



Published in final edited form as:

Neuropsychol Rehabil. 2021 June ; 31(5): 731–772. doi:10.1080/09602011.2020.1731554.

Treatment of verbal short-term memory abilities to improve language function in aphasia: A case series treatment study

Nadine Martin¹, Julie Schlesinger¹, Jessica Obermeyer², Irene Minkina¹, Samantha Rosenberg¹

¹Temple University, Philadelphia, PA

²University of North Carolina at Greensboro, Greensboro, NC

Abstract

Recent approaches to interventions for aphasia have incorporated verbal short-term memory (STM) and working memory (WM) components. We were interested in whether a treatment involving repetition of word sequences after a response delay would improve tolerance of increased verbal STM load in repetition and, consequently, improve performance on repetition and other language tasks. Eight individuals with aphasia participated. We used a single subject design with outcome measures on near-transfer tasks closely related to the treatment task and far-transfer tasks more distantly related to the treatment task. We controlled for a confound between effects of repeated exposure of treated items and effects of the STM component of repetition by minimizing repeated presentation of stimulus words in all phases of treatment. Four participants demonstrated modest acquisition effects. Some participants improved on near-transfer tasks, repetition of concrete and abstract word strings (three participants) and verbal spans (eight participants). Improvements on far-transfer tasks included naming (one participant) and discourse measures (three participants). Improvements were most evident for those participants who demonstrated a significant decline in word repetition accuracy after a response delay before treatment was initiated, suggesting that this treatment may be a good fit for individuals whose word processing deficit involves a specific difficulty in maintaining activation of linguistic representations. More studies are needed to determine who will respond to this treatment and what factors might influence the effectiveness of this treatment approach (e.g., severity of impairment).

Keywords

aphasia; verbal short-term memory; verbal working memory; treatment

Introduction

Aphasia impairs access to and retrieval of linguistic representations (McNeil & Pratt, 2000) and can affect all linguistic domains (Rosenbek, LaPointe & Wertz, 1989). Some models of language (Dell, 1986) and aphasia propose that access and retrieval abilities are supported by temporary activation and maintenance of activated language representations (Dell, 1986;

Dell, Martin, Schwartz, Saffran and Gagnon, 1997). Impairment to these processes affects performance on word retrieval in naming and repetition tasks (Dell, Schwartz, Saffran, Martin & Grewal, 1997), verbal short-term memory (STM) (Martin, Dell & Saffran, 1996; Martin & Saffran, 1997), verbal working memory (WM) (Martin, Minkina, Kohen & Kalinyak-Fliszar, 2019), and verbal learning tasks (Martin & Saffran, 1999). This study focuses on a treatment approach that targets the ability to maintain activation of semantic and phonological representations of words over the time course of word retrieval. This short-term activation maintenance ability is targeted here in the context of repetition of single words and word sequences. The need for such a treatment is motivated empirically by the observation that for some individuals with aphasia, accuracy of repetition decreases after a response delay (e.g., Kalinyak-Fliszar, Kohen, & Martin, 2011). This pattern is also observed in naming, and, as discussed below, there are individuals who show better performance on repetition and naming tasks after a response delay. In this study, we focus on improving short-term maintenance of activated words in the context of repetition of words and word sequences.

Theoretically, this treatment approach is motivated by two closely related models. First are language-based models of verbal STM which postulate a close, intricate relationship between word processing and verbal short-term memory abilities and their impairment in aphasia (e.g., Berndt & Mitchum, 1990; Craik & Lockhart, 1973; Martin & Ayala, 2004; Martin & Gupta, 2004; Martin & Saffran, 1997; R. Martin, Yaffee, & Sheldon, 1994; Saffran, 1990; Saffran & Martin, 1990). These models postulate that over the time course of language processing, words must be activated, and that activation must be maintained at adequate levels until the language task (e.g., word production) is complete. The second model is the interactive activation model of word retrieval advanced by Dell (1986) and colleagues (Dell et al., 1997; Dell et al. 2008; Martin et al., 1996; Schwartz, Dell Martin et al., 2006), which provides a finer description of the involvement of short-term maintenance processes and retrieval of linguistic representations. Dell's model postulates two components of processes that mediate access to word representations: (1) connection weight that controls the strength of activation spreading through a word's representation in the semantic-lexical-phonological network and (2) activation decay that controls short-term maintenance of activated word representations over the time course of word retrieval. The accounts of aphasia offered by these models are discussed further below.

In recent years there have been an increasing number of treatment studies using repetition tasks that incorporate some means of increasing memory load (see for reviews, Minkina, Rosenberg, Kalinyak-Fliszar & Martin, 2017; Salis, Kelly, & Code, 2015). These treatments target the processing aspects of language function (e.g., activating, maintaining activation of and retrieving the semantic, lexical and phonological representations of words). This research is in its early stages, but as we learn more about the effects of damage to these processes on word retrieval, treatments targeting these impaired processes in aphasia can be refined and should complement treatments that manipulate the linguistic content of treatment stimuli to target specific linguistic representations (e.g., semantic or phonological representations) of words (e.g., Kiran & Sandberg, 2012; Martin, Fink & Lane, 2006; Martin, 2017; Nettleton & Lesser, 1991). Although no treatment can be said to exclusively focus on linguistic representations or processing aspects of a language disorder, it is

important to gain an understanding of the contributions of each element to our treatment approaches. The treatment evaluated here (repetition of word stimuli following a response delay) minimizes repeated exposure of stimuli used in treatment and probes to allow a focus on effects of increased response delay, a manipulation used to increase memory load on verbal processing. The outcome that we sought is better tolerance of increased memory load in the context of word processing and improved ability to maintain activation of words over the course of repetition and after a time delay.

To date, treatment studies targeting STM aspects of verbal processing (e.g., Kalinyak-Fliszar et al., 2011; Koenig-Bruhin & Studer-Eichenberger, 2007; Majerus, Van Der Kaa, & Renard, 2005; Salis, 2012) have demonstrated some potential of this approach, but more research is needed for several reasons. First, it is important to determine whether this approach is appropriate for specific or all types of language deficits in aphasia. The present treatment aims to improve the ability to maintain activation of representations, but not all impairments of word retrieval are necessarily due to an impairment of this ability. Martin and Dell (2017; 2019), for example, investigated the effects of a 5-second (5-sec) response delay compared to a 1-second (1-sec) response delay on accuracy of picture naming and repetition and observed three patterns of change in accuracy after the delay of five seconds: worse performance, better performance and no change in accuracy. They hypothesized that these patterns reflected impairments to the two processing parameters of the interactive activation model. The first pattern, better performance after a response delay, was attributed to weak connection strength, leading to slowed transmission of activation and a need for more time to activate semantic and phonological representations of words. The second pattern, an increase in the decay rate of activated semantic and phonological representations of words was hypothesized to lead to difficulty in maintaining their activation after a response delay. The third pattern, no change in accuracy after five seconds, was attributed to a combination of connection weight and decay rate impairments. Martin and Dell (2019) tested these three patterns of change after a response delay in a computer implemented version of the interactive activation model of word retrieval, fitting the naming data from people with aphasia in a replication of their 2017 study. The model was able to account for these three patterns of change in accuracy after a response delay with impairments to connection weight (leading to poor transmission), decay rate (poor maintenance of activation) and no change in accuracy. This computational demonstration of the effects of three types of impairment to the hypothesized components of word processing adds to the evidence that motivates treatment of word processing deficits.

In this study, we used a treatment to improve an activation maintenance deficit (activation decays too quickly). We focused on this processing impairment for several reasons. First, this pattern was one that we had observed more often than the transmission deficit. Second, investigating treatment of temporal impairments of word processing in aphasia is in its early stages, and we wanted to focus on just one of the two aberrant patterns. In relation to that point, we did not systematically identify participants a priori who exhibited this specific impairment. Rather, we kept the sample open with respect to type of processing impairment and conducted a post hoc analysis of the participants' responses to treatment that suggests this treatment is most effective for those who show an activation maintenance deficit severe enough to reduce accuracy of repetition after a response delay. Currently, we are developing

a treatment for an activation transmission deficit (slow to activate word representations), but note also that Conroy and colleagues (Conroy, Drosopoulou, Humphreys, Halai, & Lambon Ralph, 2018) have developed a treatment approach recently that trains individuals with aphasia to speed up processing of words. Thus, interest in identifying and treating disorders of temporal processing in aphasia is increasing.

A second question to investigate is whether treatments that directly target underlying processes of word retrieval, including short-term maintenance of activation, promote generalization of improvements to items and tasks not targeted directly in treatment. In theory they should, but generalization, also referred to as skill transfer, could be limited to tasks that are closely related to the treatment task (termed “near-transfer” tasks) such as repetition, or they could have more widespread effects, leading to improvement on more distantly related “far-transfer” tasks, such as discourse. Transfer (or generalization of treatment effects from trained to untrained stimuli is of great clinical importance. And yet, there is still much that we do not understand about how to achieve this goal (see Webster, Whitworth & Morris (2015) for discussion). This study provides an opportunity to observe if generalization occurs to other tasks, some overlapping with the treatment task (e.g., repetition) and others more distantly related (e.g., discourse).

Third, interventions that incorporate a short-term activation maintenance manipulation in the treatment tasks often confound the linguistic and STM/WM components of the treatment, especially if repetition is used as the treatment task. In the context of domain specific models of STM (e.g., Cowan, 2005; 2008), this is unavoidable and in principle, is appropriate, because the short-term maintenance of activated word representations is assumed to be inherent in the access and retrieval of those representations. Thus, it makes sense to target them together in a treatment task. However, in research, it is important to isolate the variable of interest as much as possible. Single subject research designs typically include a set of treatment stimuli (e.g., word lists or sentences) and a set of untreated stimuli that serve as controls but may also improve somewhat if the treatment generalizes to other words besides the treatment stimuli. A consequence of this design is that the treated stimuli are presented repeatedly throughout treatment and therefore are subject to item-specific repetition effects. For our purposes, the use of a specific set of treatment stimuli coupled with a short-term activation manipulation (e.g., delayed repetition) would make it difficult to determine whether treatment effects resulted from improved short-term maintenance of activation or from repeated exposure of training and probe items. Thus, for this study of the repetition + response delay treatment we used a standard ABA design, but rather than using small, closed sets of trained and untrained items, we used unique items as much as possible in baselines, treatment trials, and pre-and post-treatment measures. By minimizing item exposure, we aimed to maximize the degree to which treatment effects could be attributed to improved tolerance of increased verbal STM load.

Theoretical and empirical motivation for treatment of verbal STM components of word processing disorders.

Conceptualization of short-term memory and working memory.—To explain the theoretical basis of this study, it is first necessary to clarify some terms. Our approach to

treatment focuses on improving short-term maintenance of word representations over the course of repeating words and word sequences. The ability to maintain activation of semantic, lexical and phonological representations of words in tasks such as naming, word and sentence repetition and verbal span is considered to be a form of verbal STM. This conceptualization of verbal STM aligns with that advanced by Cowan (2008), who proposes that STM is a mental ability that maintains a limited amount of information in a state that makes it temporarily accessible. Importantly, he notes further that WM is not distinct from STM, but rather, *includes* it along with other cognitive mechanisms (e.g., attention) that “make use of short-term memory” (p. 384). The treatment task used in this study, ‘repetition with a response delay’, engages verbal STM/WM processes. Although the working memory component is minimal, rehearsal can be employed during the response delay, and this constitutes a bit of “work” beyond maintaining verbal representations in a temporarily accessible state.

It is also important to distinguish the short-term activation maintenance treatment reported here from working memory training which has been the subject of considerable investigation in recent years in the cognitive and neuropsychological literature (cf., Novick, Bunting, Dougherty, Engle, 2019 for review). Working memory training can focus on domain general memory systems that support retention of all types of information or domain specific memory systems that support retention of specific information types such as language or spatial information (e.g., Klingberg, 2010). The treatment task we used in the present study is intended to stimulate verbal STM (short-term maintenance of activated words) which is domain specific. Moreover, the model that we use to motivate this treatment maintains that processing of semantic, lexical, and phonological representations of words occurs over time and that short-term maintenance of the activated representations is inherent in language processing. Thus, although this treatment may be categorized as a verbal “STM” treatment, it is focused on improving word processing, which entails a STM, or short-term activation maintenance component. Put another way, adding a STM/WM load to language tasks in treatment should foster an increase in verbal STM capacity which in turn, should improve language function, especially for those people whose performance on language and verbal STM tasks indicates a difficulty in maintaining activation of word representations. Language-based models of verbal STM (e.g., Cowan, 2005; Craik & Lockhart, 1972) provide a general framework for understanding this approach to language impairment in aphasia (Martin & Gupta, 2004). Dell’s interactive activation model of word processing provides an account of how the short-term activation maintenance deficit affects repetition and production of words. This theoretical framework and supporting empirical evidence have been reviewed in several recent papers (Minkina et al., 2017; Salis et al., 2015) and is briefly summarized below.

A theoretical framework for understanding how verbal processing engages short-term maintenance processes.—The interactive activation (IA) model of word processing (Dell, Martin, & Schwartz, 2007; Dell et al., 1997) and its extension to short-term maintenance of verbal representations in verbal span tasks (Martin & Gupta, 2004) served as our hypothesis-driven theoretical framework for understanding the role of STM in word processing. The IA model postulates two components of activation that support

accurate and efficient language processing: Strength at which a word's representations (semantic, lexical and phonological) are activated and the rate at which they decay at each stage of the activation process. In this model, there is a constant feedforward and feedback activation cycling between levels of a word's semantic and phonological representations from the moment a word's retrieval is initiated until it is retrieved. These two components, activation strength and activation decay, maintain the activation level of the target word (the one that is intended to be spoken or understood) over the time course of production, comprehension or repetition. That maintenance function takes place in the context of a field of other semantically (e.g., rabbit → squirrel) and/or phonologically related (e.g., pencil → panel) word representations that are primed to a lesser extent by spreading activation. Several factors can lead to erroneous retrieval of one of these so-called 'competitor' words, by altering temporarily the activation dynamics in the lexicon in a way that increases their activation relative to that of the target word. This model was first used to account for speech errors made by typical speakers (e.g., Dell, 1986; Dell & Reich, 1981), which have been studied at great length by researchers as "windows" into the abilities that mediate word processing (e.g., Fromkin, 1980). For example, words that are semantically or phonologically related to an intended word utterance are more likely to be retrieved in error than words that are unrelated to the intended utterance (Martin, Weisberg & Saffran, 1989). In applying the interactive activation model to aphasia, Dell and colleagues (1997) proposed that the word retrieval impairment in aphasia resulted from an alteration in the balance between activation strength (reduced connection strength) and decay processes (too-fast-decay) that support word retrieval. Martin and Saffran (1997) proposed that these same processes maintain activation of words in language tasks involving more than a single word (e.g., immediate serial digit/word recall). This idea reflects the central claim that impaired word processing and verbal STM abilities in aphasia are rooted in a common underlying processing impairment that affects short-term maintenance of activated linguistic representations.

