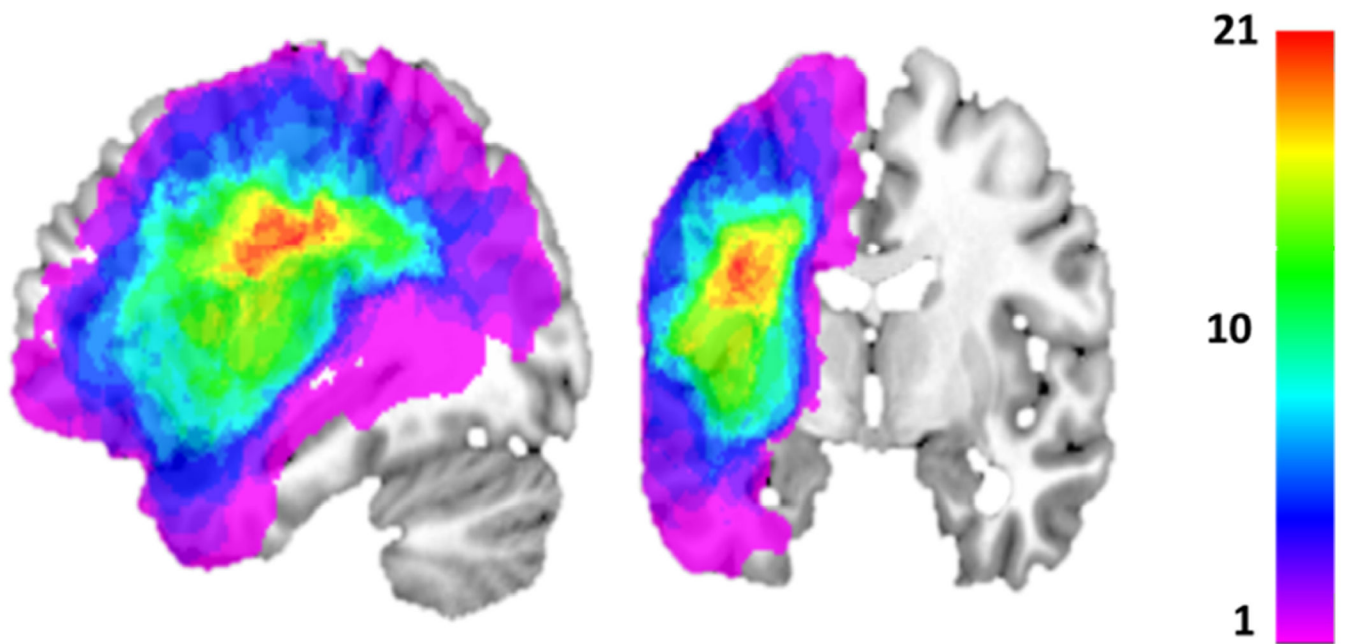


Figure 3. Lesion overlap for the 23 participants, with warm colors indicating areas of greater overlap. Highest lesion density occurs in left hemisphere posterior frontal, anterior parietal and superior temporal lobe areas.

