The Link between Verbal Short-Term Memory and Anomia Treatment Gains

Reva M. Zimmerman, MS
University of Washington, Speech and Hearing Sciences,
Seattle, WA

JoAnn P. Silkes, PhD
University of Washington, Speech and Hearing Sciences,
Seattle, WA

Diane L. Kendall, PhD
VA RR&D Puget Sound DVA Medical Center, Research Service
University of Washington, Speech and Hearing Sciences, Seattle, WA
University of Pretoria, Pretoria, South Africa

Irene Minkina, PhD
Temple University, Communication Sciences and Disorders,
Philadelphia, PA

Corresponding Author:
Reva M. Zimmerman
Department of Speech and Hearing Sciences
University of Washington
1417 N.E. 42nd St. Box 354875
Seattle, WA 98105-6246
Voice: (206) 685-2140
Fax: (206) 616-3566
E-mail: robinrm@uw.edu

Conflict of interest

The authors have no relevant conflicts of interest to declare.

Funding

Diane Kendall was funded by VA RR&D Merit Review Grant #C6572R.

Abstract

Purpose: A significant relationship between verbal short-term memory (STM) and language performance in people with aphasia (PWA) has been found across studies. However, very few studies have examined the predictive value of verbal STM in treatment outcomes. This study aims to determine if verbal STM can be used as a predictor of treatment success.

Methods: Retrospective data from 25 PWA in a larger randomized control trial of Phonomotor Treatment (PMT) were analyzed. Digit and word spans from immediately pre-treatment were run in multiple linear regression models to determine whether they predict magnitude of change from pre- to post- treatment and follow-up naming accuracy. Pre-treatment, immediately post-treatment, and three months post-treatment digit and word span scores were compared to determine if they changed following a novel treatment approach.

Results: Verbal STM, as measured by digit and word spans, did not predict magnitude of change in naming accuracy from pre- to post-treatment nor from pre-treatment to 3 months post-treatment. Furthermore, digit and word spans did not change from pre- to post-treatment or pre-treatment to 3-months post-treatment in the overall analysis. A post hoc analysis revealed that only the less impaired group showed significant changes in word span scores from pre-treatment to 3-months post-treatment.

Discussion: The results suggest that digit and word spans do not predict treatment gains. In a less severe subsample of participants, digit and word span scores can change following PMT; however, the overall results suggest that span scores may not change significantly. The implications of these findings are discussed within the broader purview of theoretical and empirical associations between aphasic language and verbal STM processing.

Aphasia is a language disorder affecting up to 2 million people in the United States (National Aphasia Association, n. d.). The disorder most commonly results from stroke, although it can occur after any type of injury to the language processing centers of the brain. The personal and social costs of aphasia are high; thus, effective treatments are needed to return people with aphasia (PWA) as close to their premorbid communicative abilities as possible. Most people who pursue treatment experience some degree of language improvement; unfortunately, there are many people who do not improve. Variable treatment response is common in aphasia treatment, regardless of the treatment type, dosage, or distribution (Nickels, 2002). Consequently, reliable predictors of aphasia treatment response have been the subject of research for decades and no clear predictors have emerged. That said, studies that included baseline cognitive measures – such as short-term memory (STM) – have shown promise as potential response predictors (Dignam et al., 2017; Goldenberg, Dettmers, Grothe, & Spatt, 1994; Lambon Ralph, Snell, Fillingham, Conroy, & Sage, 2010; Van De Sandt-Koenderman et al., 2008).

Defining Short-Term Memory

Short-term memory (STM) is loosely defined as the temporary activation of information, such that it is easily accessible for processing (Cowan, 2008). This "information" may be tactile, visuospatial, or verbal/linguistic and, although the contents of STM may be within conscious awareness, this is not necessarily the case; processing may occur subconsciously (Cowan, 1988). Cowan's model is compatible with a more language-specific view of STM, such as that proposed by Martin and Saffran (1997), wherein verbal STM *is* the activation being transmitted through an interactive activation network (Dell, 1986; Dell & O'Seaghdha, 1992). This view accounts for many linguistic behaviors observed in neurotypical and aphasic adults, without the need to describe a memory mechanism that is external to the language network (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). Furthermore, a model in which memory is integral to information processing is neurologically plausible, as described in the parallel distributed processing model of phonology (PDP, Nadeau, 2001). In this particular PDP model of language, processing cannot occur independently from memory function, as memory is the mechanism for encoding experiences and thereby strengthening neuronal and network connections. In summary, three complementary models coalesce to describe STM broadly as fleeting activation and, in the case of verbal STM, as activation that is integral to linguistic processing.

The concept of STM must be distinguished from working memory (WM), a related but distinct form of memory. WM has been conceptualized as STM plus attention (Cowan, 1988; Engle, Tuholski, Laughlin, & Conway, 1999), meaning that STM maintains information in an activated state, while attentional processes are applied to this information for the purpose of manipulating it (e.g., multiplying numbers). Furthermore, STM processes have been described as domain-specific, while WM is domain-free. That is, the temporary activation of information is inherently tied to the type of information being activated (e.g., auditory-verbal, visuospatial, or tactile; Cowan, 1988; Nadeau, 2001), while the attentional component of WM applies regardless of the stimulus type (Cowan, 1988; Engle et al., 1999). Given the distinction between these constructs, it is no surprise that STM and WM have been demonstrated to contribute differentially to measures of general intelligence and verbal abilities (Engle et al., 1999). Owing to their theoretical and empirical differences, STM and WM are often explored separately in the aphasia literature.

Short-Term Memory in Aphasia

Numerous studies have found an impairment of short-term memory in PWA using classical measures such as digit and word spans. In one study, Kasselimis et al. (2013) tested 64 subjects with aphasia and 15 subjects with left-hemisphere brain damage without aphasia on digit and spatial span tasks. They found that the PWA performed significantly worse than non-aphasic participants on all memory tasks. These memory impairments have been found to relate to performance on many types of linguistic tasks, including single word processing and sentence comprehension. At the single word level, Martin and Ayala (2004) assessed phonological, semantic, and verbal STM abilities in 64 PWA and found that the linguistic variables, especially phonological, were significantly correlated with at least one type of verbal STM measure. Likewise, for sentence processing, Pettigrew and Hillis (2014) found that the verbal STM (but not WM) scores of 47 PWA accounted for a significant proportion of the variance in their sentence comprehension scores. While these types of findings are common in the literature, it is also apparent that a high degree of individual difference exists for both the linguistic and verbal STM profiles of PWA (Majerus, Attout, Artielle, & Van der Kaa, 2015). The variety of STM profiles in aphasia has led some researchers to speculate about its predictive powers, as well as its role in language treatment. Namely, is it possible that the individual differences seen in STM abilities at baseline are related to the differences in outcomes following language treatment?

Over the past few decades, several researchers have looked to STM as a means to either predict or induce language change in PWA. Of the studies that have evaluated short-term memory as a predictor, most have found that a relationship exists between measures of visuospatial STM and language treatment gains (Goldenberg et al., 1994; Lambon Ralph et al., 2010; Seniow, Litwin, & Lesniak, 2009). Given this finding, verbal STM is the next logical domain to evaluate, as it could be a better indicator of treatment success than visuospatial memory: If temporary activation in the language network is a form of verbal STM, then measuring the integrity of that activation may provide a great deal of insight into how PWA perform in language treatment. To date, only one study has specifically included measures of verbal STM for the purposes of predicting language outcomes (Dignam et al., 2017). Dignam and her colleagues found that verbal STM and verbal learning, as measured by subtests of the Hopkins Verbal Learning Test-Revised (HVLT-R; Brandt & Benedict, 2001), were the only significant predictors of magnitude of change for treated and untreated items in PWA. Verbal STM also significantly predicted the degree of maintenance observed one month following treatment termination.