Applying the theory to treatment. Studies of aphasia interventions that target the language impairment directly have long recognized the importance of targeting specific linguistic content (e.g., semantic or phonological) affected by aphasia (e.g., Nettleton & Lesser, 1991; van Hees, Angwin, McMahon, & Copland, 2013) using methods that aim to restore access to words and sentences (e.g., Boyle & Coelho, 1995; Edmonds, Mammino & Ojeda, 2014). The intervention tested in this study does not focus directly on representation (e.g., targeting semantic or phonological representations). Rather, it focuses primarily on improving the cognitive processes that support access to words, thereby promoting stronger and more enduring activation of their linguistic representations. In terms of Dell's IA model, the treatment reported here aims to increase the ability to maintain activation in the short-term of the semantic, lexical and phonological representations of words as they are being accessed and retrieved in word processing tasks (e.g., naming, repetition, comprehension).

It is important to note that this "activation maintenance" treatment could and should be combined with a strategic selection of word stimuli that target semantic or phonological representations or phrase and sentence level stimuli, depending on the language abilities of the client. In the present study, however, we aimed to evaluate the effects of increased

demand on maintaining activation of word representations in a repetition task at a base level, without the confound of repeated exposure of items and without any strategic choice of words based on their semantic or phonological characteristics (word sequences). Thus, our treatment task is simply delayed repetition of single words or word sequences, with exposure of stimuli in all phases of testing and treatment minimized as much as possible (see methods for details).

Mechanisms of generalization. Approaches to promoting generalization to similar (near-transfer) and less similar (far-transfer) language tasks include stimulus factors (e.g., category typicality of trained items, Kiran & Thompson, 2003) and processing factors (e.g., executive processes, Yeung & Po-Law, 2009; verbal STM, Kalinyak-Fliszar et al., 2011). The delayed repetition task that we used incorporated a 1-sec, 5-sec or 10-second (10-sec) response delay between hearing the word or sequence of words and a cue to reproduce that word or sequence of words. Outcome measures included better performance on the treatment task after treatment as well as other language and verbal STM measures administered before and after treatment. We hypothesized some transfer to other tasks because the delayed repetition treatment targets fundamental ability related to word processing (input and output): *The ability to maintain access to a word's semantic, lexical and phonological representations over the course of producing, comprehending or repeating verbal stimuli*. If this ability is improved, generalization (or transfer) of treatment effects to tasks other than the treatment task should be evident to some degree. Empirical support for this hypothesis comes from studies indicating that performance on naming, repetition and lexical decision tasks is positively associated with verbal STM capacity (Martin & Ayala, 2004; Martin & Gupta, 2004).

Although improvement on other related tasks is expected, the degree of change might vary depending on the task and its similarity to the training task. We hypothesized that the delayed repetition treatment should increase verbal STM capacity (with verbal span as the outcome measure). We hypothesized further that this increase should result in greater capacity for temporary activation of words during word processing tasks and should lead to improved performance on those tasks that are similar to the treatment task (i.e., repetition) as well as transfer to better performance on tasks that engage the same pathways as repetition (e.g., auditory comprehension of words). In addition to these potential near-transfer improvements, we could also observe some far-transfer effects to tasks that engage some of the pathways required in the treatment task (e.g., output pathways engaged in word production and narrative discourse).

The present study.

Evidence indicates that repetition-based verbal STM treatments can lead to improved language abilities for some cases (for reviews see, Majerus, 2018; Minkina et al., 2017; Salis et al., 2015). However, the scope of language impairment/ability profiles that would benefit from this treatment is not known. Accordingly, we included several aphasia profiles: Broca, Wernicke, Conduction, Anomic, and Transcortical Motor. Also, five participants were classified as nonfluent and three as fluent, and severity ranged from mild to moderate levels of impairment (Aphasia Quotients on the *Western Aphasia Battery-Revised* ranging from

69.1 to 88.9). Our specific predictions for this ‘repetition after response delay’ treatment were that it would lead to improved performance on: (1) the treatment task (repeating word sequences after a response delay), (2) near-transfer tasks untreated repetition and verbal STM span tasks, and (3) far-transfer tasks, including (1) picture naming (*Boston Naming Test*; *BNT*, Kaplan, Goodglass, & Weintraub, 1983) and for one participant, the *Philadelphia Naming Test (PNT)*; Roach, Schwartz, Martin, Grewal, & Brecher, 1996) and (2) relevant content and efficiency in discourse (Nicholas & Brookshire, 1993). We chose these measures because of the evidence that verbal short-term or working memory is associated with naming (Martin & Ayala, 2004) and repetition (Martin, Saffran & Dell, 1996) in aphasia, and with discourse measures in people with traumatic brain injury (Youse & Coelho, 2005), older individuals (Salis, 2011) and people with aphasia (Cahana-Amitay & Jenkins, 2018; Salis, 2012).

Finally, we emphasize again two features of this study that depart from our previous efforts to improve language abilities with a short-term activation maintenance treatment (e.g., Kalinyak et al., 2011). First, in previous studies (e.g., Kalinyak-Fliszar et al., 2011; Kalinyak-Fliszar, Kohen & Martin, 2012), we coupled the short-term activation maintenance treatment with feedback in the form of a cueing hierarchy to facilitate the accuracy of the response. The present study included no such feedback. Second, we minimized the item-repetition confound that is inherent in a repetition based STM task. To do this, we used unique words to the extent possible for treatment stimuli as well as probes used in baseline, treatment and follow-up phases of the intervention. These modifications were intended to restrict potential sources of change in performance following treatment to the short-term activation maintenance component of repetition.

Methods

Participants.

Eight individuals with aphasia participated in this study including five males and three females. All participants had incurred a left CVA and were at least one year post-onset. They presented with aphasia according to the *Western Aphasia Battery-Revised (WAB-R)*, Kertesz, 2006). All participants were right-handed, and English was their first language. The mean age of the group was 57.88 years ($SD = 8.41$) and mean educational level was 13.63 years ($SD = 3.58$). Biographical information about the participants is shown in Table 1.

All participants voluntarily enrolled in this research program and signed a consent form that was approved by the Institutional Review Board at Temple University. Testing and treatment took place in a testing room in the Eleanor M. Saffran Center for Cognitive Neuroscience at Temple University.

Language and verbal STM abilities of participants with aphasia.—Table 2 includes details of the participants’ language and verbal STM abilities based on selected measures from the Temple Assessment of Language and Verbal Short-term Memory in Aphasia (TALSA), an assessment battery developed in our laboratory (description and normative data available in Martin, Minkina, Kohen & Kalinyak-Fliszar, 2018). We also present the average scores on these tests by five neurotypical participants who are part of an

ongoing collection of normative data from the TALSA. The average age (60.00 years, $SD = 10.54$) and education (mean = 13.33 years, $SD = 2.3$) of the controls are matched roughly in age and education to the participants with aphasia (noted above). We selected subtests from the TALSA battery that assess input lexical-phonological and lexical-semantic processing of single words as well as verbal STM span, because the treatment task, repetition after a response delay, draws from these abilities. Two tasks used to assess input lexical-phonological ability were Phoneme Discrimination and Rhyming Judgments (minimal pair judgments). Two tasks used to assess input lexical-semantic ability were Lexical Comprehension (word-to-picture matching) and Category Judgments (same or different judgment of word pairs and picture pairs). Although these tests are typically administered under two conditions that vary the memory load of the task (low and high), we present only the data from the low memory load condition of the tasks to provide a background of input phonological, lexical and semantic processing ability of the participants with aphasia and control participants.

Table 2 also shows data from the digit and word repetition span subtests of the TALSA to provide an estimate of the verbal spans of the participants. The first version of the TALSA test has been used in other studies from our laboratory (Kalinyak-Fliszar et al., 2011) and is described in detail with normative data in Martin and colleagues (2018). These background data from the TALSA test (Table 2) come from a second version of the TALSA that includes the same content as the original but is longer. We are currently running participants on this version and have not yet established normative data. However, the raw scores on each subtest can be compared to the mean scores from the five control participants to provide an estimate of semantic and phonological abilities of each participant with aphasia relative to neurotypical speakers. Controls scored at or near ceiling on lexical-semantic and lexical phonological tests and their repetition span is at or near the maximum span length (7 items) tested (average digit span 7.0, average word span 6.4).

These data provide estimates of lexical-semantic and lexical-phonological abilities of the eight participants with aphasia relative to control performance. CN, EH, KM, UP and XH show relatively good performance on both lexical-semantic and lexical-phonological tests with scores close to or exceeding .90 correct. HE demonstrates lower scores on phonological tests that use nonwords. EC and KC demonstrate relatively low scores on lexical-semantic tests and lexical-phonological tasks. These variations in scores are to be regarded only as trends towards better or worse performance on a task until a larger sample is available to establish normative data.

For participants with aphasia, the average repetition span for digits in this study was 3.23 ($SD = .74$) and for words, was 3.00 ($SD = .55$). These means are considerably lower than spans of the control participants digits: 7.00 ($SD = .00$) and words: 6.2 ($SD = 1.03$). Digit and word spans for EH and KC are less than 1 SD from the mean for this group of people with aphasia. XH's digit span is within 1 SD below the aphasia mean also, but his word span is more than 2 SD s below the mean. UP, who scores within or near the control range on the single word semantic and phonological processing tests, also shows a reduction in digit and word span over 2 SD s below the mean for control participants. These data illustrate the

pervasive reduction in measures of verbal STM in people with aphasia (e.g., Martin & Ayala, 2004).

Treatment stimuli development.

We used concrete or abstract words as treatment stimuli as well as for two laboratory-developed pre- and post-treatment measures, which are described below. The words were drawn from a listing of words rated for concreteness (Brysbaert, Warriner, & Kuperman, 2014). Only nouns that were 1-, 2-, and 3-syllable words were used. We eliminated homophones and other words that we considered to be “slang” words (e.g., phooey). The Brysbaert et al. (2014) listing was also used to set criteria for classifying stimuli as concrete or abstract. A criterion of .75 standard deviations from the mean was set to identify concrete and abstract words. Based on a rating scale of 1 to 5 (least concrete/most abstract to most concrete/least abstract), the mean rating for these words was 3.53 with a SD of 1.02. The ratings of concrete words used in this study ranged from 4.29 to 5.0 and those of the abstract words ranged from 2.77 to 1.19.

Once a corpus of words was identified that met the concrete/abstract ratings criteria, we developed the word strings (i.e., singlets, pairs, and triplets). The following criteria were set:

1. Words within a pair or triplet were not semantically related (i.e., were not members of the same semantic category (e.g., *apple orange*) and they were not associated by category membership (e.g., *reptile turtle*) or by function (e.g., *scissors paper*).
2. Additionally, the words within the pairs and triplets did not share the same initial phoneme, final phoneme or stressed vowel.
3. Only 2-syllable and 3-syllable words were used to construct the singlets, pairs, and triplets. Word pairs were either five or six syllables long. Pairs with five syllables were balanced with combinations of 2-syllable + 3-syllable and 3-syllable + 2-syllable words. Pairs with six syllables were always two 3-syllable words. Triplet strings were seven or eight syllables in length and were balanced with combinations of 2-syllable and 3-syllable words (seven syllables: 2,2,3; 2,3,2 and 3 ,2, 2 and eight syllables 3, 3, 2; 3, 2,3 and 2, 3,3).
4. We did not control for frequency or phonotactic probability of words within pairs or triplets for the treatment stimuli. This decision was made because of the large number of stimuli needed and the overriding intention to preserve, as much as possible, uniqueness of stimuli in pairs throughout the treatment protocol (see below for details).

Examples of singlets, pairs and triplets are listed below:

1. concrete singlet: *dinosaur*
2. abstract singlet: *perspective*
3. concrete pair: *salmon gazebo*
4. abstract pair: *renaissance multitude*

5. concrete triplet: *apple tuxedo porcelain*
6. abstract triplet: *levity access inclement*

Minimizing repeated exposure of words in baseline, treatment stimuli and pre- and post-treatment probes.—To minimize effects of item priming by repeated

exposure, we used new stimuli (as much as possible) in baselines, probes and training sets within cycles. It was not possible to have unique words throughout because our criteria of concreteness and syllable length limited the word pools we had to work with. Nonetheless, we achieved low levels of repeated exposure overall and no repeated exposure of stimuli within the baseline or post-treatment probe trials. The following are the criteria and counts of repeated exposure:

1. If words appeared more than once, they appeared in an early phase and then in another later phase of treatment. For word pairs and word triplets, the second appearance of the word was always combined with different words in the pair or triplet. For example, if the word “origin” appeared in an abstract triplet “origin, rumor, sigma,” it did not appear again until probe 11 as “limbo, origin, density.”
2. No items appeared more than once within baseline probes.
3. No items appeared more than once within the set of post-treatment probes.
4. Within training sets, no items appeared more than once within a session. Some items did appear again in another session that occurred later in the course of treatment.

Items used in baseline, sometimes appeared again in post-treatment probes, but, as noted in (2) and (3), only appeared once within the baseline probes and once within the post-treatment probes.

