ARTICLE IN PRESS

Journal of Neurolinguistics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Research paper

Influence of working memory on stimulus generalization in anomia treatment: A pilot study

Stacy M. Harnish[a,*], Deena Schwen Blackett[a], Alexandra Zezinka[a], Jennifer P. Lundine[a], Xueliang Pan[b]

[a] *Department of Speech and Hearing Science, The Ohio State University, United States*
[b] *Center for Biostatistics, Department of Biomedical Informatics, The Ohio State University, United States*

ARTICLE INFO

Keywords:
Aphasia
Anomia
Therapy
Working memory
Generalization

ABSTRACT

Neuropsychological testing of distinct cognitive domains holds promise as a prognostic indicator of aphasia therapy success; however, it is unclear the degree to which cognitive assessments may also predict generalization abilities. The present study aimed to assess the relationship between working memory skills and stimulus generalization from a visual picture-naming treatment to an auditory definition-naming task. Seven individuals with aphasia completed verbal and nonverbal assessments of working memory prior to participating in a cued picture-naming treatment for anomia. After treatment ended, stimulus generalization percentages were calculated for definition naming for the same items that were trained using picture naming. Scores on two nonverbal working memory measures, the backward spatial span and the 1-back, and one verbal working memory assessment, the picture span, were positively correlated with generalization percentage. These results provide preliminary evidence of the relationship between working memory and stimulus generalization. When comparing performance across working memory measures, the spatial span and the picture span were highly correlated in this sample. We propose that despite the verbal and nonverbal distinction, these tasks may have tapped into working memory similarly by relying on a shared central processing mechanism.

1. Introduction

Generalization of treatment gains can be classified into two broad categories: response generalization and stimulus generalization. Response generalization refers to generalization of therapy gains from trained items to untrained items, whereas stimulus generalization refers to the ability to generate a learned response in a different stimulus environment (i.e. from trained tasks to untrained tasks). Treatments for post-stroke word-retrieval deficits have generally resulted in poor response generalization (Coppens & Patterson, 2018; Snell, Sage, & Lambon Ralph, 2010; Wambaugh, Mauszycki, & Wright, 2014). However, there is a body of work showing that individuals perform better on naming untrained items when they share semantic properties with trained items (Kiran & Thompson, 2003) and when the trained items are less prototypical exemplars of a semantic category than are the untrained items (Kiran, Sandberg, & Sebastian, 2011). Although response generalization tends to be poor with anomia treatments, best chances occur when a strategy for word retrieval is internalized during training so that it can be used with novel exemplars (Coppens & Patterson, 2018).

---

[*] Corresponding author. Department of Speech and Hearing Science, The Ohio State University, 1070 Carmack Rd, 110 Pressey Hall, Columbus, OH 43210, United States.

E-mail address: harnish.18@osu.edu (S.M. Harnish).

https://doi.org/10.1016/j.jneuroling.2018.02.003
Received 14 July 2017; Received in revised form 2 February 2018; Accepted 6 February 2018
0911-6044/

Please cite this article as: Harnish, S.M., Journal of Neurolinguistics (2018), https://doi.org/10.1016/j.jneuroling.2018.02.003

Stimulus generalization also tends to be minimal, at best, for word-retrieval therapies (Boyle & Coelho, 1995; Boyle, 2004), many of which rely on picture naming paradigms. Demonstrating generalization of picture naming treatment gains to an untrained context is methodologically challenging (Conroy, Sage, & Lambon Ralph, 2009) due to the difficulty in providing opportunities to use words trained during picture naming for another task. For example, an untrained story retell task must provide the appropriate context so that an individual has the opportunity to produce each of the words trained during picture naming treatment during the story retell task. This challenge is compounded by the need to choose individually-tailored stimulus sets that target specific deficits in word retrieval, as well as the considerable variability in exemplars produced for a target word from even neurologically-intact individuals in response to retelling of a narrative (Armstrong, 2000). These factors make choosing specific stimulus items and modes of elicitation (e.g., narrative to retell, picture description) for measurement of stimulus generalization a challenging endeavor.