The extremely limited research on the predictive value of verbal STM in aphasia treatment is surprising for two key reasons. First, a wealth of evidence exists to demonstrate that verbal STM impairments abound in aphasia, significantly more so than in neurologically healthy controls and even people with brain injury without aphasia (Kasselimis et al., 2013; Martin & Saffran, 1997; Minkina, Rosenberg, Kalinyak-Fliszar, & Martin, 2017; Pettigrew & Hillis, 2014; Potagas, Kasselimis, & Evdokimidis, 2011; Ween, Verfaellie, & Alexander, 1996). Furthermore, some studies have demonstrated that verbal STM profiles are much more heterogeneous in PWA than in healthy controls (Majerus et al., 2015). Taken together, these findings raise the possibility that certain types of verbal STM impairment could be considered a *feature* of aphasic language performance, and that the differences in verbal memory processing in PWA could somehow be related to the variability of treatment outcomes. Second, given the linguistic demands of treatment and the prevalence of verbal STM impairment in aphasia, the characterization of participants' verbal STM abilities is theoretically and clinically important. Theoretically speaking, such descriptive information could inform current thinking about the intersection between language treatment and STM. From a clinical perspective, if PWA have difficulty maintaining activation of linguistic information during treatment, they may benefit more from treatments with carefully considered memory demands (Harris, Olson, & Humphreys, 2014; Martin & Gupta, 2004).

This Study

Given the small body of research, coupled with the theoretical and clinical importance of verbal STM in PWA, the aim of this project was to conduct a retrospective analysis of a large-scale treatment study to explore the relationship between verbal STM and aphasic language recovery in 24 PWA following an intensive anomia treatment, Phonomotor Treatment (PMT; Kendall et al., 2008). Two separate questions were under investigation here. In the first aim, verbal STM was evaluated as an indicator of anomia treatment success by assessing whether pre-treatment verbal STM abilities predicted improvement on naming of treated and untreated nouns immediately and three months after the treatment program. The measures of verbal STM used in this study were word and digit span scores at baseline, which were hypothesized to significantly predict the magnitude of treatment gains. If this finding were borne out, it would suggest that verbal short-term memory abilities at the outset of treatment are related to acquisition and generalization of linguistic gains during and following PMT.

Exploring the influence of treatment on verbal STM abilities would shed more light on the interplay between language and memory processing. Thus, the second aim of this study was to determine whether PMT influenced verbal STM immediately and three months post-treatment. The treatment consists of two general phases, wherein the participant begins by training on sounds in isolation and then progresses to sounds in sequences. After individual sounds have been introduced in isolation, participants are asked to perform phonological awareness and manipulation tasks (e.g., parsing, blending) with real and nonword stimuli in increasingly longer phoneme sequences. In essence, PMT requires repeated linguistic activation, maintenance of activation, and manipulation of phonemes and phoneme strings. The models outlined above suggest that verbal STM may be a component of the mechanism through which linguistic representations become activated. Thus, intensive stimulation of linguistic representations during PMT should result in stronger verbal STM abilities. These improvements should be reflected in increased scores on measures of verbal STM and, as such, digit and word span scores were predicted to significantly improve following PMT. To evaluate any change in span scores following PMT, measures of verbal STM from immediately post-treatment and maintenance were each compared to baseline verbal STM scores. In this study, intensive PMT was predicted to significantly improve verbal STM abilities. Such a finding would suggest that PMT may enhance the processes that underlie word-finding (Martin & Saffran, 1997).

Method

Design

This was a retrospective analysis of data from a randomized control trial (RCT) funded by the Veterans Administration (VA RR&D C6572R) comparing two treatment types for remediation of lexical retrieval abilities in aphasia (Kendall, Oelke, Allen, Torrence, & Nadeau, 2018). In that study, individuals were randomly assigned to either a control treatment (Semantic Feature Analysis, or SFA; Boyle, 2001, 2010; Coelho, McHugh, & Boyle, 2000) or experimental (Phonomotor) treatment. Due to significant overlap (44%) between the word span stimuli and SFA trained items, only the PMT group was considered in this analysis. Participants were tested at three time points: prior to treatment, immediately post-treatment, and three months following treatment termination. Participants were prohibited from engaging in any other therapeutic activities, including informal exercises at home or therapy groups, throughout the course of the study to ensure that changes in naming accuracy were attributable to the experimental treatment.

Participants

Participants were included in the RCT if they had acquired aphasia due to a left-hemisphere lesion sustained at least 6 months before enrolling in the study. Presence of aphasia was demonstrated by t-scores below cut-off on at least one of the following sections of the Comprehensive Aphasia Test (CAT; Howard, Swinburn, & Porter, 2010; Kaplan, 2001): auditory comprehension total, reading comprehension total, and writing total. Presence of anomia was determined by performance on the Boston Naming Test (Kaplan, Goodglass, Segal, & Weintraub, 2001). All participants were evaluated for apraxia of speech, the presence and severity of which were determined through analysis of spontaneous and elicited speech samples for rate, segment and intersegment durations, phonemic distortions, abnormal prosody, and effortful groping (Duffy, 2013). Participants were excluded from the study if they demonstrated severe apraxia of speech, as determined by clinical judgments on the criteria above, including a limited speech phonemic repertoire. Participants were also excluded if their medical histories revealed any neurological illness, untreated psychological illness, or chronic medical illnesses that might prevent faithful administration of the treatment protocol. Participants also had to demonstrate normal or corrected-to-normal vision and hearing to be included.

A total of 30 participants were randomly assigned to the PMT condition. Two PMT participants did not complete the protocol: one was withdrawn by the clinician after it was discovered that she suffered a traumatic brain injury just prior to the onset of treatment and the other withdrew himself due to family obligations. Any participants with missing data for a given analysis were eliminated from the analysis on a case-by-case basis. Thus, an additional 3 participants were removed for completely missing verbal STM scores. The remaining 25 participants in this analysis had a mean age of 63.7 (*SD* = 10.4), were 3.6 years post-onset (*SD* =

Table 1. Participant Characteristics

Participant	Age (years)	Sex	Education level (years)	Duration post onset (years)	Handedness	Raven's *Executive Function* (out of 36)	BNT *Lexical Retrieval* (out of 60)	CAT *Semantic Memory* t-score	CAT *Auditory Comprehension* t-score (cut-off ≤ 56)	CAT *Reading Comprehension* t-score (cut-off ≤ 60)	CAT *Writing* t-score (cut-off ≤ 57)	SAPA *Phonologic Processing* (out of 144)
1	71	F	15	1.42	R	35	49	62	59	66	69	116
2	46	F	16	1.25	R	28	23	54	58	62	69	79
3	59	F	16	5.25	R	33	42	62	50	54	56	110
4	67	M	12	2.75	R	29	46	62	60	59	59	94
5	70	F	12	6.75	R	23	6	38	39	48	44	61
6	40	F	16	1.83	R	36	5	62	51	48	52	73
7	59	M	18	2.25	R	32	6	54	54	50	55	58
8	71	F	14	0.83	R	31	36	62	58	60	52	111
9	65	M	16	8.42	R	34	4	62	44	49	53	62
10	73	M	16	2.33	R	28	6	54	49	44	43	49
11	73	F	13	2.5	R	34	12	50	50	54	53	100
12	67	M	16	4.17	R	32	20	54	53	56	57	84
13	46	M	13	7.08	L	*	14	50	45	44	41	48
14	71	F	16	3.67	R	34	41	54	52	66	62	109
15	90	F	12	6.42	L	29	26	54	58	59	60	79
16	63	M	13	4	R	35	32	62	45	56	50	77
17	60	M	14	2.75	L	19	2	62	46	50	46	50
18	62	M	16	0.83	R	35	46	54	58	55	59	105
19	50	M	12	2.08	R	29	1	62	52	51	48	79
20	67	F	16	1.67	R	32	39	62	45	54	51	93
21	70	M	12	4.42	R	21	32	54	56	49	50	55
22	66	F	12	9.42	R	22	40	62	58	54	54	94
23	65	M	14	1.42	L	22	52	62	55	54	49	88
24	59	M	18	4.17	L/R	33	16	50	47	54	48	60
25	63	M	16	2	L	32	2	62	50	42	45	35
AVE	63.7		14.6	3.6		29.9	23.9	57.0	51.7	53.5	53.0	78.8
SD	10.4		2.0	2.4		5.0	17.4	6.1	5.7	6.3	7.4	23.1

Note. M = male; F = female; BNT = Boston Naming Test; CAT Memory = Comprehensive Aphasia Test Memory Composite Score (Semantic Memory + Recognition Memory); SAPA = Standardized Assessment of Phonology in Aphasia; TALSA = Temple Assessment of Language and Short Term Memory in Aphasia; AVE = average; * = Missing score.