The total number of words used to make up pairs and triplets used in baseline, within-treatment and post-treatment probes was 3596. Of these, 1477 (.41) appeared in baseline probes and once again in the treatment stimuli or post-treatment probe stimuli. These repeats occurred more often for 3-syllable words (.47) than 2-syllable words (.32) because there were fewer of the former in the corpus of words. Although the overall proportions of recurrence seem somewhat high, Table 3 shows that recurrence of individual word stimuli across all phases of the treatment ranges from 1 to 5 exposures. These rates of recurrence varied depending on syllable length (both concrete and abstract words), with more recurrences for 3-syllable words (up to 2–3 times) than 2-syllable words (up to 1–2 times).

Pre- and Post-treatment assessments and outcome measures

The assessments and outcome measures described below relate to our hypothesis that improvements on the treatment task would increase the strength and endurance of activation of word representations and thereby increase verbal STM capacity (as evident in increased span). We have included near-transfer tasks: repetition of single word and word sequences and verbal span, as well as far-transfer tasks: naming and discourse. These are described below.

Near transfer outcome measures.

Concrete-Abstract (C-A) Word & Word Sequences Repetition Test.: Before starting treatment, participants completed a repetition task with three conditions varying the STM demands by using different response delays (i.e., 1-sec, 5-sec, and 10-sec). The stimuli in this repetition task were concrete and abstract single words, word pairs, and word triplets. The frequency of words within each sequence-length condition ranged from 0 to 25 (based on SUBTLEXus; Brysbaert, & New, 2009). There were conditions varying stimulus type, string length and response delay. There were equal numbers of concrete and abstract stimuli within each string length (singlets, 15 each, pairs, 10 pairs each and triplets, 10 pairs each). No items were used more than once within the repetition test, but stimuli were identical before and after treatment.

We administered this task before treatment with two purposes in mind. First, it served as a screening test to help identify the appropriate treatment condition for each participant, i.e., word stimuli, sequence length and response delay time that would be used in treatment (see details in section Assignment of stimulus and interval conditions to each participant). Second, performance on this test served as a pre-treatment assessment (proportion correct in each condition) that would be compared to performance on this same task following treatment.

C-A Immediate Serial Recall Span Test.: This test was developed in our laboratory. It includes word strings ranging from one to seven words with ten trials at each string length for concrete words and for abstract words. Words were categorized into concrete and abstract groups based on the concreteness ratings (Brysbaert et al., 2014) and criteria used for all word stimuli in this study (see Treatment stimuli development above). Frequency values of the words in the span task were obtained from the same source (Brysbaert et al., 2014). The mean frequency of the abstract words (n=280) was 17.50 per million and for the concrete words (n=280), 14.25 per million.

Presentations of the concrete and abstract strings were blocked. Word strings were developed so that each trial within each string length contained 1-to-3 syllable words that did not share initial or final phonemes and no primary stressed vowels. This span task was included to determine if the treatment had any effect on verbal span capacity.

Far transfer outcome measures.

Picture Naming: As naming is a common impairment in aphasia and should depend in part on stability of activation maintenance abilities that are targeted in this treatment, we looked at performance on a frequently used naming test, the 60-item *Boston Naming Test (BNT)*, Kaplan, Goodglass & Weintraub, 1983). The *BNT* was administered to all but two people before and after treatment. For one participant, EH, the test was administered only before treatment (she was unavailable for post-treatment on this test). Also, due to an administration error, another participant, XH, was administered the 175-item PNT (Roach et al., 1996) instead of the *BNT*.

Informativeness and Efficiency of the Connected Speech of Adults with Aphasia

(Nicholas & Brookshire, 1993; N & B Narratives).: This measure was used to assess discourse content and efficiency, which could potentially be improved if the ability to maintain activation of words is increased. We evaluated four measures of discourse: words/minute, correct information units (CIUs)/minute, percent CIUs (total CIUs/total words), and proportion of mazes (e.g., false starts and filled pauses). This task is the most distantly related to the training task in terms of shared specific input and output word processing pathways. And yet, the basic skill that is targeted in the repetition plus response delay treatment is the short-term maintenance of activated linguistic representations. If this basic ability improves, it is possible that discourse efficiency could also improve. Better ability to maintain activated linguistic representations may lead to production of more words and CIUs and potentially more efficient production (higher percent CIUs, and fewer mazes).

Administration schedule of pre- and post-treatment assessment.—The *C-A Immediate Serial Recall Span Test* and *C-A Word and Word-Sequences Repetition Test* were administered within two weeks before the start of treatment. The spans were administered first and presented randomly (either concrete or abstract first) in a single session. The 18 conditions of the *C-A Word and Word-Sequences Repetition Test* (single words, pairs, and triplets) were administered across three sessions, each approximately one hour in length. Six conditions were pseudo-randomly presented, such that all triplets were not presented in a single session. Test stimuli were presented orally by the examiner. To ensure consistency of administration timing, an E-prime program (Schneider, Eschman & Zuccolotto, 2002; *Psychology Software Tools, Inc.*, 2012) was developed to guide the clinician's stimulus presentation rate of the pairs and triplets as well as the timing of the response delays.

The *BNT* and the Nicholas and Brookshire Narratives (*N & B Narratives*) are a standard component of our background testing for all new participants in our research program, and thus, administration of these language measures before initiation of treatment was not as controlled as that of the *C-A Single Word and Word Sequences Repetition Test* (which were developed specifically for this treatment study). In order to ensure that administration of near and far transfer tasks obtained before and after treatment be relatively comparable, we only used data from those individuals who completed the *BNT* and the *N & B Narratives* up to six months before treatment was initiated. Five of the eight participants met these criteria (HE, KM, UP, KC and XH). For these five participants, the mean times in months (and ranges) before initiation of treatment were as follows: *BNT* 3.3 (SD 1.96, range = 1 to 6.4), *N & B Narratives* 3.1 (SD = 1.75, range = 1.5 to 6.0). The other three participants, CN, EC and EH were not included in analyses of these two far transfer tasks. All participants were in chronic stages of aphasia and had not received intervention for their aphasia during the gaps in time between assessment and initiation of the treatment reported in this study.

Scoring of repetition responses.: The following guidelines were used to score repetition responses in baselines, treatment, probes, and follow-up as well as in the pre-treatment word, pair and triplets test and the concrete and abstract spans: Repetition responses were scored as correct if they were produced without error. Articulatory distortions (but no distorted substitutions), regional pronunciations (e.g., water, /wɑtə/, typically pronounced

as /wɒdər/, “wooder”, in the Philadelphia region), and sound productions typical of African American Vernacular English (e.g., reduction of final consonant clusters) were permitted.

Post-treatment testing schedule.: Post-treatment assessments were administered in the following order: *C-A Immediate Serial Recall Span*, *Immediate C-A Single word and Word Sequences Repetition Test*, *BNT & PNT* and *N & B Narratives*. They were all administered after the immediate and one-week treatment probes and within two months following treatment (ranging from 2 weeks to 2 months).

Assignment of stimulus and interval conditions to each participant.: (Table 4). For this study, we did not select participants based on a known deficit in maintaining activation of word representations. Rather, we enrolled participants based on the presence of a moderate-to-mild aphasia (according to the *WAB-R* criteria). Current evidence suggests that the ability to maintain activation of word representations sufficiently to support single word and multiple word speaking tasks is impaired in some but not all individuals with aphasia (Martin & Dell, 2017). Nonetheless, it is not yet determined if a treatment that aims to improve the ability to maintain activation of word representations would be effective only with those who have an impairment of this ability. It is also possible that it would improve language abilities regardless of the presence or absence of a specific activation maintenance deficit.

Determining the appropriate stimuli for treatment was based in part on the participant’s performance on the pre-treatment test of repeating concrete and abstract words, pairs and triplets after three response delay conditions. The results helped to identify the combination of stimulus type (concrete or abstract), sequence length (one, two or three words) and response delay condition (1 sec, 5-sec, 10 sec) that was difficult for the participant and yet potentially could be improved with treatment. For selection of the word type, sequence length and response delay combination, our criterion was a score between .10 and .40 sequences repeated correctly in serial order. This allowed opportunity for improvement from baseline. A general rule was to choose a stimulus-response delay condition at which a decline from the 1-sec response delay condition occurred. Sometimes, there was more than one possible choice, and in these cases, we chose the condition with less accurate performance, where there seemed the most room for improvement. Table 4 shows the assignments of these conditions to each participant.

Treatment

Treatment design.—A case series design was followed, with treatment protocols for individual cases structured as a single subject ABA design with maintenance testing (A1 = Baseline, B = Treatment, A2 = Post-treatment and Follow-up). As described in the Methods, the words used in all phases of the intervention (baseline, treatment, post-treatment and follow-up) were unique as much as possible. By minimizing effects of item repetition, we eliminated the need for a set of untrained items tested before and after treatment. Only nouns were used in the stimulus set, but it should be noted that the principles of this treatment could be applied to repetition of other word classes (e.g., verbs) or repetition of phrases in future studies. Also, the primary manipulation, minimizing repeated exposure of items,

made it difficult to control for frequency and phonotactic probability of words within pairs and triplets. Finally, we did not include a control task in this study. As the treatment is intended to target a verbal maintenance process that is, to some extent, fundamental to most language tasks, it is difficult to identify a language task that could serve as a control. These limitations are addressed in the Discussion section.

Treatment schedule.—Treatment intensity was varied (two or four sessions per week) in part to accommodate participants' schedules. Thus, treatment was scheduled for four sessions per week over three weeks for some participants (CN, EH, UP, XH) and for others it was scheduled for two sessions per week over six weeks (KC, EC, HE, KM).

Treatment protocol—Figure 1 shows the overall treatment protocol (baseline, treatment, treatment probes, post-treatment probes).

Treatment task.: The treatment task was repetition of word stimuli after a response delay that was selected specifically for the participant, as noted above: concrete or abstract words, (singlets, pairs or triplets) after a 1-sec, 5-sec or 10-sec response delay. A sample of a treatment session using abstract word pairs as stimuli is shown in Figure 2.

Baseline.: Participants received a minimum of three baselines (20 word strings each) over one week. Criterion for stable performance in baseline was set at no greater than 15 percentage point change among baselines. If this criterion was not met, additional baselines were administered.

A treatment session consisted of repeating a set of 40 single words or word sequences in the condition assigned to the participant. These were presented in two 20 item sets with a 2-minute break in the middle of each set and a 10-minute break between the two sets. Treatment probes (20 word strings) in the assigned condition were administered at the beginning of each treatment session, starting with the second session. Each treatment probe consisted of unique words of the same type as the treatment stimuli and administered in the same response delay condition.

Post-treatment probes.: Three post-treatment probes were scheduled: 24-hours, 72 hours and one week following completion of treatment. This schedule was observed as closely as possible, but there were exceptions when the timing of the final treatment session led to the 24-hour or 72-hour probe occurring on a weekend. In addition to these post-treatment probes, a maintenance probe was administered one month following treatment completion.

Statistical analyses.

Analysis 1. Effects of a repetition-based verbal STM treatment (i.e., delayed recall of multiple words) on the treated task.—We calculated effect sizes for each participant for the number of strings he/she produced in serial order and in any order from (1) pre-treatment to post-treatment probes and (2) pre-treatment probes to maintenance probes. For the first effect size measure, we subtracted the mean of the pre-treatment probe scores (baselines) from the mean of the post-treatment probes (immediate recall, 24 hour, 72 hour and 1 week probes). This value was then divided by the standard deviation of the pre-

treatment measures. For the second effect size measure, we subtracted the mean of the pre-treatment probes from the 1 month probe score and again divided the number by the standard deviation of the pre-treatment measures. As there are no estimated effect size benchmarks for the ‘repetition with response delay’ task used in this study or for treatments that minimize repeated exposure of verbal stimuli, we used effect sizes estimated for alexia and agraphia reported in Beeson and Robey (2006) rather than those for lexical retrieval. The rationale for this choice was twofold. First, repetition involves both input and output processing and is in this respect more similar to reading processes than lexical retrieval. Second, because this was a novel treatment approach for which benchmarks do not exist, we used smaller effect size criteria to decrease the possibility of type II errors in this early phase treatment testing. While later stage studies may warrant more conservative benchmarks, earlier phase studies warrant moderately liberal error tolerance to allow investigators to detect existing treatment effects in smaller-scale studies (Robey, 2004). Effect sizes >2.6 were considered small, those >3.9 were considered medium, and those >5.8 were considered large (Beeson & Robey, 2006).

To assess reliability, 15% of probe responses were scored for word string accuracy ‘in serial order’ by the original scorer and a second scorer, and Cohen’s Kappa was computed (Landis & Koch, 1997).

Analysis 2. Effects of a repetition based verbal STM treatment on the C-A Single Word and Word Sequences Repetition Test.—To assess reliability, 15% of screening form responses were scored for accuracy in serial order by the original scorer and a second scorer, and Cohen’s Kappa was computed (Landis & Koch, 1997).

The proportions of correct and incorrect repetitions of single words, pairs and triplets were reported for all concrete and abstract stimuli at the 1-sec, 5-sec, and 10-sec response delay conditions. To determine if the treatment affected performance on a specific stimulus-response delay condition, we compared accuracy on repetition of word stimuli (singlets, pairs and triplets) in each response delay condition before and after treatment. We used the log odds ratio statistic, its standard error, and confidence interval (Altman, 1991) to calculate the probability of change in accuracy (in either direction) and with significance of that change determined according to Sheskin (2004). Log odds ratios are used in single case studies to compare performance before and after a treatment but are not ideally used to compare effect sizes of one case to another Fischer-Baum (2015). Our primary focus in this analysis is to compare pre- and post-training performance on assessments within a person to determine if the changes are significant.

Analysis 3. Effects of a repetition-based verbal STM treatment on the C-A Immediate Serial Recall Span Test.—Wilcoxon Signed Rank Tests were used to determine whether there were differences in performance on either or both concrete and abstract spans before and after treatment. Changes in span were noted for items reported in serial order (ISO) and in any order (IAO). While both the ISO and IAO scores indicate improvement of item recall, the ISO score indicates improved recall of order information as well.