1.1. Therapy studies investigating stimulus generalization

Studies investigating stimulus generalization have used different elicitation methods to assess the transfer of treatment gains from trained items to untrained tasks. Elicitation tasks have included story retell (Conroy et al., 2009; Faroqi-Shah & Virion, 2009; Hickin, Mehta, & Dipper, 2015; Maher et al., 2006; Rider, Wright, Marshall, & Page, 2008; Rose, Mok, Carragher, Katthagen, & Attard, 2016), picture description (Conroy et al., 2009; Rose et al., 2016), and conversation (Best et al., 2011, 2013; Boo & Rose, 2011; Conroy et al., 2009; del Toro et al., 2008; Grande et al., 2008; Greenwood, Grassly, Hickin, & Best, 2010; Rose, Douglas, & Matyas, 2002; Rose et al., 2016). These studies have produced mixed results, in part due to the methodological difficulties in demonstrating this effect, especially to more natural conversational tasks. Gross measurements focused on quantifying improvement in discourse (e.g., number of utterances, words, novel utterances, or correct information units) often provide insufficient evidence of item-specific generalization of trained items, as they do not elicit production of specific words trained in therapy. Although valuable as a measure of functional language abilities, they also leave us wondering whether the lack of demonstrated generalization may be related to additional task demands, such as syntactic planning for longer utterances. In addition, some investigations may be blurring the line between stimulus and response generalization. That is, measuring verb retrieval in conversation after a picture-naming therapy may be measuring both stimulus generalization (e.g. a different task for retrieval) and response generalization (e.g. retrieval of any verbs and not specifically the verbs trained in therapy). Similarly, when examining whether a treatment has generalized to overall language abilities using standardized measures, both the context and the specific items will vary from the trained context and items. Although these outcomes tend to demonstrate better ecological validity than does picture naming, they may not be constrained enough to demonstrate pure stimulus generalization without other confounding variables.

Some studies investigating stimulus generalization of trained items to untrained contexts have attempted to isolate target lexical items as opposed to measuring improved production of all lexical items. That is, the generalization task is designed to elicit specific words that have been trained in therapy. For example, in a case study, Hashimoto and Frome (2011) found stimulus generalization after a modified semantic feature analysis treatment from naming black and white line drawings to naming the same items in colored photographs in their natural setting (e.g. all animals in a zoo setting). Rider et al. (2008) investigated generalization of confrontation naming gains after semantic feature analysis on discourse production in closed-set contexts. Target words came from story retellings and procedural explanations. They found an increase in the number of target words produced in discourse after therapy. Similarly, Conroy et al. (2009) investigated generalization of improved confrontation naming after picture naming therapy to picture supported narratives, using the Cookie Theft picture and the Cinderella story, and to unsupported narratives, using the Cinderella story without the book. Unsurprisingly, they found a stepwise decrement in naming from picture naming to supported narrative to unsupported narrative. However, picture naming accuracy in the confrontational naming task significantly predicted accuracy in the connected language tasks. In these studies, some of the previously mentioned confounding variables posed by discourse/conversational tasks were more tightly controlled because items trained in therapy were also targeted during stimulus generalization tasks.

One criticism of picture-naming therapies is that the task is not functionally relevant. Picture-naming therapy may change a subset of the neural circuitry involved in lexical retrieval (Kleim & Jones, 2008), resulting in improved picture naming but without much generalization to word retrieval in other contexts. If we measure success in anomia therapy with task specific acquisition of a lexical item, then we fail to meet the critical need for translation of that skill to functional communication tasks that may actually impact communicative abilities and quality of life. Intensive restorative therapy aims to rehabilitate the impaired language processes (as opposed to therapy that seeks to compensate for a language loss using specific strategies), and can be a powerful tool for rehabilitation of language (Cherney, Patterson, & Raymer, 2011; Dignam, Rodriguez, & Copland, 2016; Harnish et al., 2014). However, if we achieve improved lexical retrieval in the context of a picture stimulus, we cannot assume that improved retrieval will occur in the absence of the picture. Thus, to facilitate real change in communicative competence, we must either assess and treat generalization of those lexical items across contexts or equip patients with portable picture stimuli to facilitate retrieval in the context in which they were trained. Without these steps, we are falling short of our goal to rehabilitate communication abilities by discharging patients with improved picture naming skills, for example, but without any translation to functional skills.