2.4), and 44% were female (*n* = 11). Semantic abilities were quantified using the semantic memory subtests of the CAT (Howard, Swinburn, & Porter, 2010; Kaplan, 2001, mean = 29.3, *SD* = 17.4), nonverbal reasoning skills were measured with Raven's Progressive Matrices (Raven, Court, & Raven, 1998; mean = 29.9, *SD* = 5.0), and phonologic abilities were measured by the Standardized Assessment of Phonology in Aphasia (Kendall et al., 2010; mean = 78.8, *SD* = 23.1). See Table 1 for each participant's demographic profile and baseline scores on standardized assessments.

Stimuli and Treatment

Phonomotor treatment was delivered on an intensive schedule: 2 hours/day, 4-5 days/week, for six consecutive weeks, for a total of 56-60 hours of therapy. Treatment stimuli consisted of single phonemes, as well as 1-, 2- and 3-syllable real and non-words with phoneme strings of low phonotactic probability (PP) and high neighborhood density (ND; see Vitevich & Luce, 1998, for criteria). The trained real words consisted of 40 words selected using the MRC Psycholinguistic Database (Coltheart, 1981) to control for variables such as age of acquisition, imageability, and frequency. Furthermore, the trained real words were not the same as the word stimuli in the word span task (described below). Trained items were those directly targeted in therapy, while untrained items were only presented during the pre-treatment, immediate post-treatment, and follow-up assessment periods.

Outcome Measures

Verbal short-term memory

The Temple Assessment of Language and Short-Term Memory in Aphasia (TALSA; Martin, Minkina, Kohen, & Kalinyak-Fliszar, 2018) was administered prior to treatment, immediately post-treatment and at maintenance testing. The TALSA has been used in an increasing number of studies (Martin, Kohen, Kalinyak-Fliszar, Soveri, & Laine, 2012; McNeil et al., 2018; Peñaloza et al., 2017; Tuomiranta, Grönroos, Martin, & Laine, 2014) to test verbal STM in PWA by using carefully designed stimuli that balance linguistic complexity and memory load. This assessment tool consists of a variety of subtests to measure different areas of verbal STM (e.g., semantic versus phonologic). Six subtests from the TALSA were administered to participants: rhyming judgments, category judgments, rhyming triplet judgments, synonymy triplet judgments, and digit and word spans. Rhyming judgments were administered as a measure of phonological STM, and category judgments were selected as a measure of semantic STM. These tasks required participants to maintain activation of a single word for either one second or five seconds before being presented with a second word; participants made a judgment (e.g., the words rhyme, or they do not rhyme) with a single button press after both words were presented. On the other hand, rhyming triplet and synonymy triplet judgments were administered as measures of phonological and lexical-semantic working memory, respectively (Martin et al., 2018). In the triplet tasks, the participants heard and saw three words and pointed to the two rhyming words (in the rhyming triplet task) or to the two synonyms (in the synonymy task). The digit and word spans were administered as more traditional measures of verbal STM in both normal and disordered populations (Jones and Macken, 2015). Participants heard strings of numbers and words and then pointed to numbers or pictures on a screen in the exact order that they heard them. On the rhyming judgments, category judgments, rhyming triplet judgments, synonymy triplet judgments subtests, participant performance was generally at ceiling; thus, only the digit and word span measures were included in this analysis.

The digit span stimuli and administration are generally consistent with those used in normal populations (Jones & Macken, 2015; Martin et al., 2018). The digits 1-9 were used and were not repeated within a sequence. The word span stimuli are specifically what differentiate the TALSA span tasks from other word spans. The stimuli were carefully chosen so that they were high in imageability and frequency, and they were matched for number of syllables to the digits to enable better comparison across the two tasks. The TALSA is also unique in its scoring (described below); because the precise point of breakdown is captured in the participants' scores, the subtests provide a more accurate descriptor of digit and word span performance than the traditional method of two consecutive failures at any given string length.

The digit and word span subtests were administered using E-Prime (Psychology Software Tools, Pittsburgh, PA), once at baseline, once immediately following treatment, and once three months post-treatment. Administration and scoring of the spans followed the recommendations from Shelton, Martin, and Yaffee (1992). Participants were instructed to point to each string of digits or pictures in the same order that they heard them after the spans were administered. Participants indicated responses by pointing to digits or pictures on an array that appeared following stimulus presentation. The location of the digits and pictures changed on every presentation to ensure that participants could not rely on spatial memory to aid with verbal retention. Span lengths ranged from 2-7 items, and each length level consisted of 10 trials.

Participants' scores were derived by calculating the longest string length at which at least 5/10 of the strings were recalled, plus .5 times the proportion of the strings recalled in the next longer string length. To illustrate, consider a person who accurately recalls 6 out of 10 strings at the digit span length of 3, but then only correctly recalls 4 out of 10 strings at the span length of 4. First, because the participant did not correctly recall at least five items at span length 4 (i.e., did not recall at least 50% of the strings), the digit span task would be discontinued. Then, the score would be calculated as follows: 3 + .5(4/10) = 3.2. This same procedure would then be completed for the word span task. Using this method to calculate participants' scores yields a much more sensitive measure of their actual abilities; having scores from 10 trials per list length captures individual variability, while adding the proportion of lists correctly recalled at the highest list length achieved indicates the participants' maximal performance once variability has been accounted for.

Confrontation naming of trained and untrained nouns

Lexical retrieval abilities were quantified by accuracy in confrontation naming of pictures of trained and untrained stimuli. Forty trained and twenty untrained nouns were randomized and administered in a single confrontation naming task. The untrained nouns had similar phonotactic probability and neighborhood density characteristics as the trained items. The confrontation naming task was administered at three time points: prior to treatment, immediately post-treatment, and three months post-treatment. Both trained and untrained noun sets were administered three times at each time point, and these three scores were then averaged to create single mean trained and untrained noun naming scores for the time point.

During administration, the clinician asked participants to name a pictured object using only one word. Only the first naming attempt was considered, and the response was only scored correct if the entire word was produced accurately. A response was considered incorrect if any of the following were present: phonologic errors such as substitutions, additions, omissions, or transpositions; semantic errors; mixed phonologic and semantic errors; neologisms; unrelated real words; no response; and/or circumlocution. Participant responses were scored offline by trained research assistants. Intra-rater and inter-rater analyses were performed on 10% of the total corpus using Cohen's kappa to determine consistency among raters. The intra-rater reliability for the raters was found to be kappa = .99 ($p < .001$), and inter-rater reliability was found to be kappa = .94 ($p < .001$), indicating almost perfect agreement (Landis & Koch, 1977).

Results

Digit spans are historically significantly better than word spans across participants and studies (Jones & Macken, 2015); therefore, to determine whether the subtests from the TALSA should be analyzed separately or combined into a composite measure, paired Wilcoxon signed-rank tests were run for each time point. Statistically significant differences were found at all three time points: at baseline, digits ($M = 2.72$; $SD = 0.96$) and word ($M = 2.24$; $SD = 0.78$); $Z = -2.83$, $p = .005$, means differed; immediately post-treatment, digits ($M = 2.81$; $SD = 1.09$) and words ($M = 2.42$; $SD = 0.80$); $Z = -2.02$, $p = .044$, differed; and follow-up, digits ($M = 2.79$; $SD = 1.05$) and words ($M = 2.44$; $SD = 0.83$); $Z = -1.99$, $p = .047$, differed. Thus, mean digit and word span scores were analyzed separately.

Aim 1

This aim sought to answer the question of whether span scores predict magnitude of naming performance change for treated and untreated nouns immediately and three months post-treatment. Sequential entry multiple regression was used to evaluate this aim. The dependent variable was the mean trained noun naming accuracy scores immediately and 3 months post-treatment. All independent variables, including the mean pre-test naming scores, were standardized for ease of interpretation. Standardized mean trained noun naming accuracy at baseline was represented by the slope intercept. Span scores from the pre-treatment were standardized and, given the significant difference between scores on each measure, they were entered into the models separately to determine whether one measure was a better predictor of treatment change than the other.