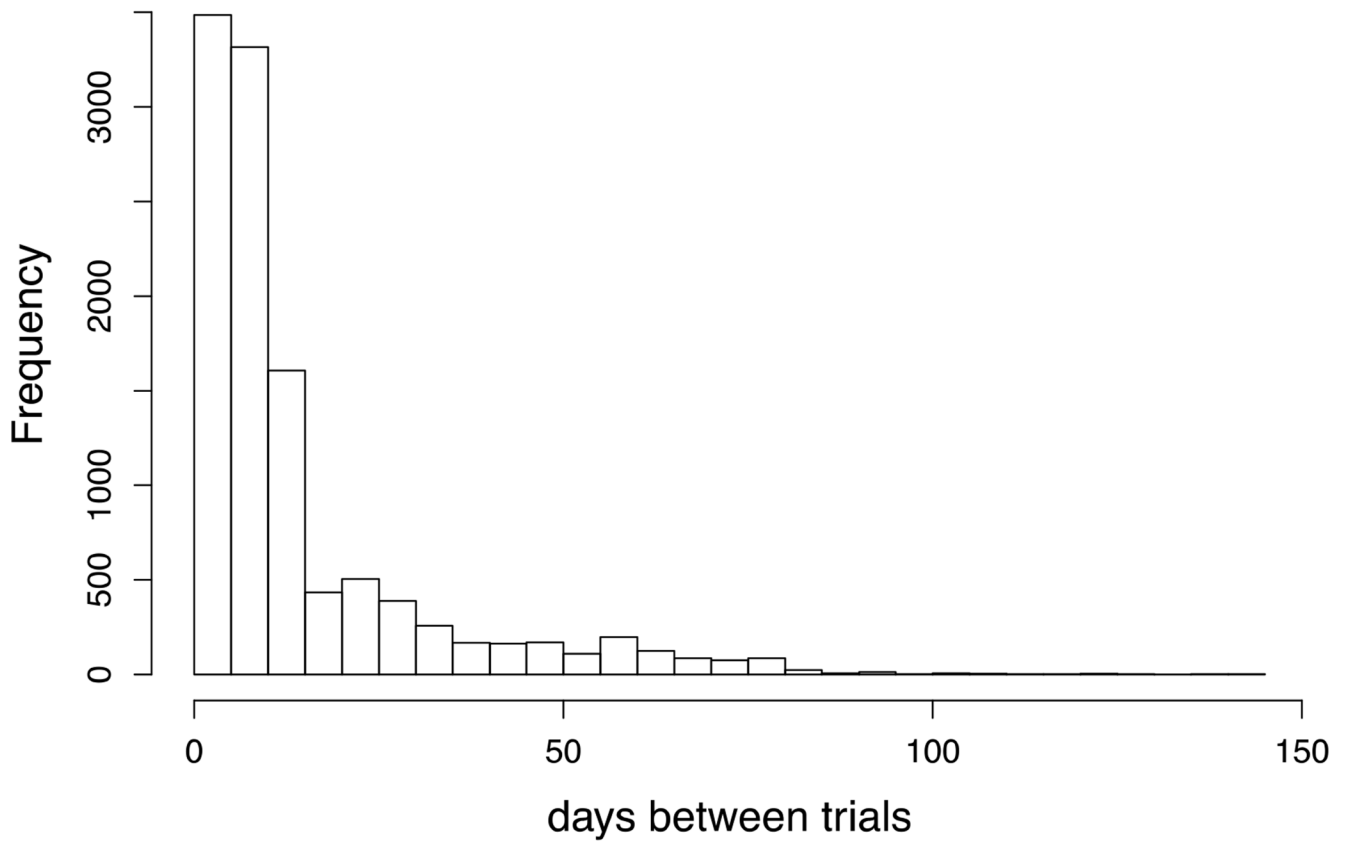


Figure 4. The distribution of intervals (number of days) between training trials across the 23 participants and all training trails. Because items were tested and trained on the same trials, these values served both as ISI's and RI's.