1.2. Definition-naming as a measure of stimulus generalization for picture-naming therapies

Definition naming offers a well-controlled opportunity to demonstrate stimulus generalization by providing the opportunity to name each item trained with picture naming in a different, arguably more functionally-relevant, modality. There are shared and distinct task demands for picture naming and definition naming. The shared demands are retrieval of lexical-semantic and phonological features of the items to be named and production of the motor response. Task differences are related to the stimulus input modality. Picture naming relies on a visual stimulus that is usually present during the entire lexical retrieval process. By contrast, definition naming relies on auditory comprehension of a fading auditory stimulus, whereby the rate of processing is forced on the listener. For this reason, definition naming may rely more heavily on cognitive processes, such as verbal working memory, than would picture naming.

Definition naming is not a functional skill, per se, but it could be argued that it bridges a gap between skills required for picture naming and lexical retrieval in conversation. Both definition naming and lexical retrieval in conversation likely require a greater processing load than picture naming due to management of multiple demands, such as comprehension of a fading verbal stimulus while simultaneously searching for a word. Moreover, definition naming and conversation both require the speaker to take a verbal turn as a response to auditory input from another participant (Sacks, Schegloff, & Jefferson, 1974). Hence, the cognitive demands during definition naming may more closely resemble those required for lexical retrieval in conversation than the demands required for picture naming. Nevertheless, there will likely still be a gap between definition naming and production of these items in discourse during conversation due to the need to generalize from smaller linguistic units to larger units, the contribution of inter-personal factors (Carragher, Conroy, Sage, & Wilkinson, 2012), and lack of inherent scaffolding that is encountered in definition-naming (and picture-naming) tasks (Boo & Rose, 2011). As a first step in investigating these processes, however, definition naming may help control for these additional demands posed by conversational discourse tasks by better isolating the output processes that were trained in therapy (i.e. single-word lexical retrieval and production) with a novel input stimulus in a more conversationally-functional modality (i.e. auditory).

1.3. Working memory and aphasia

Working memory has been defined as a multi-component system consisting of short-term storage components (i.e. short-term memory) and attentional processes, known as the central executive, which manipulates stored information (Cowan, 2008). These components allow information, such as partial results, to be temporarily maintained by the storage components while being monitored and "worked on" by the central executive to plan and carry out behavior (Cowan, 2008). It has been well established that individuals with aphasia tend to have difficulty with some cognitive tasks involving verbal working memory (Burgio & Basso, 1997; Potagas, Kasselimis, & Evdokimidis, 2011), nonverbal working memory (Burgio & Basso, 1997; Lang & Quitz, 2012; Mayer & Murray, 2012; Potagas et al., 2011; Seniow, Litwin, & Lesniak, 2009), and attention and executive functioning skills (Caspari, Parkinson, LaPointe, & Katz, 1998; Glosser & Goodglass, 1990; Hula & McNeil, 2008; Korda & Douglas, 1997). Impairments in short-term memory and working memory in individuals with aphasia have been well-investigated in the literature (Martin & Ayala, 2004; Mayer & Murray, 2012; Salis, Kelley, & Code, 2015; Wright & Fergadiotis, 2012) using a variety of assessment procedures, such as span tasks and n-back tasks (See Salis et al., 2015 for a review.)