In the first block of the regression, pre-treatment naming abilities accounted for significant variation of trained nouns immediately post-treatment, $R^2 = 0.86$ ($R^2_{adjusted} = 0.85$), $F(1, 22) = 0.95$, $p < .001$, follow-up trained nouns naming accuracy $R^2 = 0.90$ ($R^2_{adjusted} = 0.89$), $F(1, 22) = 0.95$, $p < .001$, untrained nouns immediately post-treatment, $R^2 = 0.96$ ($R^2_{adjusted} = 0.96$), $F(1, 22) = 0.98$, $p < .001$, and follow-up untrained noun naming accuracy, $R^2 = 0.94$ ($R^2_{adjusted} = 0.94$), $F(1, 22) = 0.97$, $p < .001$. Once pre-treatment naming ability was controlled for, TALSA digit and word spans did not account for any additional variation of trained nouns immediately post-treatment, $R^2_{change} = 0.00$ ($R^2_{adjusted} = 0.84$, $R^2_{total} = 0.86$), $F_{change}(2, 20) = 0.14$, $p > .05$, follow-up trained noun naming accuracy $R^2_{change} = 0.00$ ($R^2_{adjusted} = 0.88$, $R^2_{total} = 0.90$), $F_{change}(2, 20) = 0.003$, $p > .05$, untrained nouns immediately post-treatment, $R^2_{change} = 0.00$ ($R^2_{adjusted} = 0.96$, $R^2_{total} = 0.96$), $F_{change}(2, 20) = 0.15$, $p > .05$, or follow-up untrained noun naming accuracy, $R^2_{change} = 0.01$ ($R^2_{adjusted} = 0.94$, $R^2_{total} = 0.95$), $F_{change}(2, 20) = 0.99$, $p > .05$.

When all predictors were entered into the model, and holding all other variables constant, results demonstrated that average immediate post-treatment accuracy of trained noun naming was 20.87 ($SE = 0.94$), $t(20) = 22.18$, $p < .001$, follow-up accuracy of trained noun naming was 17.70 ($SE = 0.75$), $t(20) = 23.63$, $p < .001$, immediate post-treatment accuracy of untrained noun naming was 8.79 ($SE = 0.24$), $t(20) = 35.96$, $p < .001$, and finally, follow-up accuracy of untrained nouns was 8.82 ($SE = 0.28$), $t(20) = 31.83$, $p < .001$. Furthermore, pre-treatment naming accuracy also uniquely predicted immediate and follow-up untrained noun naming (all slope coefficient *t*-test *p*s < 0.01; see Tables 2-5). However, TALSA predicted neither accuracy of immediate post-treatment and follow-up trained noun naming, nor accuracy of immediate and follow-up untrained noun naming (all slope coefficient *t*-test *p*s > 0.05). Overall, these results suggest that pre-treatment naming ability is the most significant factor for predicting treatment

Table 2.

Multiple Linear Regression with Sequential Predictor Entry for Trained Nouns at Immediate Post-Treatment

	Block 1								Block 2							
	R^2_{change}		R^2_{total}		R^2_{adj}	*b*		sr^2	R^2_{change}	R^2_{total}		R^2_{adj}	*b*		sr^2
Model Fit	0.86	***	0.86	***	0.85				0.00	0.86	***	0.84			
Coefficients															
Intercept						20.79	***						20.87	***	
zTrained Nouns Pre-Tx						10.43	***	0.86					10.56	***	0.52
zTALSA Digits Pre-Tx													0.40		0.00
zTALSA Words Pre-Tx													-0.73		0.00

Note. Block 1 *F*-change test *df* = 1, 22; Block 2 *df* = 2, 20. Pre-Tx = Pre-treatment. TALSA = Temple Assessment of Language and Short-Term Memory in Aphasia. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.
Multiple Linear Regression with Sequential Predictor Entry for Trained Nouns at 3-Month Follow Up

	Block 1							Block 2						
	R^2_{change}		R^2_{total}		R^2_{adj}	*b*		*sr²*	R^2_{change}	R^2_{total}		R^2_{adj}	*b*		*sr²*
Model Fit	0.90	***	0.90	***	0.89				0.00	0.90	***	0.88			
Coefficients															
Intercept						17.71	***						17.70	***	
zTrained Nouns Pre-Tx						10.05	***	0.90					10.08	***	0.53
zTALSA Digits Pre-Tx													-0.09		0.00
zTALSA Words Pre-Tx													0.05		0.00

Note. Block 1 *F*-change test *df* = 1, 22; Block 2 *df* = 2, 20; Block 3 *df* = 2, 18. Pre-Tx = Pre-treatment. TALSA = Temple Assessment of Language and Short-Term Memory in Aphasia. * p < 0.05, ** p < 0.01, *** p < 0.001.

Table 4.
Multiple Linear Regression with Sequential Predictor Entry for Untrained Nouns at Immediate Post-Treatment

	Block 1								Block 2							
	R^2_{change}		R^2_{total}		R^2_{adj}	b		sr^2	R^2_{change}	R^2_{total}		R^2_{adj}	b		sr^2
Model Fit	0.96	***	0.96	***	0.96				0.00	0.96	***	0.96			
Coefficients															
Intercept						8.77	***						8.79	***	
zUntrained Nouns Pre-Tx						5.65	***	0.96					5.65	***	0.56
zTALSA Digits Pre-Tx													0.17		0.00
zTALSA Words Pre-Tx													-0.18		0.00

*Note.*Block 1 *F*-change test *df* = 1, 22; Block 2 *df* = 2, 20; Block 3 *df* = 2, 18. Pre-Tx = Pre-treatment. TALSA = Temple Assessment of Language and Short-Term Memory in Aphasia. * p < 0.05, ** p < 0.01, *** p < 0.001.

Table 5.

Multiple Linear Regression with Sequential Predictor Entry for Untrained Nouns at 3-Month Follow Up

	Block 1								Block 2						
	R^2_{change}		R^2_{total}		R^2_{adj}	*b*		sr^2	R^2_{change}	R^2_{total}		R^2_{adj}	*b*		sr^2
Model Fit	0.94	***	0.94	***	0.94				0.01	0.95	***	0.94			
Coefficients															
Intercept						8.76	***						8.82	***	
zUntrained Nouns Pre-Tx						5.36	***	0.94					5.45	***	0.57
zTALSA Digits Pre-Tx													0.39		0.00
zTALSA Words Pre-Tx													-0.58		0.00

Note. Block 1 *F*-change test *df* = 1, 22; Block 2 *df* = 2, 20; Block 3 *df* = 2, 18. Pre-Tx = Pre-treatment. TALSA = Temple Assessment of Language and Short-Term Memory in Aphasia. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

gains, while verbal short-term memory abilities, as measured by TALSA digit and word span tasks, are not.

Aim 2

This aim explored whether PMT improved verbal STM. Wilcoxon matched pairs signed-rank tests with Dunn-Sidak adjusted *p*-values were conducted to compare pre-treatment TALSA span abilities to immediate post-treatment and follow-up span lengths for digits and words separately. There were no significant differences in pre-treatment digit and immediate post-treatment digit spans, $Z = -0.02$, adjusted $p > .05$, or pre-treatment versus 3-month post-treatment digit spans, $Z = -0.47$, adjusted $p > .05$. Likewise, no differences were found between pre-treatment and immediate post-treatment $Z = -2.00$, adjusted $p > .05$, or pre-treatment compared to three months post-treatment $Z = -1.51$, adjusted $p > .05$, word span scores. This suggests that Phonomotor Treatment does not improve verbal short-term memory abilities, as measured by the TALSA digit and word span tasks. These results are summarized in Tables 6-7.