Analysis 4. Effects of a repetition-based verbal STM treatment on picture naming (BNT, and for participant XH, Philadelphia Naming Test, PNT).—Six of the eight participants' picture naming data were obtained within six months of the initiation of treatment. For these individuals, we conducted a case by case analysis comparing the difference in proportions of correct responses before and after treatment using the log odds ratio statistic, standard error and confidence interval to determine the probability of change in naming performance after treatment (Altman, 1991) and determined the significance of that change according to Sheskin (2004).

Analysis 5. Effects of a repetition based verbal STM treatment on the N & B Narratives.—We evaluated pre- and post-treatment discourse production using the stimuli developed by Nicholas and Brookshire (1993). These stimuli are made up of ten discourse prompts which include a combination of personal narratives (2), single picture descriptions (4), sequential picture descriptions (2) and procedural discourse (2). For each participant, discourse production was audio recorded and then orthographically transcribed into Systematic Analysis of Language Transcripts software (Miller & Iglesias, 2012) by trained research assistants. Transcriptions included words, pauses of greater than two seconds, and mazes. Mazes were defined as filled pauses (e.g., uh, um) and false starts. Once transcribed, discourse samples were coded for words and CIUs, defined as words that are accurate, relevant to the content and not repeated (Nicholas & Brookshire, 1993). The primary discourse outcome was the proportion of CIUs relative to total number of words produced in the narratives (%CIUs). This measure was selected to determine if relevant content words improved in discourse. Additionally, words per minute (WPM), CIUs/minute (CIUs/Min), and the proportion of mazes over total words were evaluated to determine if discourse efficiency improved.

Point-to-point transcription reliability was completed by trained research staff for 18.5% of transcripts, which resulted in 94.6% reliability. Reliability was calculated by dividing total correctly transcribed items by the total possible. Point-to-point coding reliability was completed for words and CIUs on 29% of transcripts. Coding reliability for words was 98.6% and was calculated by dividing total agreed upon words by total possible. Reliability for CIUs was calculated similarly with total agreed upon CIUs divided by total possible with inter-rater reliability of 88.69%.

Results

Analysis 1. Effects of a repetition-based verbal STM treatment (i.e., delayed recall of multiple words) on the treated task (Figures 3–10)

Reliability.—A Kappa coefficient of .80 (substantial agreement) was found for intra-rater scoring of string accuracy in serial order and .67 (substantial agreement) for inter-rater scoring of string accuracy in serial order (Landis & Koch, 1977).

Visual Analysis of acquisition and post-treatment probe data.—Figures 3 through 10 show the probe data from the baseline, treatment (acquisition) and post-treatment phases. Two measures were tracked in the probe analysis, proportion of strings correct and proportion of words correct. Each of those measures was tracked additionally for accuracy

of strings in serial order (ISO) and in any order (IAO). As the figures show, some individuals had more difficulty retaining the order of words in a string (CN, EC, EH, KM) and others not so much (especially HE and KC). Another notable feature of these probe data is that acquisition is overall modest, though small gains are noted in proportion of words correct IAO for participants CN, HE, UP and XH. In the post-treatment probes, four participants show an increase in accuracy for string and/or words in probes at 24 hours, 1 week and/or one month after treatment. This interesting pattern of further increases in accuracy in the post-treatment follow up probes will be addressed in the Discussion.

Changes in pre- to post-treatment probes: Effect sizes. (Table 5)—We determined effect sizes for four measures obtained at baseline, the immediate post-treatment probes and the one-month follow-up probe: Proportion of strings (singlets, pairs or triplets) correctly produced in serial order (ISO) and in any order (IAO), as well as the proportion of correct items produced ISO and IAO. Three participants showed small to medium effect sizes for some of the above measures. KM showed a medium effect (4.04) for strings ISO and a small effect (2.84) for strings IAO. HE showed a small effect (3.18) for total correct items ISO. Lastly, KC showed a small effect (2.62) for correct items IAO.

Of the participants who demonstrated small and moderate effect sizes at the immediate post-treatment probe, only KC retained this effect size one month later. However, CN, who showed no significant changes immediately post-treatment, and then demonstrated better performance on three of the four measures at the one-month probe point: A small effect of proportion of strings correct in serial order (reflecting improved recall of item and order information), a large effect of word recall in any order (reflecting improved item recall), and a small effect of proportion of words repeated accurately in any order.

Analysis 2 (Table 6). Effects of treatment following a response delay on the C-A Single Word and Word Sequences Repetition Test

Reliability.—For scoring of the immediate post-treatment probes, we found a Kappa coefficient of .89 (almost perfect agreement) for intra-rater scoring of accurate repetition of words and word strings in serial order as well as a Kappa coefficient of .85 (almost perfect agreement) for inter-rater accuracy on scoring of these same stimuli (Landis & Koch, 1977).

Pre- and post-treatment comparison of accuracy.—We first calculated the proportion of correct repetitions (pre- and post-treatment) for each stimulus type (concrete and abstract single words, pairs and triplets) x string length (singlet, pair, triplet) x response delay condition (1-sec, 5-sec, 10-sec). We then used the log odds ratio, its standard error and confidence interval to calculate the probability of change in proportion correct for each stimulus type after treatment (Altman, 1991), with the significance of that change calculated according to Sheskin (2004).

Of the eight participants, five showed one or more instances of significant change (four improved and one getting worse). The changes are shown in Table 6.

CN improved on abstract word pairs after a 10-sec response delay and showed a trend for improvement ($p = .07$) for concrete word pairs after a 10 second response delay. KM

improved on concrete word pairs after 10 seconds. UP improved on concrete triplets after a 1-sec response delay and abstract pairs after a 10-sec response delay. XH also showed improvements on two stimulus conditions, concrete and abstract single words after a 5-sec response delay. EC demonstrated a trend of improvement in repetition of concrete pairs after a 1-sec response delay. Finally, HE showed a significant decline in accuracy for repetition of abstract single words after a 1-sec response delay.

With each participant repeating 18 word sequences, this may seem to be relatively few instances of significant improvement after the response delay. However, it is important to note that for most of the participants, repetition of single words was at ceiling, and repetition of the word triplets was near floor levels. For single words, the mean was high at .83 (SD = .14) correct (median = .87, range = .40–1.00). For word pairs, the mean was .40 (SD = .23) correct (median = .40, range = 0.00 – 1.00). For word triplets, the mean was quite low .09 (SD = .14) correct (median = 0.00, range = was 0.00 - .60). Thus, we would not expect much change in comparisons when performance was near ceiling or floor.

Analysis 3 (Table 7). Effects of a repetition-based verbal STM treatment on the C-A Immediate Serial Recall Span Test.

As Table 7 shows, span for concrete words increased for eight participants (from .20 to .60) and span for abstract words increased for five participants (from .20 to 1.00). Increases greater than .50 were observed for three participants, EC, KM and XH. These observations indicate qualitative improvements in span capacity for some of the participants. A group analysis using the Wilcoxon Signed Rank Tests provides quantitative support for these improvements. There were statistically significant increases in median (Mdn) span (0.30) for concrete words ISO: pre-treatment $Mdn = 2.70$ and post-treatment, $Mdn = 3.00$ ($z = -2.39$, $p = .017$). No significant differences were observed for the following pre- and post-treatment comparisons: concrete IAO span, abstract ISO span and abstract IAO span.

Analysis 4 (Table 8). Effects of a repetition-based verbal STM treatment on picture naming (BNT (Kaplan et al., 1983), and for participant XH, Philadelphia Naming Test, (PNT, Roach et al., 1996)).

Table 8 shows the pre- and post-treatment scores on the *BNT* for HE, KM, KC, and UP. There are two notes of protocol error: EH did not complete the *BNT* at post-testing and XH was administered the *PNT* instead of the *BNT*. There were small increases in accuracy on the *BNT* following treatment for some participants. We conducted individual analyses comparing differences in proportion correct before and after treatment (using log odds ratios and z tests). Only XH's change in performance (increase of .11) was significant (odds ratio = .62, 95% confidence interval = .41 to .95, $z = 2.18$, $p = .029$).

Analysis 5 (Table 9). Effects of a repetition-based verbal STM treatment on Nicholas & Brookshire (1993) discourse data.

For the discourse samples, we calculated proportions of CIUs relative to total number of words, words per minute (WPM), CIUs per minute (CIUs/min), and proportions of mazes that were produced before and after treatment (Table 9). Six of the eight participants completed all ten N & B prompts before and after treatment; however, a data collection error

occurred for two participants (EH and KC). Thus, only one discourse elicitation prompt was completed at post-treatment for those two participants. Also, two participants, CN and EC, completed the N & B discourse samples around a year prior to initiation of treatment (14 months for CN and 11 months for EC). Although they did not participate in interventions during that period, any gains observed in content and efficiency could not be attributed to the treatment and their data were removed from the analysis.

A meaningful increase in percentage of CIUs was interpreted as a pre- to post-treatment change larger than twice the standard error of measurement (SEM) reported by Brookshire and Nicholas (1994), based on the performance of 20 people with aphasia from first to third administration of the N & B stimuli (1993). A change of greater than twice the SEM (4.2%) was considered improvement that could be attributed to treatment (Edmonds et al., 2014; Wambaugh & Ferguson, 2007). Using this benchmark, three participants demonstrated improvement in percentage of CIUs, which indicates an increase in relevant content words in relation to total words.

Participant KM produced an average of 44.44% CIUs before treatment. Based on a doubling of the SEM, it would be predicted that KM's post-treatment percentage of CIUs would fall between 40.24% and 48.64%. At post-treatment, KM's percentage of CIUs increased to 54.95% which exceeded what would be predicted due to repeated administration of the stimuli (Brookshire & Nicholas, 1994).

Participant UP's average percent of CIUs was 44.27% at pre-treatment. At post-treatment, UP's percent CIUs increased to 54.96%. This change was greater than twice the SEM (40.07% – 48.47%).

At pre-treatment, participant XH produced an average of 25.36% CIUs. XH's percentage of CIUs improved by greater than twice the SEM, to 31.86%, suggesting an increase in relevant content words in discourse.

To assess improvements in discourse efficiency we evaluated WPM, CIUs/Min, and the proportion of mazes (% Mazes). A change of greater than twice the SEM was used as a benchmark to indicate improvement on WPM and CIUs/Min. Therefore, a change of 11 WPM (Brookshire & Nicholas, 1994) or a change of 8 CIUs per minute (Nicholas & Brookshire, 1993) was interpreted as an increase from pre- to post-treatment (Edmonds, et al., 2014; Wambaugh & Ferguson, 2007). Two participants made improvements in efficiency, KC and HE. KC increased in WPM from 53.78 at pre-treatment to 74.42 WPM at post-treatment; however, these results should be interpreted with caution due to the sampling error at post-treatment (i.e., only one discourse sample was collected). HE also improved in discourse efficiency as indicated by a large reduction in percent mazes in discourse from pre- (34.65%) to post-treatment (17.18%).

Discussion

In the context of this case series treatment study, we aimed to determine if a repetition based verbal STM intervention that minimized repeated exposure to individual treatment stimuli but preserved its verbal STM component (repetition of unique word stimuli after a time

delay), would effectively increase verbal STM capacity and improve performance on measures of language function. This design departs from traditional single subject treatment designs, which often include a fixed set of items for training and a fixed control set of items that are not trained. Howard, Best and Nickels (2015) note that repeated probing of training items can lead to direct influences of the probing itself in addition to effects of the treatment. In the context of an IA model of word processing, the rationale for this design was that repeated exposure to a finite set of semantically or phonologically unrelated word stimuli would increase the level of activation of this cohort of word representations (e.g., Martin, Fink, Laine, & Ayala, 2004) relative to other related words in the lexical network. Consequently, this imbalance of activation favoring the trained items could result in limited generalization to untreated items. By eliminating the repeated exposure of specific items in the treatment task, we intended for repetition of verbal stimuli after a response delay to provide practice in activating and maintaining representations of words over a time delay and improve the short-term maintenance component of word processing. Thus, the design was intended to identify a potential contribution of STM processes to a 'repetition with response delay' treatment.

Although the data provide some support for using minimal item exposure in a repetition-based verbal STM treatment for aphasia, for some participants, no appreciable changes in performance were observed. On the treatment task, only four participants (KM, HE, KC, CN) showed effect sizes ranging from small to large post-treatment probes. Results of the outcome measures provided more evidence that this treatment had positive effects on performance of language and verbal STM tasks for some participants.

Table 10 summarizes the outcomes of our pre- and post-treatment language and verbal STM measures for each person. While some participants show no changes following this treatment (e.g., EC and EH), others showed improvements in some, but not all outcome measures. Three participants (KC, UP and XH) improved significantly on repetition of concrete or abstract single words and word pairs. On the verbal span task, concrete word spans increased for all participants and abstract word spans increased for five participants.

Individual differences are apparent on the picture naming tests (*BNT*, *PNT*) and discourse measures as well. For picture naming, the only participant who showed any significant improvement following treatment was XH. For the narrative discourse task, individual differences also were present. The proportion of CIUs increased for three people (KM, UP, XH) and this difference was greater than our criterion of two times the SEM reported by Brookshire & Nicholas (1994) from pre- to post-treatment. Also, KC showed an increase in words per minute and HE showed a reduction in the % Mazes score.

What can we learn from these outcomes? STM treatments that involve repetition of single and multiple word sequences with or without a time delay involve three processing steps: Access to, short-term maintenance of, and retrieval of semantic and phonological representations of words. In typical single subject treatment designs, another factor is introduced, repeated exposure of a specific set of items. This creates a confound for our hypothesis that increasing short-term retention load by adding a response delay to the repetition task improves verbal span capacity that in turn improves language function.

Related to this hypothesis, Nickels (2002) demonstrated that for some people with aphasia, naming accuracy improves with repeated attempts and without feedback. Recently, Creet, Morris, Howard and Nickels (2019) showed that for some people, repeated naming improved accuracy even if naming attempts were six weeks apart. Creet and colleagues (2019) suggest that the ability to monitor language output is one variable that contributes to this ability.