As briefly mentioned above, there is emerging evidence showing that working memory may be related to the ability to respond to aphasia therapy (Harnish & Lundine, 2015; Lambon Ralph, Snell, Fillingham, Conroy, & Sage, 2010; Seniow et al., 2009). Seniow et al. (2009) found that nonverbal working memory abilities, as measured by the Benton Visual Retention Test (Sivan, 1992) predicted response to an anomia therapy in the acute stages of recovery. Similarly, Harnish and Lundine (2015) and Lambon Ralph et al. (2010) reported relationships between memory abilities and anomia therapy gains. Specifically, Harnish and Lundine found that nonverbal working memory, as measured by the backward version of the Spatial Span subtest of the Wechsler Memory Scale (Wechsler, 1997), was a significant predictor of anomia treatment effect size in eight individuals with chronic aphasia. Lambon-Ralph et al. assessed 33 people with chronic aphasia on a variety of language and cognitive assessments and found that the best predictor of anomia therapy gains was a language factor as measured by the Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 2001) and a cognitive factor comprising scores on assessments of reasoning and problem-solving, including the Test of Everyday Attention (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994), Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtis, 1993), Rey Complex Figure immediate and delayed copy (Meyers & Meyers, 1995), and Pyramids and Palm Trees (Howard & Patterson, 1992). These studies show that there is some evidence in the literature for the relationship between memory and other cognitive skills in predicting response to anomia therapy and may be worth further investigating and perhaps targeting as a key component of rehabilitation.

Whereas previous literature has shown that neuropsychological testing of distinct cognitive domains holds promise as a prognostic indicator of aphasia therapy success (Harnish & Lundine, 2015; Lambon Ralph et al., 2010; van de Sandt-Koenderman et al., 2008), it is unclear the degree to which cognitive assessments may also predict generalization abilities. The present study aims to assess the impact of working memory skills on stimulus generalization by investigating the relationship between measures of verbal and nonverbal working memory and generalization from a picture naming treatment to definition naming.

2. Method

2.1. Participants

This study was approved by the local institutional review board prior to participant enrollment. All individuals ($n = 7$) experienced a left hemisphere middle cerebral artery ischemic stroke and were at least six months post-onset. Table 1 contains descriptions of participant lesions as obtained from review of medical records. They had a diagnosis of aphasia as determined by a Western Aphasia Battery (Kertesz, 2007) Aphasia Quotient of less than 94.7, but auditory comprehension was within two standard deviations of the mean for the Auditory Verbal Comprehension Score. This criterion was used to ensure that participants' auditory

Table 1
Participant lesion characteristics.

Participant	Lesion Characteristics
P1	(CT) .8 × .5 cm right frontal parietal white matter (age indeterminate); left frontal, occipital, and parietal lobes, left caudate nucleus
P2	(CT) Left temporal and parietal lobe and temporal operculum; left insula grey and white matter
P3	(MRI) Left frontal lobe cortex and subcortical white matter, insular cortex, and lateral margin of left basal ganglia; remote infarcts on left frontal and parietal watershed, centrum semiovale and corona radiata
P4	(CT) Left frontal lobe near Sylvian fissure
P5	(MRI) Cortical and subcortical left frontal and temporal lobes, small extension into left parietal lobe
P6	(MRI) Left frontotemporal, parietal, and posterior temporal lobes
P7	(MRI) Posterior portion of left middle cerebral artery territory; subcortical left temporal, frontal, and parietal lobe high signal changes