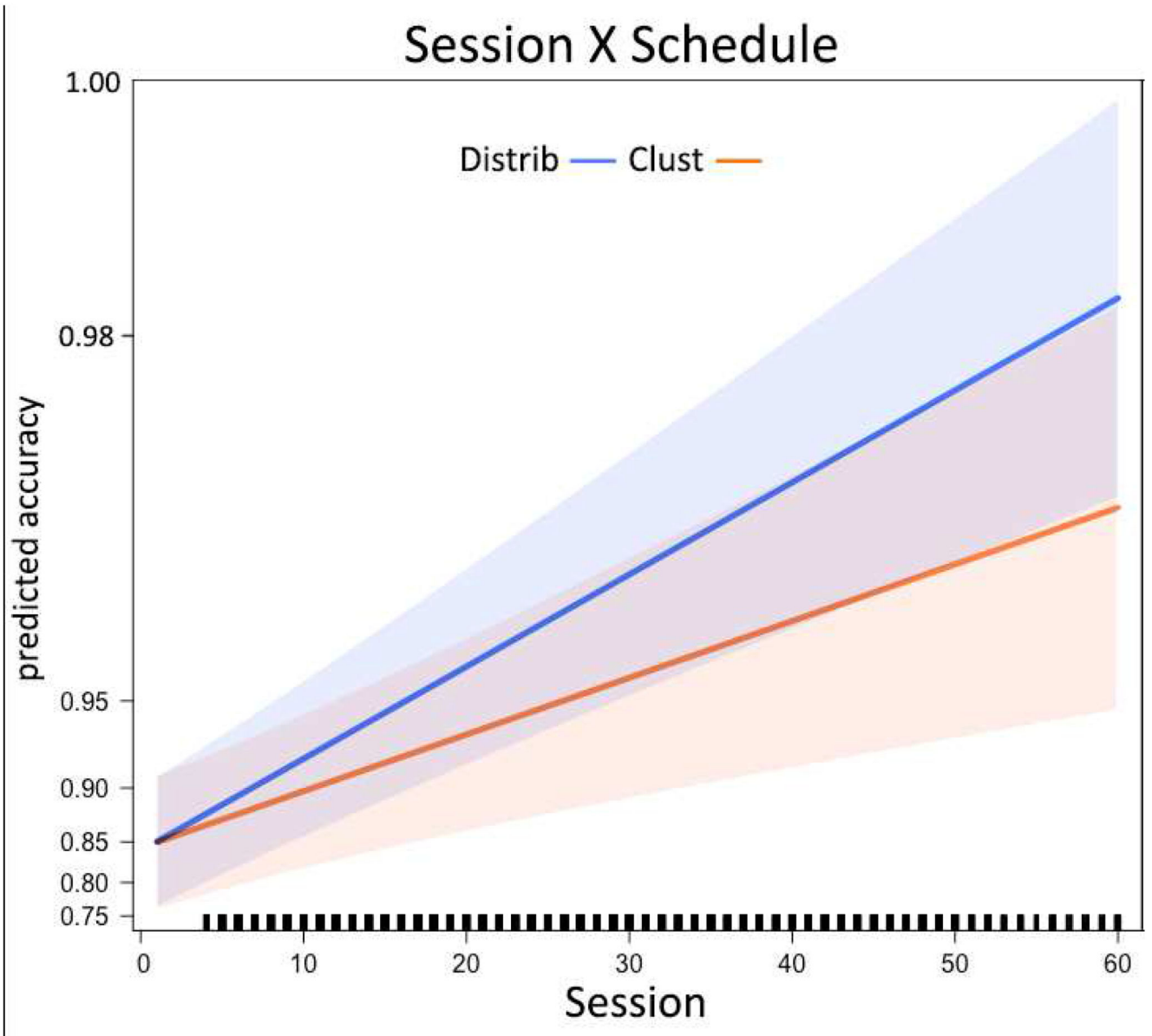


Figure 5.

The effect of training schedules on learning rates. The figure depicts the model-predicted improvement in spelling (y-axis) across training sessions (x-axis), for the Distributed (blue) and Clustered schedules (orange). The shaded errors reflect the 95% confidence interval around the fixed effects, as returned by the R package *effects* (Fox & Weisberg, 2018).