In this treatment study, we controlled for repeated exposure of training and probe items as much as possible, with the assumption that by doing so, the treatment primarily involved practice in access, maintenance and retrieval of verbal representations. The treatment yielded effect sizes for the acquisition stage (probes between baseline and post-treatment probes) of treatment that, overall, were modest. For example, although effect sizes for XH's acquisition stage and 1-month probes were small, he showed improvements on most post-treatment language measures. In contrast, effect sizes for HE and EH were significant for some measures, but no increases in CIUs were observed in narrative discourse. Another pattern of effect sizes was observed in CN's case. Although he did not show any significant effect sizes in the post-treatment probe one week after treatment, he did at the 1-month probe point. This pattern is of interest and could reflect a consolidation of learning effect (e.g., Kendall, Oelke, Brookshire, & Nadeau, 2015), but this possibility needs further investigation.

The positive effects of repeated attempts on successful naming (Nickels, 2002; Creet et al., 2019) contrast with the quite modest improvements observed in repetition accuracy when we minimized the opportunities to repeat an item more than once during the treatment protocol. Although naming and repetition are different tasks, they overlap at the retrieval stage between the lexicon and phonological network. Thus, it is possible that repeated opportunities to name or repeat a word will influence the outcome of a treatment for word retrieval impairments.

In our previous studies, when we have used repeated exposure of a specific set of 'trained items' in conjunction with the 'repetition with delay' training (Kalinyak-Fliszar et al., 2011; Kalinyak-Fliszar, Kohen, & Martin, 2012), improvements in the acquisition stage were more robust than those observed in the present study. As noted earlier, one treatment version (Kalinyak-Fliszar, et al., 2011) included a cueing hierarchy, which may have contributed to the more robust acquisition effects. Although we have not yet directly compared this 'repetition with delay' treatment with and without repeated items in the training set, it is reasonable for future studies to investigate the hypothesis that item repetition plays a role in effects of repetition-based treatments (with or without a response delay), and these effects could be observed in acquisition, post-treatment performance on the treatment tasks and generalization to non-treatment tasks.

The mechanism of generalization effects in word retrieval abilities postulated by the IA model of lexical processing (e.g., Dell et al., 1997) is spreading activation to other related words in the lexicon each time a word representation is activated. Such effects are short-term during training, but ideally, over a longer term of treatment would lead to longer lasting changes in strength of activation. In relation to the treatment reported here, this model assumption raises several research questions. As noted above, Martin et al. (2004) proposed

that the interference they observed in their contextual priming repetition treatment was due to repeated exposure to a specific set of items creating a cohort of highly activated words in the lexicon that overshadowed effects of spreading activation to other words in the lexicon. If this effect is real, would it be subject to set size, with smaller sets creating a stronger cohort of highly active items in the lexicon that cannot be surpassed by activation of competing lexical representations? Larger sets of repeated items might allow more time between repeated exposure, which could allow for activation of trained items to rise to a level that makes them more accessible, but at the same time, would allow for related words activated via spreading activation also to reach accessible levels. It is possible that a middle ground approach that included some repeated item exposure might have resulted in greater improvement following this treatment. We are currently working with a second format of this treatment that addresses the issue of degree of repeated exposure in training and probe stimuli.

In addition to generalization of effects in word processing ability, there is the related issue of transfer of improvements to closely related tasks (e.g., repetition and repetition span) and distantly related tasks (e.g., naming and discourse). The mechanism for this kind of transfer would be real change in the strength and maintenance of activation that supports word retrieval and verbal STM performance. To the extent that we observe changes in outcome measures such as verbal span, naming and discourse, we can infer that such longer lasting changes in activation strength and/or activation maintenance ability have occurred.

Another lesson from this study relates to the individual responses to the treatment. The participant sample in this study, with diverse types of aphasia and underlying neural damage profiles, showed several different types of response to this treatment. One participant, XH, showed improvements in all language measures, picture naming, and narrative discourse. He also showed increases in concrete and abstract word spans and greater accuracy on repetition of concrete and abstract words. For XH, then, this treatment influenced language function beyond the treatment task. KC and UP also showed modest increases on some discourse measures, repetition of word-sequences, and concrete and abstract word spans. Other participants showed little improvement on the outcome measures following this treatment. As it is unlikely that one treatment approach is beneficial for all types of aphasia, it is important to determine as closely as possible the neural, linguistic and cognitive profiles associated with good response to a treatment. Thus, future studies of this treatment approach to improve short-term maintenance of activated word representations should aim to identify language impairment profiles that respond to this treatment.

Limitations of this study.

In the Methods section, we mentioned several limitations in the design or implementation of this study. First, for some participants (CN, EC, KM and EH) the pre-treatment language measures (*BNT* and *N & B Narratives*) were not always administered close in time to the initiation of treatment. Although these individuals were in chronic phases of aphasia and received no interventions between the pre-treatment testing and start of treatment, we cannot confidently attribute changes observed in post-treatment language measures to the 'repetition with response delay' treatment. Thus we eliminated three participants from the

two far-transfer tasks, picture naming and narratives. Second, the design of this study did not include a control task. Selection of an appropriate control task for this treatment is challenging, as most language tasks engage to some extent the activation maintenance processes that are targeted in this treatment. However, use of a nonlinguistic control, such as the Five-Point Test (Regard, Strauss, & Knapp, 1982), would provide some control for a general learning effect that is not modality specific, and this will be incorporated into future studies. Third, treatment stimuli were not controlled for frequency or phonotactic probability. Our primary variable of concern was the uniqueness of items throughout all components of the treatment protocol (baseline, training, probes). The large number of stimuli needed to avoid repeated exposure of words in any phase of treatment made it difficult to control for these two variables. Finally, we did not test a priori whether the people enrolled in this study had a specific deficit in maintaining activation of words in the short-term. As a first pass in testing this treatment, we left open the possibility of this treatment being effective regardless of the presence or absence of an activation-maintenance impairment. These limitations are being addressed in variations of this treatment that we are currently testing.

Conclusions and Future Directions.

This study represents a first step in disentangling the effects of verbal STM (short-term maintenance of word representations) from item-specific effects in a repetition-based treatment for word processing in aphasia. Predicted transfer of treatment effects to verbal span and narrative discourse were observed in some of the participants, especially those who showed a deficit before treatment in the short-term maintenance of activated word representations (the specific target of this treatment). These outcomes suggest that for some individuals with aphasia, focusing on STM processes that support language function could be a useful treatment approach.

That a treatment may not be the right one for all individuals with aphasia is an important point that most researchers and clinicians recognize. Furthermore, as we develop our treatments, it is necessary to learn why a treatment works for some individuals and not for others. Recent guidance for development of approaches to treatment by the Medical Research Council in the United Kingdom (Craig, Dieppe, Macintyre, Michie, Nazareth, & Petticrew, 2006) notes that interventions in health services are often complex and as such, require careful and systematic evaluations that consider these complexities before they are implemented. The treatment approach reported here is a good example of a complex intervention and as such leads to as many (if not more) questions as answers. For example, it will be important to determine factors that influence who benefits from this treatment approach including type and severity of language impairment and presence of a sensitivity to increased verbal STM load. A second question concerns our minimizing the repeated exposure of probe and treatment stimuli. The acquisition effects observed in this treatment study were low compared to those in our previous studies using a similar (though not identical) treatment paradigm (Kalinyak-Fliszar et al., 2011). A more controlled treatment design is needed to determine the extent to which item-specific effects (repeated exposure to specific items) contribute to acquisition effects in repetition-based treatments. It will be important also to compare this treatment approach with another approach that does not

increase memory load in the treatment task. Such a comparison could provide insight into our hypothesis that increasing tolerance of verbal STM memory load is a key element of improvements in language abilities following this treatment. Finally, although not investigated directly in this study, this treatment should, in theory, impact functional communication abilities in a positive way. Conversational discourse involves instances in which verbal representations must be actively maintained ‘in mind’ before an utterance is produced. Investigation of this possibility is an important future goal.

Acknowledgments.

Research reported in this publication was supported by *National Institute on Deafness and other Communication Disorders Center* of the National Institutes of Health under award number R01DC013196. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We are very grateful to the participants who participated in this study.

References

- Altman DG (1991) Practical statistics for medical research. London: Chapman and Hall.
- Beeson PM, & Robey RR (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology review*, 16(4), 161–169. [PubMed: 17151940]
- Bemdt RS, & Mitchum CC (1990). Auditory and lexical information sources in immediate recall: Evidence from a patient with deficit to the phonological short-term store. In Vallar G & Shallice T (Eds.), *Neuropsychological impairments of short-term memory*. Cambridge: Cambridge Univ. Press, pp. 115–144.
- Boyle M & Coelho C (1995). Application of semantic feature analysis as a treatment for aphasic dystonia. *American Journal of Speech-Language Pathology*, 4, 94–98.
- Brookshire RH, & Nicholas LE (1994). Speech sample size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 37(2), 399–407.
- Brybaert M & New B (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41 (4), 977–990. [PubMed: 19897807]
- Brybaert M, Warriner AB, & Kuperman V (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavioral Research Methods*, 46 (3), 904–911.
- Cahana-Amitay D & Jenkins T (2018). Working memory and discourse production in people with aphasia. *Journal of Neurolinguistics*, 48, 90–103.
- Conroy P, Drosopoulou CS, Humphreys GF, Halai AD, Lambon Ralph MA (2018). Time for a quick word? The striking benefits of training speed and accuracy of word retrieval in post-stroke aphasia. *Journal of Neurolinguistics*, 48, 1815–1827.
- Cowan N (2005). *Working Memory Capacity*. Hove, East Sussex, UK: Psychology Press.
- Cowan N (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, 169, 323–338. [PubMed: 18394484]
- Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, & Petticrew M (2006). Developing and evaluating complex interventions: new guidance. Medical Research Council, UK, www.mrc.ac.uk/complexinterventionsguidance.
- Craik FI, & Lockhart RS (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11 (6), 671–684.
- Creet E, Morris J, Howard D & Nickels L (2019). Name it again! Investigating the effects of repeated naming attempts in aphasia. *Aphasiology*, 33, 1202–1226.
- Dell GS, Martin N, & Schwartz MF (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, 56, 490–520. [PubMed: 21085621]

- Dell GS (1986) A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93:283–321. [PubMed: 3749399]
- Dell GS, Schwartz MF, Martin N, Saffran EM & Gagnon DA (1997). Lexical access in aphasic and non-aphasic speakers. *Psychological Review*, 104 (4), 801–838. [PubMed: 9337631]
- Edmonds LA, Mammino K, & Ojeda J (2014). Effect of verb network strengthening treatment (VNeST) in persons with aphasia: Extension and replication of previous findings. *American Journal of Speech-Language Pathology*, 23 (2), S312–S329. [PubMed: 24687125]
- Fromkin V (Ed.) (1980). *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. San Francisco: Academic Press.
- Fischer-Baum S (2015). How much better? The challenge of interpreting interactions in intervention studies. *Aphasiology*, 29 (5), 609–614.
- Howard D, Best W & Nickels L (2015). Optimising the design of intervention studies: critiques and ways forward *Aphasiology*, 29 (5), 26–562.
- Kalinyak-Fliszar M, Kohen F & Martin N, (2011). Remediation of language processing in aphasia: Improving activation and maintenance of linguistic representations in (verbal) short-term memory. *Aphasiology*, 25 (10), 1095–1131. [PubMed: 22791930]
- Kalinyak-Fliszar M, Kohen FP, & Martin N (2012). Improving the maintenance of word representations in short-term memory to improve language function: Acquisition and generalization effects. *Clinical Aphasiology Conference*, Lake Tahoe, CA, May 20–25, 2012.
- Kaplan E, Goodglass H Weintraub S (1983). *Boston Naming Test*. Lea & Fibiger, Philadelphia.
- Katz RC & Wertz RT (1997). The Efficacy of Computer-Provided Reading Treatment for Chronic Aphasic Adults. *Journal of Speech, Language and Hearing Research*, 40 (3) 493–507.
- Kendall D, Oelke M, Brookshire C, & Nadeau S (2015). The influence of phonomotor treatment on word retrieval abilities in 26 individuals with chronic aphasia: An open trial. *Journal of Speech, Language, and Hearing Research*, 58, 798–812.
- Kertesz A (2006). *Western Aphasia Battery-Revised (WAB-R) Pro-Ed*, Austin, TX.
- Kiran S & Sandberg C (2012). Treating communication problems in individuals with communication disorders. In Peach R and Shapiro L (Eds.) *Cognition and Acquired Language Disorders: A Process-Oriented Approach* (pp. 299–325). St. Louis, MO: Elsevier Mosby.
- Kiran S & Thompson C (2003). The role of semantic complexity in treatment of naming deficits: Training categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech Language and Hearing Research*, 46, 608–622.
- Klingberg T (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14 (7), 317–324. [PubMed: 20630350]
- Koenig-Bruhin M, & Studer-Eichenberger F (2007). Therapy of short-term memory disorders in fluent aphasia: a single case study. *Aphasiology*, 21(5), 448–458.
- Landis JR, & Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. [PubMed: 843571]
- Majerus S, Van Der Kaa MA, Renard C, Van der Linden M, & Poncelet M (2005). Treating verbal short-term memory deficits by increasing the duration of temporary phonological representations: A case study. *Brain and Language*, 95(1), 174–175.
- Martin N, Fink R, Laine M & Ayala J (2004). Immediate and short-term effects of contextual priming on word retrieval. *Aphasiology*, 18, 867–898.
- Martin N (2012). Managing communication deficits associated with memory disorders. In Peach R and Shapiro L (Eds.) *Cognition and Acquired Language Disorders: A Process-Oriented Approach* (pp. 275–297) St. Louis, MO: Elsevier Mosby.
- Martin N (2016). Disorders of word production. In Papathanasiou I, Coppens P, Potagas C (Eds.) *Aphasia and Related Neurogenic Communication Disorders*, 2nd edition (pp. 169–174). Sudbury, MA: Jones and Bartlett Publishers.
- Martin N, & Ayala J (2004). Measurements of auditory-verbal STM span in aphasia: Effects of item, task, and lexical impairment. *Brain and language*, 89 (3), 464–483. [PubMed: 15120538]
- Martin N, Fink R, Laine M (2004). Treatment of word retrieval with contextual priming. *Aphasiology*, 18, 457–471.