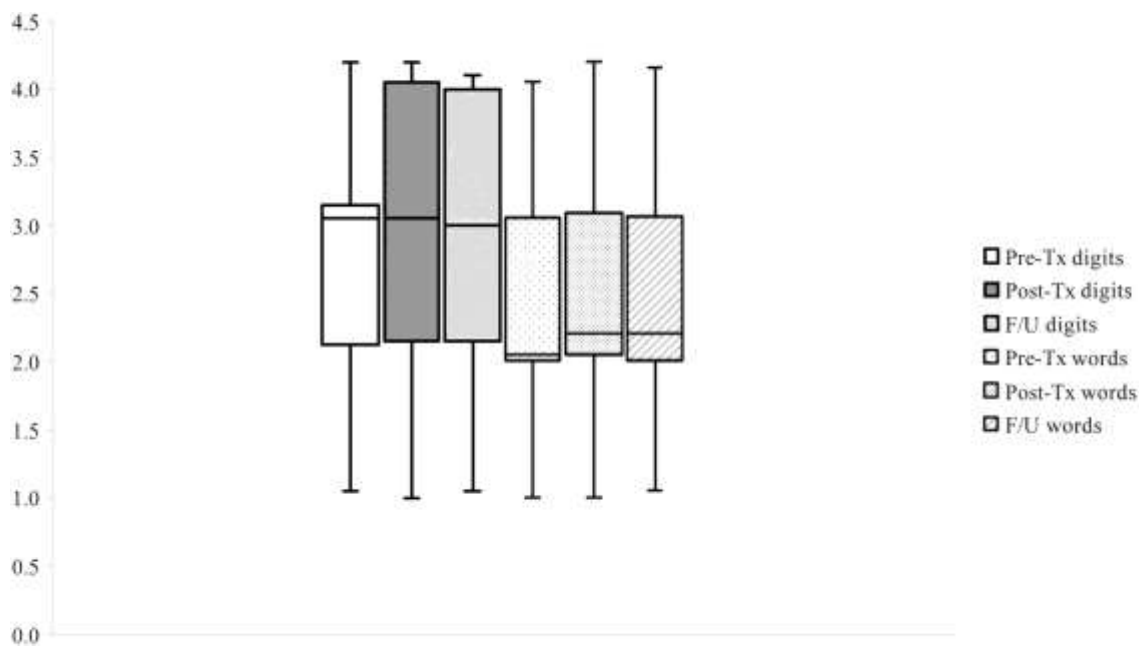

Figure 1. Overall performance on digit and word span tasks pretreatment (Pre-Tx), immediately posttreatment (Post-Tx), and at 3-month follow-up (F/U).

Table 6.
Median (Range) of Digit and Word Span scores at each time point, as well as the change between immediate post-treatment and 3-month follow up from pre-treatment.

	Period			**Changes**	
	Pre-Tx	Post-Tx	F/U	Post-Tx - Pre-Tx	F/U - Pre-Tx
Digit	*n* = 23	*n* = 21	*n* = 20	*n* = 21	*n* = 20
	3.05 (1.05, 4.20)	3.05 (1.00, 4.20)	3.00 (1.05, 4.10)	0.00 (-0.05, 0.00)	-0.05 (0.00, -0.10)
Word	*n* = 25	*n* = 22	*n* = 20	*n* = 22	*n* = 20
	2.05 (1.00, 4.05)	2.20 (1.00, 4.20)	2.20 (1.05, 4.15)	0.15 (-0.00, 0.15)	0.15 (0.05, 0.10)

Note. Pre-Tx = Pre-treatment assessment. Post-Tx = Post-treatment assessment. F/U = 3-month follow up assessment.

Table 7.
Wilcoxon Signed-rank test p*-values for changes in digit and word span by period*

	p-value
Digit Span	
Post-Tx vs Pre-Tx	0.981
F/U vs Pre-Tx	0.637
Word Span	
Post-Tx vs Pre-Tx	0.045
F/U vs Pre-Tx	0.132

Note. Pre-Tx = Pre-treatment assessment. Post-Tx = Post-treatment assessment. F/U = 3-month follow up assessment. Dunn-Sidak adjusted $p = 0.025$

Post-hoc analyses

A post-hoc Wilcoxon signed-rank test was conducted to determine whether differences existed between those with milder anomia and those with moderate-to-severe anomia on digit and word span changes following PMT. Group membership was determined using a median split-half method, where the median score on the BNT at baseline was determined (*Mdn* = 23.0); all those scoring at or above the median were categorized as “high” and all those scoring below the median were categorized as “low.” For digit spans, no differences were found between baseline, immediate post-treatment, or follow-up digit span scores for either group (all Dunn-Sidak adjusted *p*s > 0.05, see Tables 8-9). For word spans, no differences were found between baseline and immediate post-treatment scores for either group, nor were word spans significantly different between baseline and follow-up for the low group. A significant increase was found between baseline and follow-up word span scores for the high group ($Z = -2.34$, adjusted $p = .019$). That is, the only significant change in verbal STM ability, as measured by word span scores, was for those with less severe anomia three months after treatment termination.

Table 8.
Median (Range) of Digit and Word Span scores at each time point by group and change between immediate post-treatment and 3-month follow up from pre-treatment.

		Period			**Changes**	
		Pre-Tx	Post-Tx	F/U	Post-Tx - Pre-Tx	F/U - Pre-Tx
High Group	**Digit**	*n* = 13	*n* = 13	*n* = 12	*n* = 13	*n* = 12
		3.15 (1.20, 4.20)	3.10 (1.05, 4.20)	3.58 (1.05, 4.10)	-0.05 (-0.05, 0.00)	0.43 (-0.15, -0.10)
	Word	*n* = 13	*n* = 13	*n* = 12	*n* = 13	*n* = 12
		3.00 (2.00, 4.05)	3.05 (2.00, 4.20)	3.05 (2.00, 4.15)	0.05 (0.00, 0.15)	0.05 (0.00, 0.10)
Low Group	**Digit**	*n* = 10	*n* = 8	*n* = 8	*n* = 8	*n* = 8
		2.15 (1.05, 3.05)	2.13 (1.00, 3.10)	2.20 (1.10, 3.00)	-0.02 (-0.05, 0.05)	0.05 (0.05, -0.05)
	Word	*n* = 12	*n* = 9	*n* = 8	*n* = 9	*n* = 8
		2.00 (1.00, 3.05)	2.05 (1.00, 3.00)	2.00 (1.05, 2.20)	0.03 (0.00, -0.05)	0.00 (0.05, -0.85)

Table 9.
Wilcoxon Signed-rank test p-*values for changes in digit and word span by period, by group*

	High Group p-value	Low Group p-value
Digit Span		
Post-Tx vs Pre-Tx	0.857	0.595
F/U vs Pre-Tx	0.953	0.500
Word Span		
Post-Tx vs Pre-Tx	0.140	0.255
F/U vs Pre-Tx	0.019*	0.257

Note. Pre-Tx = Pre-treatment assessment. Post-Tx = Post-treatment assessment. F/U = 3-month follow up assessment.
Dunn-Sidak adjusted $p = 0.025$