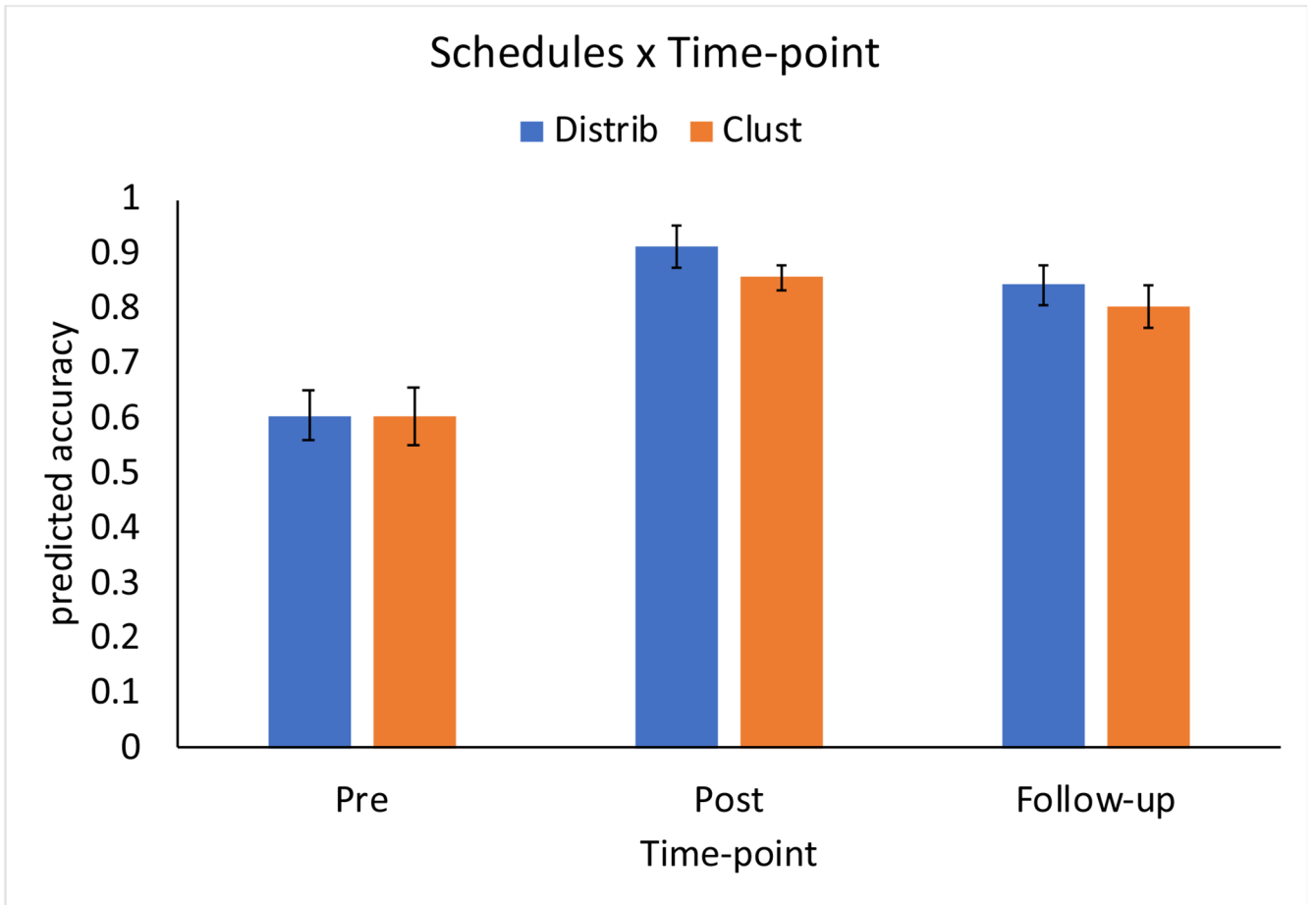


Figure 6. The effect of training Schedules on spelling accuracy for trained words at Pre- and Post-treatment and at a 3-month Follow-up. Depicted are the model-predicted accuracy improvements; error bars reflect 95% confidence intervals around the fixed effects estimates (Fox & Weisberg, 2018).