- Martin N & Dell GS (2017). Adding temporal dynamics to models of impaired language production. Academy of Aphasia, Baltimore, MD, November 4 – 6, 2017.
- Martin N & Dell GS (2019). Maintenance versus transmission deficits: The effect of delay on naming performance in aphasia. *Frontiers in Human Neuroscience*. November 27, 2019 |10.3389/fnhum.2019.00406
- Martin N & Gupta P (2004) Exploring the relationship between word processing and verbal STM: Evidence from associations and dissociations. *Cognitive Neuropsychology*, 21, 213–228. [PubMed: 21038201]
- Martin N, Minkina I, Kohen FP & Kalinyak-Fliszar M (2019). Assessment of linguistic and verbal short-term memory components of language abilities in aphasia. *Journal of Neurolinguistics*. 48, 199–225.
- Martin N, & Saffran EM (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, 14 (5), 641–682.
- Martin N & Saffran EM (1999). Effects of word processing and short-term memory deficits on verbal learning: Evidence from aphasia. *International Journal of Psychology*, 34 (5/6), 330–346.
- Martin N, Weisberg RW & Saffran EM (1989). Variables Influencing the Occurrence of Naming Errors: Implications for Models of Lexical Retrieval, *Journal of Memory and Language*, 28, 462–485.
- McNeil MR, & Pratt SR (2001). Defining aphasia: Some theoretical and clinical implications of operating from a formal definition. *Aphasiology*, 15 (10–11), 901–911.
- Miller J, & Iglesias A (2016). *Systematic Analysis of Language Transcripts (SALT), Version 16* [Computer Software] Madison, WI: SALT software, LLC.
- Miller LM & Roodenrys S (2009). The interaction of word frequency and concreteness in immediate serial recall, *Memory & Cognition*, 37 (6), 850–865 [PubMed: 19679864]
- Minkina I, Rosenberg S, Kalinyak-Fliszar M, & Martin N (2017). Short-term memory and aphasia: From theory to treatment. *Seminars in Speech and Language*, 38 (1), 17–28. [PubMed: 28201834]
- Nettleton J, & Lesser R (1991). Therapy for naming difficulties in aphasia: Application of a cognitive neuropsychological model. *Journal of Neurolinguistics*, 6, 139–157.
- Nickels L (2002). Improving word finding: Practice makes (closer to) perfect? *Aphasiology*, 16, 1047–1106.
- Nicholas LE, & Brookshire RH (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 36 (2), 338–350.
- Novick JM, Bunting MF, Dougherty MR & Engle RW (Eds.) (2019.) *Cognitive and Working Memory Training: Perspectives from Psychology, Neuroscience and Human Development*. New York, NY: Oxford University Press.
- Psychology Software Tools, Inc. [E-Prime 2.0] (2012). Retrieved from <http://www.pstnet.com>.
- Regard M, Strauss E, & Knapp P (1982). Children's production on verbal and non-verbal Fluency tasks. *Perceptual and Motor Skills*, 55, 839–844. [PubMed: 7162920]
- Roach A, Schwartz MF, Martin N, Grewal RS & Brecher A (1996). The Philadelphia Naming Test: Scoring and Rationale. *Clinical Aphasiology*, 24, 121–133.
- Robey R (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders*, 37, 401–411. [PubMed: 15231420]
- Rosenbek JC, LaPointe LL, & Wertz RT (1989). *Aphasia: A clinical approach*. Pro Ed. Erlbaum Associates Inc.
- Saffran EM (1990). Short-term memory impairment and language processing. In Caramazza A (Ed.), *Advances in cognitive neuropsychology and neurolinguistics*. Hillsdale NJ: Lawrence Erlbaum.
- Saffran EM and Martin N (1990). Neuropsychological evidence for lexical involvement in STM. In Vallar G & Shallice T (Eds.), *Neuropsychological Impairments of Short-term Memory*. Cambridge: Cambridge University Press.
- Salis C (2011) Understanding of auditory discourse in older adults: The effects of syntax and working memory, *Aphasiology*, 25(4), 529–539.

- Salis C (2012). Short-term memory treatment: Patterns of learning and generalisation to sentence comprehension in a person with aphasia. *Neuropsychological Rehabilitation*, 22(3), 428–448. [PubMed: 22300398]
- Salis C, Kelly H & Code C (2015). Assessment and treatment of short-term and working memory impairments in stroke aphasia: A practical tutorial. *International Journal of Language and Communication Disorders*, 50(6), 721–736. [PubMed: 26123369]
- Schneider W, Eschman A & Zuccolotto (2002). *E-Prime User's Guide*. Psychology Software Tools, Inc.
- Sheskin DJ (2004) *Handbook of parametric and nonparametric statistical procedures*. 3rd ed., Boca Raton: Chapman & Hall /CRC.
- van Hees S Angwin A, McMahon K & Copland D (2013). A comparison of semantic feature analysis and phonological components analysis for the treatment of naming impairments in aphasia. *Neuropsychological Rehabilitation*, 23:1, 102–132, [PubMed: 23098246]
- Wambaugh JL, & Ferguson M (2007). Application of semantic feature analysis to retrieval of action names in aphasia. *Journal of Rehabilitation Research & Development*, 44 (3), 381–394. [PubMed: 18247235]
- Webster J, Whitworth A, & Morris J (2015). Is it time to stop “fishing”? A review of generalisation following aphasia intervention, *Aphasiology*, 29 (11), 1–25.
- Yeung O, Law S-P & Yau M (2009). Treatment generalization and executive control processes: Preliminary data from Chinese anomic individuals. *International Journal of Language and Communication Disorders*, 44 (5), 784–794. [PubMed: 19565397]
- Youse YM & Coelho CA (2005) Working memory and discourse production abilities following closed-head injury, *Brain Injury*, 19:12, 1001–1009. [PubMed: 16263642]

Baseline (BL)			12 Treatment (T) sessions				Post-Treatment probes			
BL1	BL2	BL3	T1	T2	T3	T4.....T12	24 hour	72 hour	One week	One month
3 baselines within a week's time, but more baselines if change among baselines is greater than 15%. 20 word strings for each baseline			No probe	Probe 1	Probe 3	Probe 4.....12				
			Cycle 1:	Cycle 1:	Cycle 1:	Treatment as in sessions preceding.				
			Set A (40 stimuli)	Set A (40 stimuli)	Set A (40 stimuli)					
			2 minute break	2 minute break	2 minute break					
			Set B (40 stimuli)	Set B (40 stimuli)	Set B (40 stimuli)					
			10 minute break	10 minute break	10 minute break					
			Cycle 2:	Cycle 2:	Cycle 2:					
			Set C (40 stimuli)	Set C (40 stimuli)	Set C (40 stimuli)					
			2 minute break	2 minute break	2 minute break					
			Set D (40 stimuli)	Set D (40 stimuli)	Set D (40 stimuli)					
			End of session	End of session	End of session					

Note 1. Baseline, treatment and probe stimuli were some combination of (1) concrete or abstract words, (2) 1-, 2- or 3-word sequence and (3) response delay of 1-sec, 5-sec or 10 sec and were different for each participant depending on his/her performance on each permutation of these three stimulus conditions.

Note 2. To the extent possible, unique stimuli were used in each baseline, training in each session, probes each session and each post-treatment probe. (See Treatment Stimuli Development in Methods section).

Figure 1.
 Schedule of Treatment Components

Example of treatment stimuli and administration: Abstract word pairs			
Participant ID / Condition:			
Examiner:			
Date:			
Cycle 1			
Set A		Set B	
quotient	transference	imprudence	distortion
relation	fundraising	leadership	vestige
genocide	spousal	repayment	candor
deficit	conundrum	mockery	inference
rendezvous	excellence	detection	reprimand
probation	morals	delinquent	suffrage
quackery	solace	jealousy	motto
zillionth	fostering	legalese	hankering
covenant	escrow	gratitude	engagement
disturbance	immersion	decency	boycott
Cycle 2			
Set C		Set D	
derangement	pinnacle	psychosis	treachery
salvation	mock	ascension	genre
dysfunction	workmanship	detriment	irony
courage	inclement	rascal	prevention
excuser	accord	subsistence	psyche
magnitude	contempt	enrichment	legion
achievement	decision	safety	infidel
mystery	cleansing	pretender	rejection
interlude	aversion	farewell	assessment
persistence	reckoning	rhetoric	wizardry

Figure 2.
Example of Treatment Session

Figure 3a

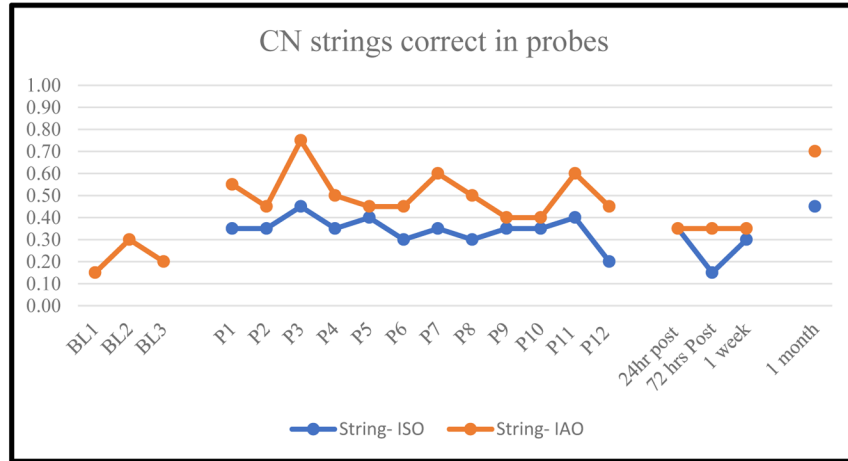


Figure 3b

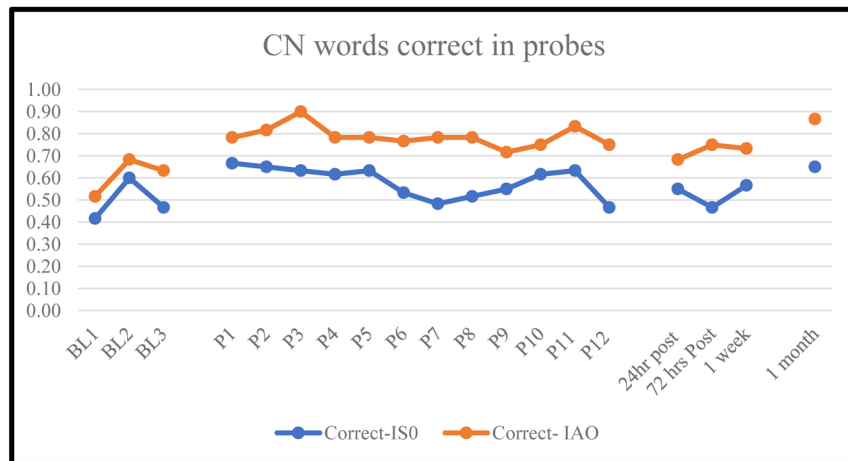


Figure 3. CN Proportion of Word Strings (3a) and Words (3b) Correct in Baseline, Treatment and Post-treatment Probes

Figure 4a

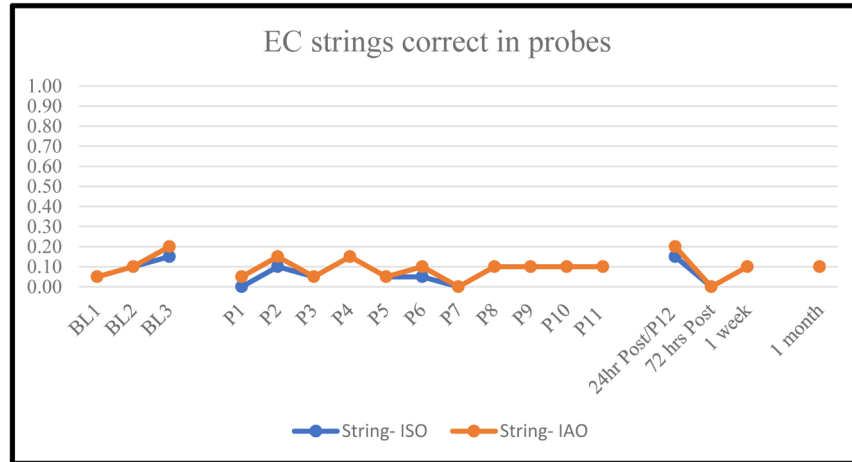


Figure 4b

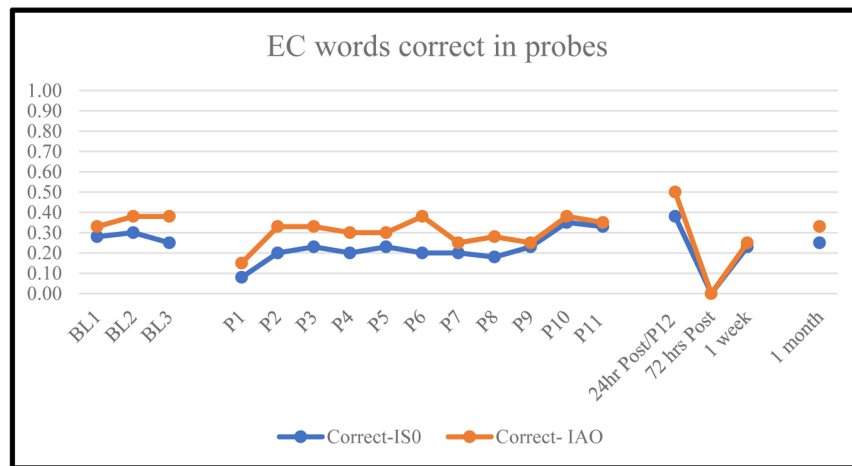


Figure 4. EC Proportion of Word Strings (4a) and Words (4b) Correct in Baseline, Treatment and Post-treatment Probes