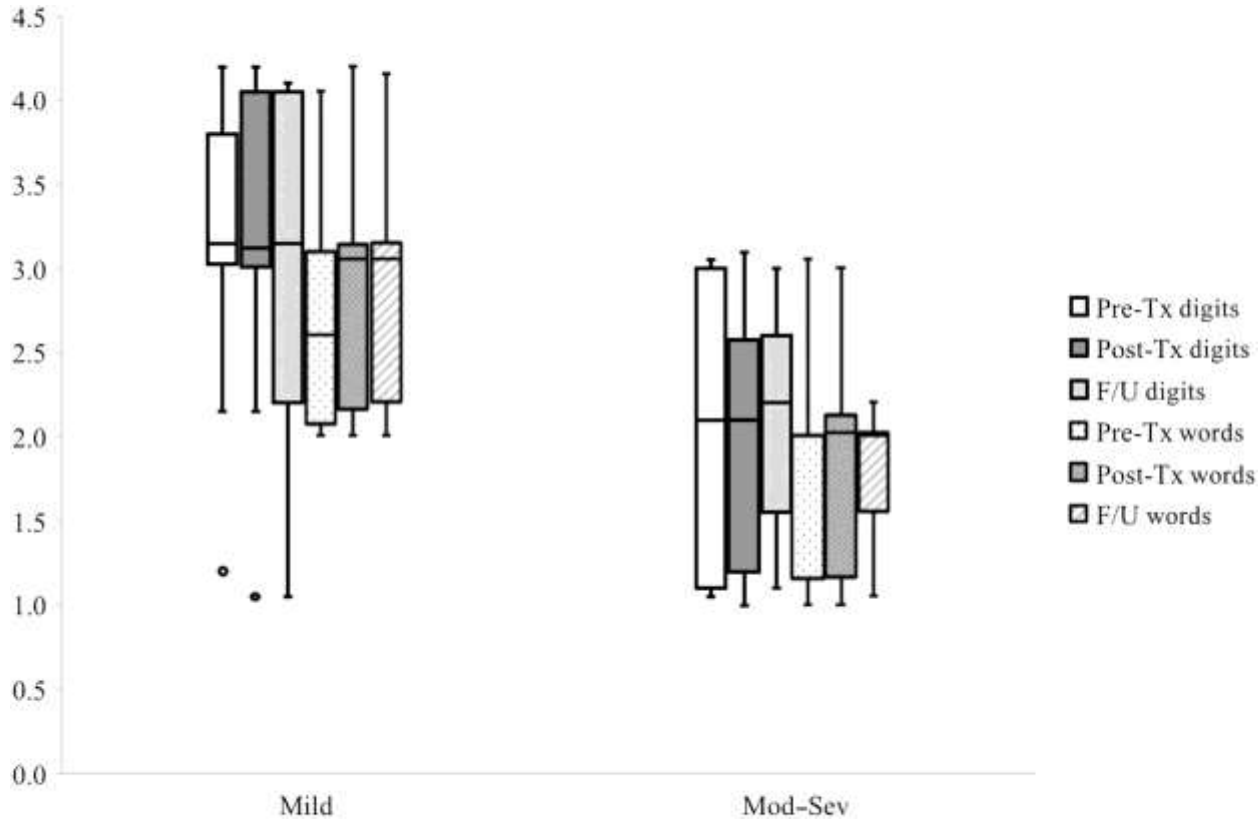

Figure 2. Performance by group on digit and word span tasks pretreatment (Pre-Tx), immediately posttreatment (Post-Tx), and at 3-month follow-up (F/U). Mod–Sev = moderate–severe.

Discussion

The aims of this study were to determine (1) whether verbal STM predicted the magnitude of treatment gains following treatment and (2) whether the treatment, PMT, improved verbal STM. Participants completed confrontation naming and digit and word span tasks prior to treatment, immediately following treatment termination, and at the 3-month follow up. Results indicate that digit and word span scores did not predict magnitude of treatment gains for trained or untrained words immediately post-treatment or at follow up. Results also suggest that PMT did not increase verbal STM, as measured by digit and word span scores, immediately post-treatment or at follow-up. However, a post-hoc analysis of span changes for aphasic subgroups based on severity revealed that PWA with milder anomia did demonstrate significantly increased word spans at follow up.

Verbal STM as a Predictor of Anomia Treatment Gains

For Aim 1, digit and word spans were not predictive of magnitude of treatment change. Two possible reasons for this finding are that the verbal STM tasks (a) rely on different processes than the outcome measure and/or (b) do not measure capacity for verbal learning. On a very basic level, the span tasks administered in this study relied on receptive abilities, while naming entails expression. Across the literature, PWA have been found to demonstrate dissociations in receptive and expressive skills (Dick et al., 2001). Thus, it is possible that the pointing condition of the digit and word spans were poor predictors in this study because they do not yield any information about the integrity of verbal output processing (Martin & Ayala, 2004; Martin & Saffran, 2002). Furthermore, digit and word spans scores may be a static representation of verbal STM in receptive modalities, instead of a representation of the dynamic and complex processes that undergird verbal learning. That is, a digit or word span score at baseline may only provide a snapshot of input capacity at that particular point in time, when what is needed is an indicator of acquisition processes. If this is indeed the case, it would explain why span scores are highly correlated with general naming ability at any given point in time but are not predictive of treatment outcomes (Best et al., 2013; Martin & Ayala, 2004). A more sophisticated measure may be required if the goal is to determine the predictive value of verbal STM in anomia treatment.

Other subtests of the TALSA, such as the two memory load conditions of the confrontation naming test, may be better indicators of verbal STM in the output domain. The TALSA includes an adapted version of the Philadelphia Naming Test (Roach, Schwartz, Martin, Grewal, & Brecher, 1996) that includes 90 of the original 175 items. Participants are shown a line drawing and then, after either one second or five seconds, they receive a prompt to name the item. The benefit of this measure would be to gain insight into activation maintenance across the two conditions in an output only modality. We suggested above that the verbal STM task needs to more closely align with the outcome measure. The TALSA naming task is nearly identical to the outcome measure (i.e., they are both confrontation naming tasks), with the key difference being that the two interval conditions provide insight into how time influences processing. Martin and colleagues have found that for some PWA, the 5-second interval improves naming performance, but for the majority of subjects whose performance changed across conditions, naming accuracy decreases from the 1-second to the 5-second condition (Martin & Dell, 2017; Martin et al., 2018).

A possible explanation for the change in performance may be related to two key mechanisms of activation transmission in an interactive activation model of language: For those whose performance increased with the delay, connection strength between linguistic representations may have been weak, resulting in slowed activation transmission. Therefore, the extra time may have allowed for sufficient activation to reach the target lexeme and its constituent components to allow for successful retrieval. On the other hand, for those whose naming accuracy decreased over time, the decay rate of activation may be too rapid to allow for successful retrieval. In either case (i.e., weak connection strength or overly rapid activation decay), differential performance between conditions could be a more fruitful predictor of treatment response. That is, the change or lack thereof in naming performance across conditions may serve as a proxy for individuals' processing (i.e., verbal STM) abilities, which could in turn be used as a predictor variable in future treatment studies.

Another alternative measure of verbal STM in this context may be the Hopkins Verbal Learning Test-Revised (HVLT-R; Brandt & Benedict, 2001), which does seem to have a significant predictive relationship with treatment outcomes for PWA (Dignam et al, 2017). The HVLT-R consists of a training phase and a recognition/recall phase. In the training phase, the subject hears a list of twelve words (one every two seconds), then recalls all of the words they can remember in any order. The subject then repeats this process two more times. The recall accuracy from each of the three trials are combined to yield a "total recall" score. After approximately 20 minutes, the examiner asks the subject to recall as many words as possible. The recall accuracy from this trial is the "delayed recall" score and reflects transfer into long-term memory of the test items. The "retention" score is the proportion of words from the participant's best training trial score that he/she can recall after the delay (i.e., delayed recall score divided by highest training trial score). The test also consists of a recognition trial, in which participants are told to listen to a list of words and identify any words that recognize from earlier trials. This yields the "recognition hits" score.

In the context of using the HVLT-R as a predictor of treatment gains, using the score from the first trial would be the only relevant measure, as the subsequent trial scores could reflect the influence of long-term memory. It must be noted that using the test in this way warrants cautious interpretation of results. While the first trial score may yield some information – although not likely any more than what the word span provides – the real value of the HVLT-R is the insight it offers into verbal learning. Because anomia treatment is fundamentally a form of teaching, having a measure of how much and how quickly a participant can glean verbal information may be better suited to unravelling the relationship between memory and treatment success.

It is worth discussing whether measures of WM may have yielded different results than the digit and word span tasks used in this study. Working memory span tasks such as operation (Turner & Engle, 1989) or reading (Daneman & Carpenter, 1980) spans could be used both in research and clinical settings. In an operation span, subjects see a math equation with the solution provided and are asked to indicate whether the answer is correct or incorrect. Following the equation, they see a word written in lowercase letters, with the instruction to remember the word for later recall. Following a random number of equation-word items, the participant is prompted to recall all of the words that they saw. The score is the number of words correctly recalled in serial order. In the original reading span task, subjects are instructed to read a sentence aloud and remember the word at the end of the sentence for later recall. After a set of sentences, the subject is prompted to recall all of the sentence-final words in serial order. Subjects are presented three sets of increasing numbers of sentences, from 2-6. The score is calculated as the level for which the subjects correctly recalled all of the words in order on two of three sets. For PWA, this task could be modified to minimize a decrement in performance due to language and/or motor speech impairment, for example by having subjects listen to instead of read sentences. This alteration would make the task a listening span, rather than a reading span.

Both operation and reading spans – but not backward digit spans – have been correlated with fluid intelligence (Engle et al, 1999), which is thought to underlie problem-solving and reasoning abilities, as well as adaptability (Cattell, 1963, 1971; Horn & Cattell, 1967). Furthermore, Engle and colleagues also found that these tasks specifically rely on central executive functioning and attention, which allows for maintenance of information in the face of interference. Thus, operation and reading/listening span scores derived from PWA prior to treatment could potentially measure their ability to maintain activation of treatment targets long enough for consolidation into long-term memory, as well as their ability to apply this new learning to novel tasks. Such a measure could serve as a better predictor of anomia treatment gains than simple span tasks alone.