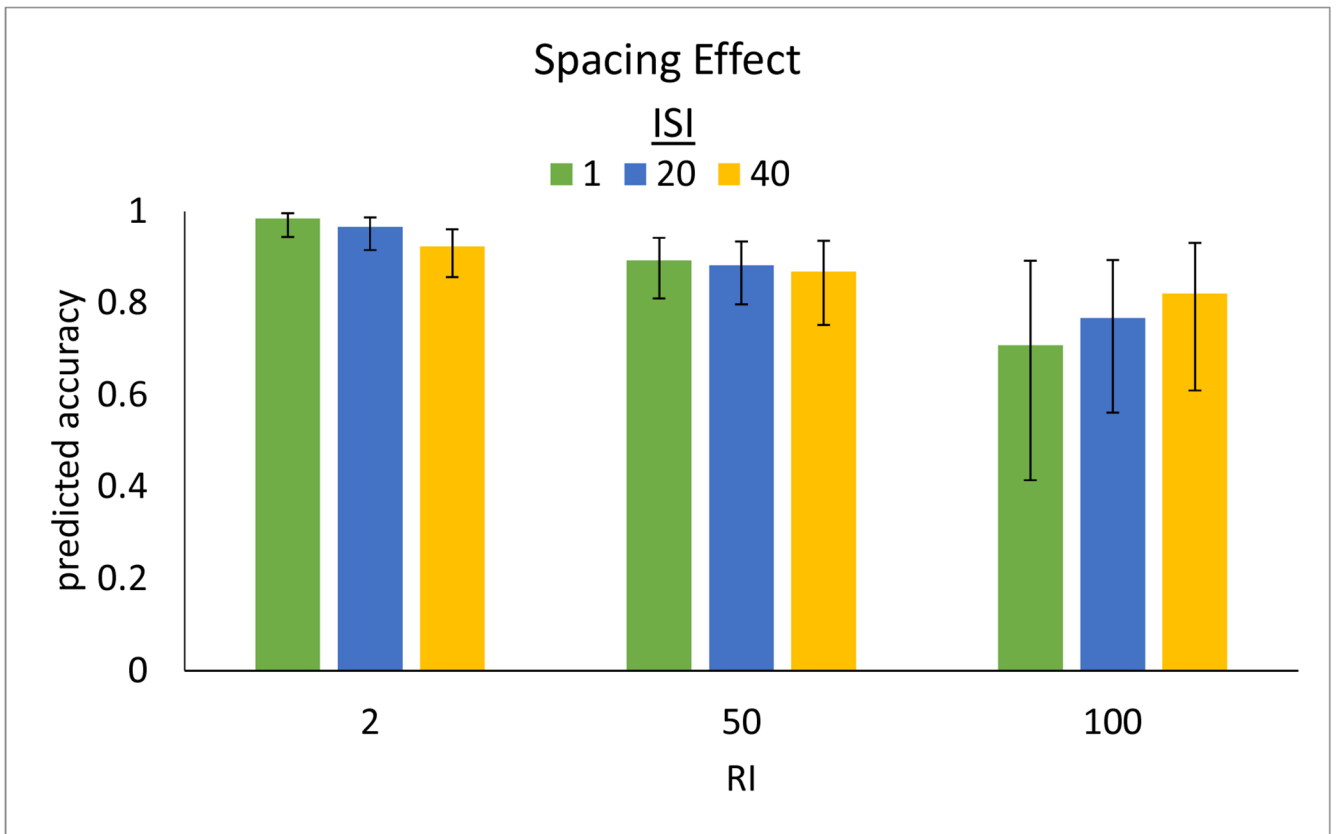


Figure 7. Depiction of the results of the LMEM analysis of spacing of study, showing model-predicted spelling accuracy (y-axis) for 3 levels of RI (x-axis; 2, 50, and 100 days) across 3 levels of ISI (1, 20, and 40 days in green, blue, and yellow respectively).

Table 1.

Demographic, lesion, and assessment data for the 23 participants. P-values and percentiles represent comparisons with normative samples.