Figure 5a

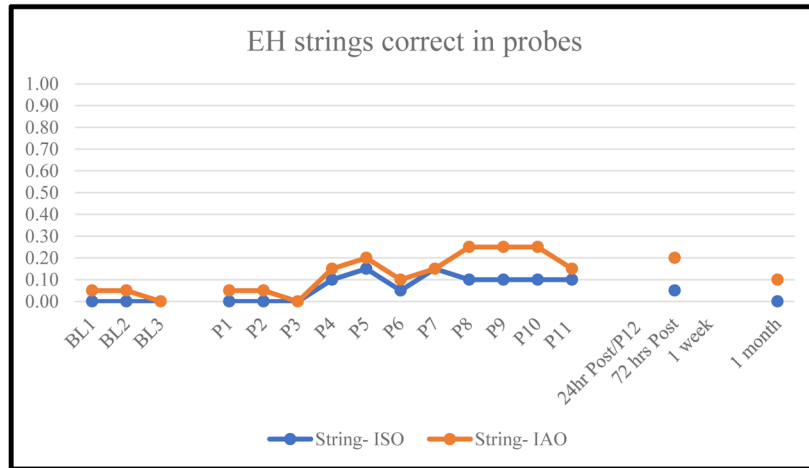


Figure 5b

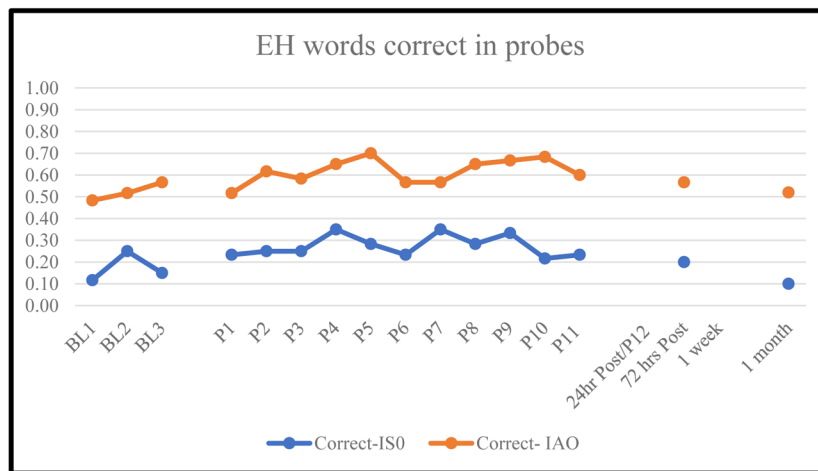


Figure 5. EH Proportion of Word Strings (5a) and Words (5b) Correct in Baseline, Treatment and Post-treatment Probes

Figure 6a.

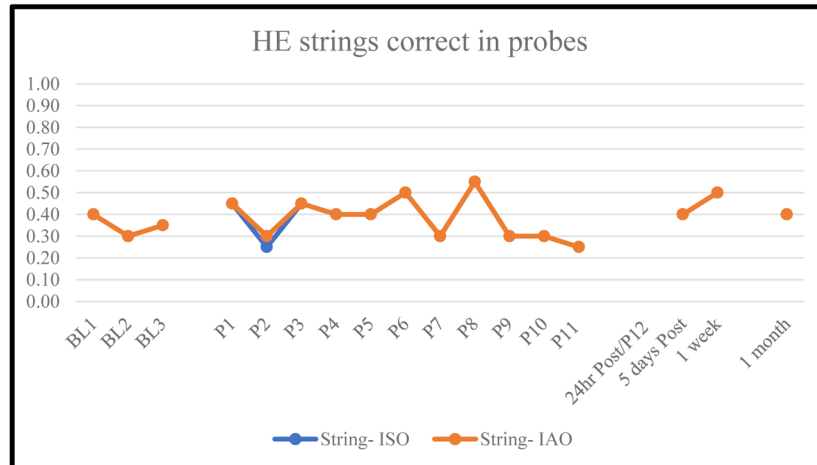


Figure 6b.

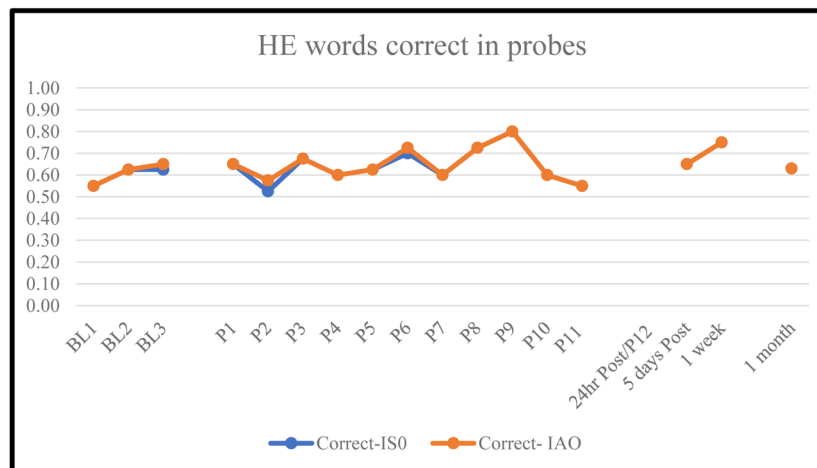


Figure 6. HE Proportion of Word Strings (6a) and Words (6b) Correct in Baseline, Treatment and Post-treatment Probes

Figure 7a

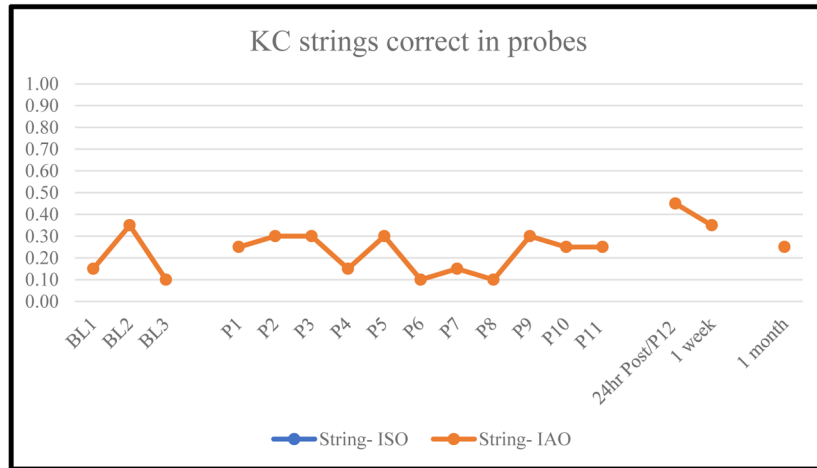


Figure 7b

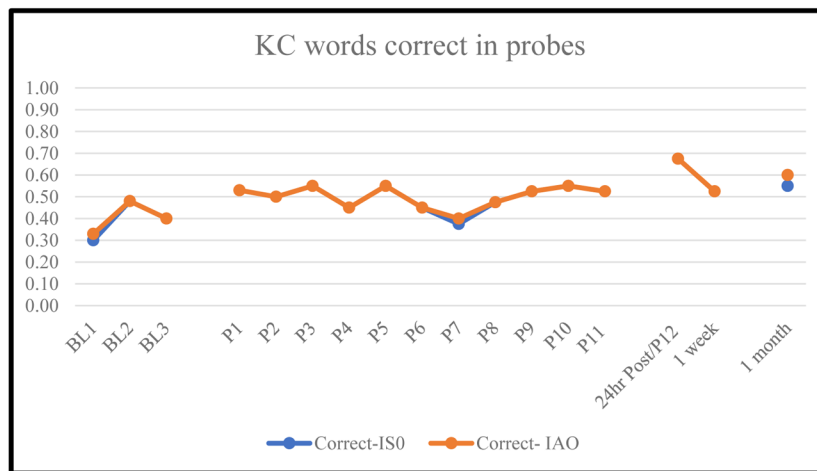


Figure 7. KC Proportion of Word Strings (7a) and Words (7b) Correct in Baseline, Treatment and Post-treatment Probes

Figure 8a

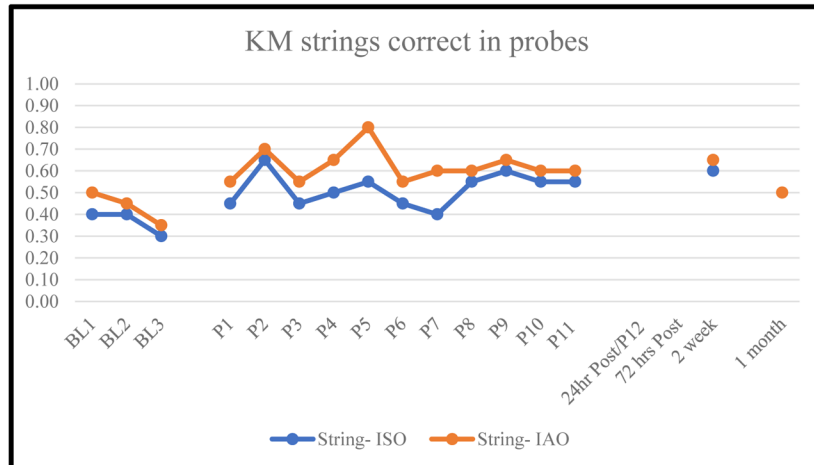


Figure 8b

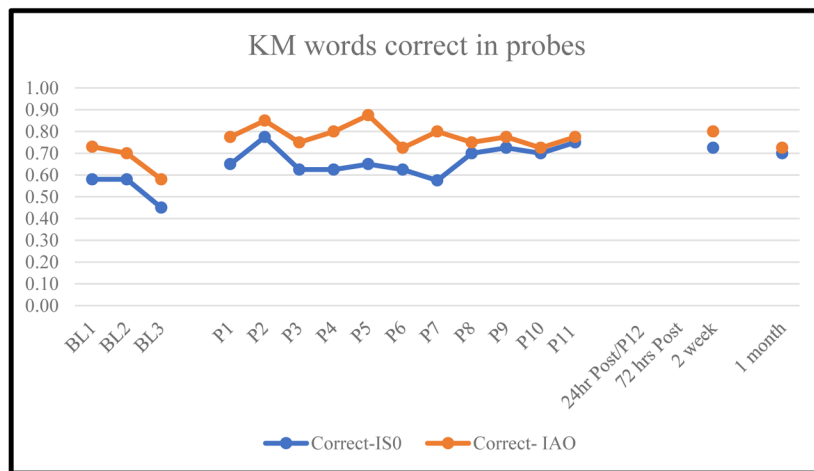


Figure 8. KM Proportion of Word Strings (9a) and Words (9b) Correct in Baseline, Treatment and Post-treatment Probes

Figure 9a

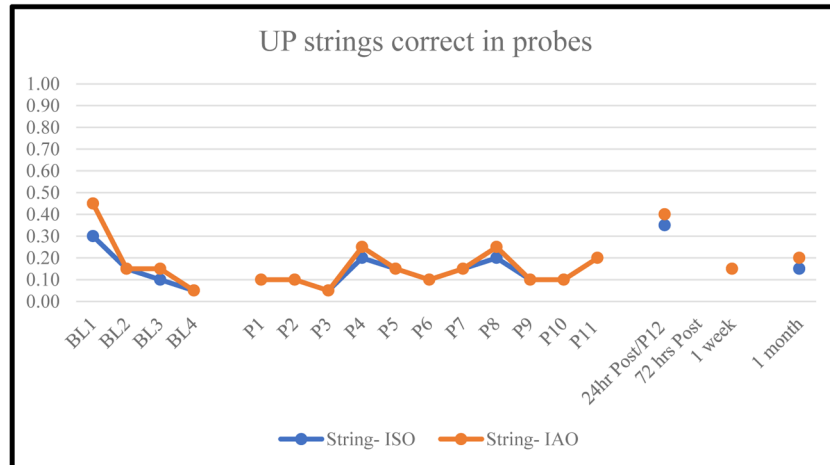


Figure 9b

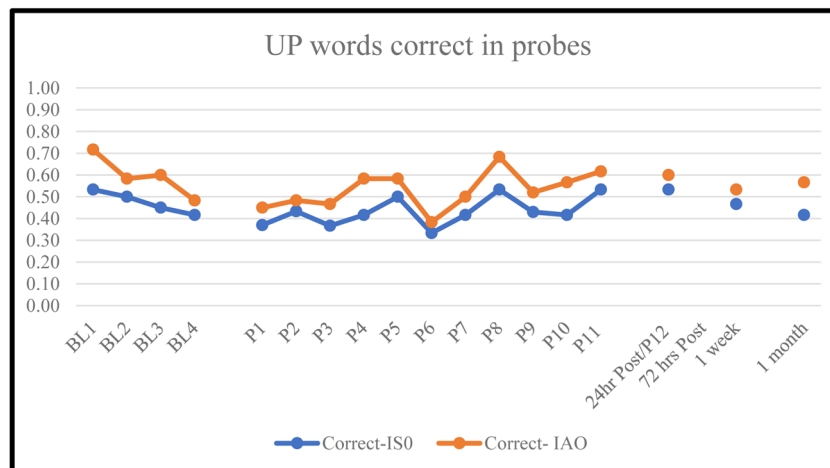


Figure 9. UP Proportion of Word Strings (9a) and Words (9b) Correct in Baseline, Treatment and Post-treatment Probes

Figure 10a

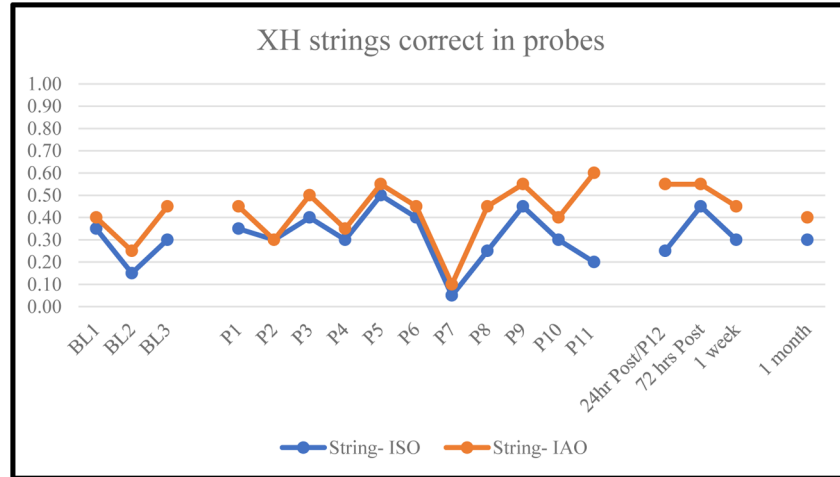


Figure 10b

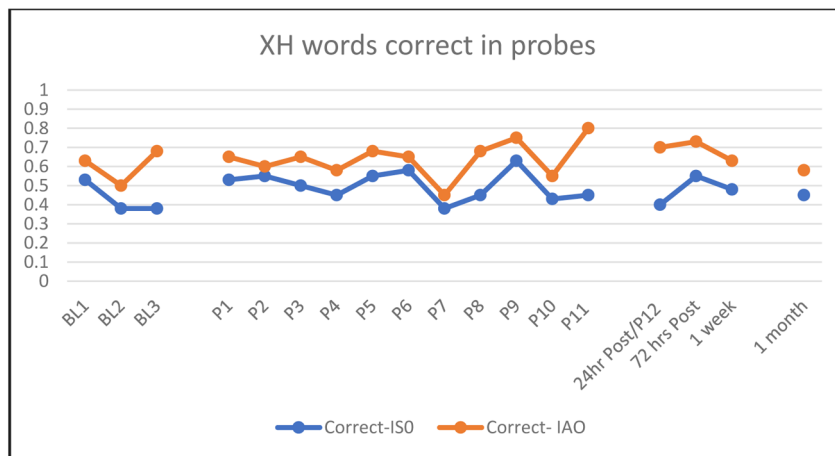


Figure 10. XH Proportion of Word Strings (10a) and Words (10b) Correct in Baseline, Treatment and Post-treatment Probes

Table 1.