VSTM Changes Following Anomia Treatment

For Aim 2, there was no significant change overall in word and digit span performance immediately following treatment or at follow up, likely because the sample was too small and heterogeneous to reveal any improvements. An additional consideration is that span scores have proven to be relatively stable from pre-to-post-treatment assessments in other studies (Best et al., 2013; Harnish & Lundine, 2015). Thus, it is possible that performance on these measures is less likely to change except in the case of some fundamental shift in processing abilities, such as improved verbal STM resulting from a memory-focused treatment. Digit span scores, in particular, may be less susceptible to changes in linguistic processing, as evidence suggests that digit span abilities may tap into a specific form of long-term sequence knowledge (Jones & Macken, 2015). That is, examinees may unconsciously rely on knowledge of number sequences, gained from experiences with random number sets in daily life, which results in longer digit spans. Despite increasing evidence that long-term knowledge may be propping up performance on this task, digit spans continue to be the gold standard for testing STM.

Both of these explanations are supported by the post hoc analysis, which looked at whether digit and word span scores changed over time in the less severe (or “high”) versus the more severe (or “low”) group. No significant changes were found for either group immediately post-treatment or for the low group’s digit and word span scores at follow up. However, the high group demonstrated significant change at the follow-up assessment point. Namely, word span scores for the less impaired PWA were significantly increased from baseline at three months post-treatment, but digit span scores were not. Other changes were noted for the high group, such as digit scores 3 months post-treatment and word scores immediately post-treatment, but these did not reach significance. It is unclear whether these changes are due to variable intra-individual performance across time points and/or differences in sample sizes across times points (all available data were used to increase power). This issue warrants further investigation in the future. Regardless, the fact that the group showed improvement suggests the possibility that people with less severe anomia experienced some fundamental change in their ability to process and maintain verbal information. Milder participants progressed to more difficult stimuli in PMT than those who were more severe, which provided more intense “training” for verbal STM processes. Furthermore, it is likely that the high group participated more in everyday conversations than the low group following PMT, which continued to strengthen their language processing abilities. This mechanism of change following PMT is tentative and requires more investigation. Nevertheless, the shift in abilities at follow-up was obscured in the original analysis of the larger group but, by dividing the participants into high and low groups, the data became slightly less heterogeneous, thereby revealing that significant changes to span capacity can occur.

It is interesting that the significant change only occurred at three months post-treatment and only for word spans. This finding is consistent with some of the extant literature on PMT, in that participants have demonstrated continued improvement beyond treatment termination (Kendall et al., 2018; Kendall et al., 2008; Lai, unpublished data). Post-treatment language improvements may result from the treatment re-establishing the basic building blocks of language (i.e., phonemes and phoneme sequencing), which then leads to improved lexical and semantic processing. As the connections between representations are strengthened, the activation between them improves, such that milder PWA may be able to participate in daily conversation more. This participation may, in turn, fortify language representations and activation strength. As a result, the course of activation from phonology to semantics and back may serve to enhance participants' ability to maintain activation in a span task. It is unclear why only word span changed significantly, given the ubiquity of numbers in everyday contexts. It is possible that this difference may reflect some difference in semantic processing for digit versus non-digit words, but this is a topic that requires more exploration in future studies.

Limitations and Future Directions

Although aspects of this study have been very illuminating, very clear limitations exist; namely, the measure of verbal STM and the sample size. As has been discussed, a more appropriate test of verbal STM may yield different and/or more predictive results than in the current study. Because the current study is a retrospective analysis of data from a larger randomized control trial focused on anomia treatment efficacy, the verbal STM measures were very limited in their scope. This limitation negatively impacted the ability of the current study to adequately characterize the verbal STM abilities of this sample. In future studies, it would be prudent to use measures that better reflect the skills necessary to complete PMT. For example, more appropriate TALSA subtests might include TALSA naming test, as well as phonological probe memory spans. As noted above, the difference in performance between the 1- and 5-second interval conditions of the naming test could indicate the role of memory in an output modality. Probe memory spans are thought to rely more on WM, as participants are required to maintain increasingly longer strings of words for comparison to a probe word. This ability would also be necessary in a protocol like PMT, which utilizes intensive phonological awareness tasks. Martin et al. (2018) reported that PWA perform considerably worse than neurologically healthy controls on naming and probe span tasks, which would imply that these tests are also more sensitive than the measures that were excluded from this analysis: rhyming judgments, category judgments, rhyming triplet judgments, and synonymy triplet judgments.

The second limitation in this study, like many other aphasia studies, is the small sample size. A post-hoc power analysis was conducted using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) to reveal an achieved power of .27. Even with its larger sample relative to the extant aphasia literature, this study has a severely limited ability to reject the null hypothesis. An a priori power analysis with G*Power suggests that at least 59 participants would be needed to detect a small effect size (0.20; Cohen, 1988) with three predictor variables, as in the current analysis, and if a TALSA naming test change score, and phonological probe spans, plus baseline naming scores were to be used. The feasibility of such a large-scale study is questionable, given the resources required by both the research team and any potential participants, but this is nevertheless a goal to aspire to.

Conclusion

This project was a theoretically motivated attempt to determine (a) whether verbal STM is a predictor of the magnitude of change that participants experience in word retrieval abilities following PMT and (b) whether PMT improves verbal STM. The analyses in this study have suggested that PMT does not improve digit and word span scores in a group with a wide range of severities, and that these scores do not predict change following PMT. However, those with milder anomia did demonstrate significant changes in their verbal STM ability, which has interesting implications for the mechanisms of PMT. Ultimately, these results have revealed an underlying question that is less about the relationship between PMT and verbal STM, and more about whether the measures used to represent verbal STM are adequate to arrive at the answers we seek. Through careful consideration and selection of appropriate verbal STM measures, future studies may paint a clearer picture of the significance of verbal STM as it relates to anomia treatment success.

Acknowledgements

The authors would like to acknowledge the participants and their families for their time and efforts. The second author was funded by VA RD&D Merit Review Grant #C6572R. We are very grateful to Megan Moldestad and Wesley Allen for data collection, Janaki Torrence for project management, James Proudfoot for his statistical input, Kerry Lam for her assistance in recruiting and scheduling participants, and Jenny Erpenbeck for managing the data analysis team. We acknowledge the students at the University of Washington who were involved in data analysis and the Aphasia Network (Oregon) for participant recruitment. The contents of this publication do not represent the views of the U.S. Department of Veterans Affairs or the United States government.

References

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J., & Howard, D. (2013). Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex, 49*(9), 2345-2357. https://doi.org/10.1016/j.cortex.2013.01.005

Boyle, M. (2001). Semantic Feature Analysis: The Evidence for Treating Lexical Impairments in Aphasia. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders, 11*(2), 23. https://doi.org/10.1044/nnsld11.2.23

Boyle, M. (2010). Semantic Feature Analysis Treatment for Aphasic Word Retrieval Impairments: What's in a Name? *Topics in Stroke Rehabilitation, 17*(6), 411-422. https://doi.org/10.1310/tsr1706-411

Brandt, J., & Benedict, R. H. B. (2001). *Hopkins Verbal Learning Test - Revised*. Odessa, FL: Psychological Assessment Resources.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

Coelho, C. A., McHugh, R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology, 14*(2), 133-142. https://doi.org/10.1080/026870300401513

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed. ed.). Hillsdale, N.J.: Hillsdale, N.J. : L. Erlbaum Associates.

Coltheart, M. (1981). *MRC Psycholinguistic Database.* Retrieved from http://websites.psychology.uwa.edu.au/school/MRCDatabase/mrc2.html

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163-191. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*(7), 323-338. https://doi.org/10.1016/S0079-6123(07)00020-9

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283-321.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287-314.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801-838.

Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N., & Gernsbacher, M. A. (2001). Language deficits, localization, and grammar: evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological review*, *108*(4), 759.