ID	Sex	Age (yrs)	Education (yrs)	Time post stroke (months)	Lesion volume (cc)	Number of treatment sessions	Spelling severity Percent Correct	CMTF Percent Correct (Crawford p-value)	Doors Percent Correct (Percentile)	Corsi Blocks Spatial Span (Crawford p-value)
DTE*	F	80	18	14	64.7	40	64	75 (0.683)	71 (25)	5 (0.363)
JRE	F	75	18	207	91	42	69	63 (0.678)	67 (10-25)	5 (0.363)
AES*	F	59	16	209	225	48	67	n/a	92 (95-99)	5 (0.363)
LHT*	M	74	16	16	78	56	29	61 (0.613)	79 (75)	6 (0.879)
PQS*	M	54	18	17	92.8	17	58	81 (0.44)	63 (10-25)	4 (0.097)
THD	M	67	18	76	215	21	95	53 (0.294)	71 (25)	4 (0.097)
TTR	F	46	16	21	113.9	21	66	63 (0.678)	75 (50)	5 (0.363)
CIE	F	62	14	85	45.9	17	51	53 (0.294)	58 (10-25)	5 (0.363)
DSK*	M	67	16	59	135.5	12	70	39 (0.054)	88 (95)	5 (0.363)
JGL*	F	72	16	32	36.3	48	91	63 (0.678)	71 (25)	5 (0.363)
RFZ	M	60	18	46	55.3	16	100	n/a	42 (1-5)	1 (<0.001)
RHH	M	45	16	82	108.8	18	100	79 (0.496)	58 (10-25)	3 (0.017)
RHN*	F	75	19	27	7.7	16	90	49 (0.188)	58 (10-25)	5 (0.363)
CCN	M	40	18	38	100.7	23	91	49 (0.188)	33 (1-5)	4 (0.097)
ABS	M	58	18	97	138.9	19	94	68 (0.961)	79 (75)	5 (0.363)
AEF*	F	55	16	101	214.6	25	90	54 (0.338)	58 (10-25)	5 (0.363)
ESG	M	62	16	38	102.8	27	53	79 (0.496)	100 (99)	6 (0.879)
FCE	M	64	12	119	42.2	19	74	63 (0.678)	83 (90)	6 (0.879)
KMN*	M	55	15	28	52	48	59	57 (0.436)	75 (50)	6 (0.879)
KST*	M	61	14	46	18.2	29	83	71 (0.894)	33 (1-5)	2 (0.002)
MSO*	M	45	18	103	166.9	30	36	50 (0.22)	75 (50)	6 (0.879)
TCl	F	69	12	45	67	22	61	49 (0.188)	79 (75)	5 (0.363)
TCK	M	69	16	68	21.1	26	83	40 (0.066)	83 (90)	4 (0.097)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

* indicates individuals participating only in Study 1; Spelling Severity = percent correct on the Length List of the JHU Dysgraphia Battery (Goodman & Caramazza, 1985); CMFT = Cambridge Memory Face Test (Duchaine & Nakayama, 2006); Doors = Doors and People (Baddeley et al., 1994); Corsi Blocks (Kessels et al., 2000); Crawford refers to a modified independent samples t-test (Crawford and Garthwaite, 2002).

Table 2.

Performance on language assessments before and after treatment. Trained words = performance on individualized word sets targeted in treatment. Untrained words = Length List from JHU Dysgraphia Battery (Goodman & Caramazza, 1985) administered immediately prior to treatment. Oral reading = PALPA 35 (Kay et al., 1996); Reading comprehension = PALPA 51 (Kay et al., 1996). Single word auditory comprehension = Northwestern Naming Battery (Thompson et al., 2012). Picture semantics = Pyramids and Palm Trees (Howard & Patterson, 1992). Spoken word production = Northwestern Naming Battery (Thompson et al., 2012). Bonferroni corrected threshold for multiple comparisons is $p < 0.007$.

	Pre-Treatment Percent Accuracy (SD)	Post-Treatment Percent Accuracy (SD)	Paired t-test
Spelling: Trained words	62 (12)	90 (14)	<0.001
Spelling: Untrained words	73 (20)	77 (22)	0.004
Oral Reading	69 (25)	74 (23)	0.073
Reading Comprehension	66(16)	67 (17)	0.656
Auditory Comprehension	95 (6)	97 (6)	0.069
Picture Semantics	94 (3)	94 (4)	0.823
Spoken Picture Naming	66 (25)	70 (21)	0.181