Biographical information for the eight participants

ID	Sex	Age	MPO ¹	Education	Etiology	WAB-R ² Aphasia Quotient	WAB-R Aphasia Classification
CN	M	52	43	10 yrs	Left MCA ³ aneurysm and left internal carotid artery (ICA) occlusion, affecting left insular cortex, posterior 2/3s of the inferior frontal gyrus, inferior middle frontal gyrus, anterior margin of the angular gyrus and inferior insula. Temporal lobe is intact.	77.9	Transcortical Motor
EC	F	68	333	18 yrs	Left CVA ⁴ aneurysm involving the left frontal, parietal lobes with dilatation of the left frontal horn, left lateral ventricle, left atria and left temporal horn with enlargement of the left sylvian fissure.	80.3	Broca's
EH	F	52	135	13 yrs	Left CVA, left hemisphere stroke with damage to the insular cortex, middle and inferior frontal gyrus extending to the parietal lobe. Temporal lobe is intact.	84.8	Anomic
HE	F	63	91	16 yrs	Left CVA aneurysm in the region of the left MCA bifurcation/trifurcation, large hypoattenuating lesions within the left temporal, frontal and parietal lobes into the frontoparietal lobes.	80.3	Transcortical Motor
KC	M	56	161	15 yrs	Left CVA involving the middle and superior temporal gyrus, middle and inferior frontal gyrus, inferior parietal lobe (supramarginal and angular gyrus), and extending down to the lateral ventricle with damage to the basal ganglia. Temporal pole is preserved.	69.1	Transcortical Sensory
KM	M	71	173	16 yrs	Left CVA infarct affecting MCA territory and some of the ACA ⁵ territory; extensive damage to frontal portions of the temporal and parietal lobes, down to lateral ventricles, sparing superior middle and frontal lobes. Insula and basal ganglia are severely damaged.	77.6	Broca's
UP	M	53	98	14 yrs	Left CVA affecting the posterior 2/3s of the inferior frontal gyrus, subcortical white matter beneath the middle and superior frontal gyri, and the anterior superior insula cortex. The temporal lobe was intact.	88.9	Anomic
XH	M	48	19	7 yrs	Left MCA infarct with damage to the middle and inferior frontal gyrus and anterior insula. Some damage extending to the basal ganglia (head of the caudate).	77	Conduction

¹ months post-onset,²WAB-R = Western Aphasia Battery-Revised (Kertesz, 2006),³MCA = middle cerebral artery,⁴CVA = cerebral vascular accident,⁵ACA = anterior cerebral artery

Table 2.

Language and verbal short-term memory (STM) abilities of participants with aphasia and five neurotypical control participants: Proportion correct on TALSA assessments of lexical-phonological and lexical-semantic abilities as well as repetition span for digits and words

Participant	Lexical-Phonological				Lexical-Semantic		Verbal Short-term Memory	
	Phoneme Discrimination		Rhyming Judgment Pairs		Lexical Comprehension	Category Judgment	Repetition Span (in serial order)	
	Word (n=44)	Nonword (n=44)	Word (n=60)	Nonword (n=60)	Word-Picture Matching (n=48)	Word Pairs (n=60)	Digits	Words
CN	0.95	1.00	0.90	0.97	0.96	0.95	3.40	2.80
EC	0.89	0.98	0.93	0.88	0.92	0.90	3.00	3.20
EH	0.98	0.98	0.93	0.88	1.00	0.87	2.40	2.80
HE	1.00	0.86	0.95	0.83	0.98	0.97	4.80	3.40
KC	0.73	0.56	0.82	0.82	0.88	0.87	2.80	3.40
KM	0.93	0.98	0.87	0.87	0.98	0.98	3.40	3.20
UP	1.00	1.00	0.90	0.98	0.96	1.00	3.40	3.40
XH	0.98	0.98	0.90	0.89	1.00	0.97	2.60	1.80
Controls (n=5)								
Mean	0.99	0.98	0.98	0.99	1.00	1.00	7.00	6.20
Standard Deviation	0.013	0.034	0.03	0.02	0.01	0.01	0.00	1.03

Table 3.

Rates of repeated use of words that appeared more than once in pairs and in triplets

Stimulus concreteness and syllable length	Word Pairs	Word Triplets
Concrete 2-syllable words	1 time	2 times
Concrete 3-syllable words	4 times	5 times
Abstract 2-syllable words	1 time	2 times
Abstract 3-syllable words	3 times	2 times

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

Assignment of stimulus type (concrete/abstract), stimulus length (pair, triplet), interval time (5 or 10 seconds) and intensity of treatment (two or four times per week)

ID	Stimulus	Condition	Time Interval	Intensity
CN	Concrete	Triplets	5 seconds	4x/wk
EC	Abstract	Pairs	5 seconds	2x/wk
EH	Concrete	Triplets	5 seconds	4x/wk
HE	Abstract	Pairs	10 seconds	2x/wk
KC	Abstract	Pairs	5 seconds	2x/wk
KM	Concrete	Pairs	10 seconds	2x/wk
UP	Abstract	Triplets	5 seconds	4x/wk
XH	Abstract	Pairs	5 seconds	4x/wk

Table 5.

Effect sizes for change between baseline and (1) the immediate post-treatment probe and (2) the 1-month maintenance probe for strings and words correct in serial order (ISO) and in any order (IAO)

ID	Immediate Post-Treatment Probe				1 Month Maintenance Probe			
	String		Word		String		Word	
	ISO	IAO	ISO	IAO	ISO	IAO	ISO	IAO
CN	0.65	1.75	0.34	1.33	3.06 [*]	6.33 ^{***}	1.65	3.05 [*]
EC	0.50	0.44	1.13	0.40	0.00	-0.22	-1.06	-1.15
EH	1.73	5.77 ^{**}	0.66	1.06	0.00	0.96	-1.72	-0.03
HE	2.00	2.00	3.18 [*]	2.18	1.00	1.00	0.75	0.35
KC	1.51	1.51	2.29	2.62 [*]	0.38	0.38	1.73	2.62 [*]
KM	4.04 [*]	2.84 [*]	2.51	1.64	2.31	0.87	2.18	0.69
UP	0.93	0.43	0.48	-0.31	0.00	0.00	-1.13	-0.31
XH	0.64	1.44	0.54	0.90	0.32	0.32	0.23	-0.25

Note: Effect sizes calculated with the following formulas: Mean post-treatment-Mean baseline/SDbaseline (Beeson & Robey, 2006) 1 Month Accuracy-Mean baseline/SD baseline

* = small effect (> 0.2);

** = medium effect (> 0.5);

*** = large effect (> 0.8)

Table 6.

Participants whose pre-treatment performance on repetition of concrete or abstract singlets, pairs and triplets at 1-sec, 5-sec and 10-sec response delays changed significantly or at trend level after treatment

Participant ID	Stimulus type	Sequence Length	Response Delay	Proportion Correct			Confidence Interval	z value	p value
				Before Treatment	After Treatment	Log odds ratio			
Significant improvement ($p < .05$)									
CN	Abstract	Pairs	10 sec	0.10	0.50	0.13	0.02 to 0.82	2.18	0.03
KM	Concrete	Pairs	10 sec	0.10	0.50	0.13	0.02 to 0.82	2.18	0.03
UP	Concrete	Triplets	1 sec	0.10	0.60	0.10	0.02 to 0.63	2.46	0.01
UP	Abstract	Pairs	10 sec	0.10	0.50	0.13	0.02 to 0.82	2.18	0.03
XH	Abstract	Singlets	5 sec	0.40	0.80	0.17	0.05 to 0.85	2.15	0.03
XH	Concrete	Singlets	5 sec	0.60	1.00	0.05	0.00 to 0.94	2.00	0.05
Trends ($p < .07$)									
CN	Concrete	Pairs	10 sec	0.40	0.70	0.24	0.05 to 1.13	1.80	0.07
EC	Concrete	Pairs	1 sec	0.20	0.50	0.22	0.04 to 1.11	1.84	0.07
HE	Abstract	Singlets	1 sec	0.87	0.53	5.69	0.94 to 34.46	1.891	0.06

No significant changes or trends: KC, EH

Table 7.

Repetition spans for concrete and abstract words, recalled in serial order, pre- and post-treatment

ID	Concrete Words			Abstract Words		
	Pre-treatment	Post-treatment	Proportion change	Pre-treatment	Post-treatment	Proportion change
CN	3.00	3.20	0.20	1.80	1.80	0.00
EC	1.60	2.20	0.60	1.80	1.80	0.00
EH	2.60	3.00	0.40	2.20	2.60	0.40
HE	3.00	3.00	0.00	2.60	3.00	0.40
KC	2.40	2.80	0.40	2.40	1.60	-0.80
KM	2.80	3.00	0.20	1.80	2.40	0.60
UP	3.20	3.40	0.20	2.40	2.60	0.20
XH	1.80	2.40	0.60	1.80	2.80	1.00
Mean	2.55	2.88	0.33	2.10	2.33	0.23
SD	0.58	0.40	0.21	0.34	0.52	0.53
Median	2.70	3.00	0.30	2.00	2.50	0.30

Table 8.

Proportion correct on pre- and post-treatment picture naming tests (4 participants, BNT¹ and 1 participant PNT²)

Participant ID ³	Naming test administered	Number of test items	Proportion Correct		Change in accuracy (+/-)	z test comparing proportion correct pre- and post-treatment
			Pre-treatment	Post-treatment		
HE	BNT	60	0.68	0.68	0.00	NS ⁴
KC	BNT	60	0.57	0.63	0.07	NS
KM	BNT	60	0.65	0.60	-0.05	NS
UP	BNT	60	0.68	0.72	0.03	NS
XH	PNT	175	0.55	0.66	0.11	$z = 2.18, p = .03$

¹Boston Naming Test (Kaplan, Goodglass & Weintraub, 1983).

²Philadelphia Naming Test (Roach, Schwartz, Martin, Grewal & Brecher, 1996).

³Three participants excluded from this analysis because dates of testing were too distant from initiation of treatment.

⁴NS = Not significant

Table 9.

N & B discourse results from pre- to post-treatment

Participant	Time Period	Total words	Total CIUs	%CIUs	Total Time (sec)	Words/Min	CIUs/Min	Total mazes	%Mazes
HE	Pre-tx	531	259	48.78%	1283	24.84	12	184	34.65%
	Post-tx	623	320	51.36%	1567	23.85	13.42	107	<i>17.18%</i>
KC [*]	Pre-tx	327	57	17.43%	365	53.78	9.38	20	6.12%
	Post-tx	325	66	20.31%	262	74.42	15.11	14	4.20%
KM	Pre-tx	457	205	44.44%	1132	24.22	10.87	91	19.91%
	Post-tx	384	211	54.95%	1021	22.57	12.4	83	21.61%
UP	Pre-tx	777	344	44.27%	513	90.88	40.23	38	4.89%
	Post-tx	837	460	54.96%	656	76.58	42.09	53	6.33%
XH	Pre-tx	844	214	25.36%	821	61.7	15.64	182	21.56%
	Post-tx	722	230	31.86%	706	61.34	19.54	151	20.91%

Notes. (1) CIUs = correct information units, %CIUs = percent correct information units, %Mazes = percent of mazes out of total words.

(2) Bold and italicized result for HE indicates substantial decrease in mazes compared to total words.

(3) Post-treatment results in bold typeface indicate an increase of greater than twice the standard error of the mean reported in Brookshire and Nicholas (1994) for percent CIUs (2.1) and/or words per minute (5.5).

* (4) Data collection error, only one discourse sample was collected.

(5) Three participants were excluded from this analysis because dates of testing were too distant from initiation of treatment.

Table 10.

Summary of changes made on pre- and post-treatment assessments

Participant	Repetition of Concrete/Abstract Stimuli		Concrete and Abstract Word Spans		Naming	Discourse Measures		
	Significant improvement on one or more stimuli conditions after treatment	Yes	Concrete	Abstract		Change in proportion correct	Increased % CIU	Increased words/minute
CN	Yes	0.20	0.00	0.00		--		
EC	Trend	0.60	0.00	0.00		--		
EH	No	0.40	0.40	0.40		--		
HE	Trend	0.00	0.40	0.40	0			-17.18%
KC	No	0.20	0.60	0.07			20.64%	
KM	Yes	0.40	-0.80	-0.05			10.51%	
UP	Yes	0.20	0.20	0.03			10.69%	
XH	Yes	0.60	1.00	0.11*			6.50%	

Notes. (1) Bolded text indicates notable or significant improvements following treatment.

* (2) p = .03