Dignam, J., Copland, D., Brien, K., Burfein, P., Khan, A., & Rodriguez, A. D. (2017). Influence of Cognitive Ability on Therapy Outcomes for Anomia in Adults with Chronic Poststroke Aphasia. *Journal of Speech, Language, and Hearing Research*, *60*(2), 406-421. https://doi.org/10.1044/2016_JSLHR-L-15-0384

Duffy, J. R. (2013). *Motor speech disorders : substrates, differential diagnosis, and management* (3rd ed. ed.). St. Louis, Mo.: St. Louis, Mo.: Elsevier Mosby.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working Memory, Short-Term Memory, and General Fluid Intelligence: A Latent-Variable Approach. *Journal of Experimental Psychology: General*, *128*(3), 309-331. https://doi.org/10.1037/0096-3445.128.3.309

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G * Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Goldenberg, G., Dettmers, H., Grothe, C., & Spatt, J. (1994). Influence of linguistic and non-linguistic capacities on spontaneous recovery of aphasia and on success of language therapy. *Aphasiology, 8*(5), 443-456. https://doi.org/10.1080/02687039408248669

Harnish, S. M., & Lundine, J. P. (2015). Nonverbal Working Memory as a Predictor of Anomia Treatment Success. *Am J Speech Lang Pathol, 24*(4), S880-894. https://doi.org/10.1044/2015_ajslp-14-0153

Harris, L., Olson, A., & Humphreys, G. (2014). The link between STM and sentence comprehension: A neuropsychological rehabilitation study. *Neuropsychological Rehabilitation, 24*(5), 678-720. https://doi.org/10.1080/09602011.2014.892885

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26,* 107-129.

Howard, D., Swinburn, K., & Porter, G. (2010). Putting the CAT out: What the Comprehensive Aphasia Test has to offer. *Aphasiology, 24*(1), 56-74. https://doi.org/10.1080/02687030802453202

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, *144*(C), 1-13. https://doi.org/10.1016/j.cognition.2015.07.009

Kaplan, E. (2001). Boston naming test. In H. Goodglass, E. Kaplan, O. Segal, & S. Weintraub (Eds.), (Second edition. ed.). Austin, Texas: Austin, Texas Pro-ed.

Kaplan, E., Goodglass, H., Segal, O., & Weintraub, S. (2001). Boston naming test. In (Second edition. ed.). Austin, Texas: Austin, Texas Pro-ed.

Kasselimis, D. S., Simos, P. G., Economou, A., Peppas, C., Evdokimidis, I., & Potagas, C. (2013). Are memory deficits dependent on the presence of aphasia in left brain damaged patients? *Neuropsychologia*, *51*(9), 1773-1776. https://doi.org/10.1016/j.neuropsychologia.2013.06.003

Kendall, D. L., del Toro, C., Nadeau, S. E., Johnson, J., Rosenbek, J., & Velozo, C. (2010). *The development of a standardized assessment of phonology in aphasia*. Paper presented at the Clinical Aphasiology Conference, Isle of Palms, SC.

Kendall, D. L., Oelke, M., Allen, W., Torrence, J., & Nadeau, S. E. (2018). *Phonomotor versus Semantic Feature Analysis Treatment for anomia in 58 persons with aphasia: A randomized controlled trial*. Manuscript under revision.

Kendall, D. L., Rosenbek, J. C., Heilman, K. M., Conway, T., Klenberg, K., Gonzalez Rothi, L. J., & Nadeau, S. E. (2008). Phoneme-Based Rehabilitation of Anomia in Aphasia. *Brain and Language*, *105*(1), 1-17. https://doi.org/10.1016/j.bandl.2007.11.007

Lambon Ralph, M. A., Snell, C., Fillingham, J. K., Conroy, P., & Sage, K. (2010). Predicting the outcome of anomia therapy for people with aphasia post CVA: Both language and cognitive status are key predictors. *Neuropsychological Rehabilitation*, *20*(2), 289-305. https://doi.org/10.1080/09602010903237875

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Majerus, S., Attout, L., Artielle, M. A., & Van der Kaa, M. A. (2015). The heterogeneity of verbal short-term memory impairment in aphasia. *Neuropsychologia*, *77*, 165-176. https://doi.org/10.1016/j.neuropsychologia.2015.08.010

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. https://doi.org/10.1016/j.bandl.2003.12.004

Martin, N., & Dell, G. S. (2017). Adding temporal dynamics to models of impaired language production. Baltimore, MD. Academy of Aphasia, November 4-6, 2017.

Martin, N., & Gupta, P. (2004). Exploring the relationship between word processing and verbal short-term memory: Evidence from associations and dissociations. *Cognitive Neuropsychology*, *21*(2-4), 213-228. https://doi.org/10.1080/02643290342000447

Martin, N., Kohen, F., Kalinyak-Fliszar, M., Soveri, A., & Laine, M. (2012). Effects of working memory load on processing of sounds and meanings of words in aphasia. *Aphasiology, 26*(3-4), 462-493. https://doi.org/10.1080/02687038.2011.619516

Martin, N., Minkina, I., Kohen, F. P., & Kalinyak-Fliszar, M. (2018). Assessment of linguistic and verbal short-term memory components of language abilities in aphasia. *Journal of Neurolinguistics*. https://doi.org/10.1016/j.jneuroling.2018.02.006

Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14*(5), 641-682.

Martin, N., & Saffran, E. M. (2002). The relationship of input and output phonological processing: An evaluation of models and evidence to support them. *Aphasiology*, *16*(1-2), 107-150. https://doi.org/10.1080/02687040143000447

McNeil, M. R., Fassbinder, W., Dickey, M. W., Hula, W. D., Zhou, L., . . . Forman, S. L. (2018). *Aphasic Comprehension: Conflict Resolution and Short-Term Memory*. Manuscript in preparation.

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. https://doi.org/10.1055/s-0036-1597261

National Aphasia Association. (n.d.) Aphasia FAQs. Retrieved from: https://www.aphasia.org/aphasia-faqs/.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, *16*(10-11), 935-979. https://doi.org/10.1080/02687030244000563

Pettigrew, C., & Hillis, A. E. (2014). Role for memory capacity in sentence comprehension: Evidence from acute stroke. *Aphasiology*, *28*(10), 1258-1280. https://doi.org/10.1080/02687038.2014.919436

Peñaloza, C., Mirman, D., Cardona, P., Juncadella, M., Martin, N., Laine, M., & Rodríguez-Fornells, A. (2017). Cross-situational word learning in aphasia. *Cortex*, *93*, 12-27. https://doi.org/10.1016/j.cortex.2017.04.020

Potagas, C., Kasselimis, D., & Evdokimidis, I. (2011). Short-term and working memory impairments in aphasia. *Neuropsychologia*, *49*(10), 2874-2878. https://doi.org/10.1016/j.neuropsychologia.2011.06.013

Raven, J. C., Court, J. H., & Raven, J. E. (1998). Raven's coloured progressive matrices. In *Coloured progressive matrices*. San Antonio, Tex.: San Antonio, Tex. : Harcourt Assessment.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. Clinical Aphasiology, 24, 121–133.

Seniow, J., Litwin, M., & Lesniak, M. (2009). The relationship between non-linguistic cognitive deficits and language recovery in patients with aphasia. *J Neurol Sci*, *283*(1-2), 91-94. https://doi.org/10.1016/j.jns.2009.02.315

Tuomiranta, L., Grönroos, A.-M., Martin, N., & Laine, M. (2014). Vocabulary acquisition in aphasia: Modality can matter. *Journal of Neurolinguistics*, *32*, 42-58. https://doi.org/10.1016/j.jneuroling.2014.08.006

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127-154.

Van De Sandt-Koenderman, W. M. E., Remerie, S. C., Van Der Voort-Klees, Y. A., Wielaert, S. M., Ribbers, G. M., Van Harskamp, F., . . . Duivenvoorden, H. J. (2008). MAAS (Multi-axial Aphasia System): Realistic goal setting in aphasia rehabilitation. *International Journal of Rehabilitation Research, 31*(4), 314-320. https://doi.org/10.1097/MRR.0b013e3282fc0f23

Vitevich, M. S., & Luce, P. A. (1998). Phonotactics, phonological neighborhoods, and working memory in spoken word recognition. *The Journal of the Acoustical Society of America, 104*(3), 1758-1758. https://doi.org/10.1121/1.423698

Ween, J. E., Verfaellie, M., & Alexander, M. P. (1996). Verbal memory function in mild aphasia. *Neurology, 47*(3), 795-801.