comprehension skills were sufficient to understand the instructions for all tasks. Participants experienced anomia, as evidenced by a score between 4 and 44 on the Boston Naming Test (Kaplan et al., 2001). They were all pre-morbidly right-handed, as determined by the Edinburgh Handedness Inventory (Oldfield, 1971) and were native speakers of English. Individuals participated in a screen for apraxia of speech that included tasks from the Apraxia Battery for Adults II (Dabul, 2000) (diadochokinetic rate, repetition of words of increasing length, repetition of multisyllabic words), a description of the Cookie Theft picture (Goodglass, Kaplan, & Barresi, 2001), and automatic speech tasks (i.e., counting from 1 to 10 and reciting the days of the week). Two certified speech-language pathologists with experience in adult neurogenic disorders, but no affiliation with this study, listened to recordings of the apraxia screen. These raters judged participant performance for the existence and severity of apraxia of speech using a Likert scale. Individuals with severe-to-profound apraxia of speech were excluded. If there was disagreement as to whether a participant had severe-to-profound apraxia of speech, a third certified speech-language pathologist judged the severity of apraxia of speech. Additional exclusion criteria were severe depression (Beck, Steer, & Brown, 1996), uncorrected vision or hearing problems, or suspected diffuse injury or disease of the brain.

Twenty-one individuals were recruited to participate in the study. Seven individuals met the inclusion criteria and completed all study activities. Demographic data are presented in Table 2.

2.2. Assessment

Participants' language and cognitive abilities were assessed prior to beginning therapy and are presented in Tables 2 and 3.

2.2.1. Working memory and language assessments

Working memory abilities have been investigated using the spatial span and the digit span, which are nonverbal and verbal span tasks, respectively. These tasks can be administered in the forward condition, where participants repeat numerical digits or touch a series of blocks in the same order as the administrator, or in the backward condition, where participants repeat digits or touch blocks in the reverse order as the administrator. From a theoretical standpoint, it would make sense that backward span tasks, both verbal and nonverbal, would rely more heavily on working memory than do forward span tasks, due to the mental manipulation involved in reversing the order of stimuli. Indeed, there is evidence that in healthy individuals, the backward condition of the digit span is more difficult than a forward digit span (Kessels, van den Berg, Ruis, & Brands, 2008), potentially indicating that the two tasks place different demands on the memory system; the forward digit span recruiting verbal working memory, such as Baddeley and Hitch's (1974) proposed phonological loop, and the backward digit span recruiting verbal working memory plus additional attentional oversight, such as the central executive. However, data on spatial span performance in healthy individuals show similar performance

Table 2
Participant demographic information and language assessment scores.

Participant	P1	P2*	P3	P4*	P5*	P6	P7*
Age (years)	72	**36**	62	**40**	**67**	78	**70**
Years of Education	20	**17**	12	**14**	**14**	12	**13**
Months Post Stroke	178	**33**	14	**12**	**21**	74	**10**
BNT[1], pre	14	**33**	5	**9**	**38**	10	**6**
WAB-AQ[2]	60	**71.8**	57.4	**63.8**	**76.5**	39	**48.1**
Classification[2]	Broca's	**Conduction**	Broca's	**Anomic**	**Conduction**	Broca's	**Broca's**
AV	8.2	**8.3**	7.1	**7.7**	**8.45**	6.1	**7.65**
Spontaneous Speech	9	**13**	9	**11**	**15**	6	**9**
Repetition	5.7	**5.6**	7.8	**7.2**	**5.8**	3.6	**2.5**
Naming/Word Finding	7.1	**7.5**	5.7	**6**	**9**	3.8	**4.9**

Note. * identifies participants who generalized from picture naming to naming to definition.

1. BNT = Boston Naming Test. Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test.* Philadelphia: Lea & Febiger.

2. WAB = Western Aphasia Battery, AQ = Aphasia Quotient. AV = Auditory Verbal Comprehension Score, Kertesz, A. (1982). *Western Aphasia Battery.* New York: Grune & Stratton.

Table 3
Participant scores on cognitive assessments and naming probes.

	Spatial Span	1-back	Picture Span	Listening Span	Baseline Picture Naming Average % correct (*SD*)	Baseline Definition Naming Average % correct (*SD*)	CPNT Acquired	Definition Naming Acquired	Generalization of CPNT Acquired to Definitions	Generalization of Items not Acquired during CPNT to Definitions	Tau-*U* Effect for CPNT
P1	0	−0.18	1	10	30 (8.66)	25 (14.14)	4/10 (40%)	1/12 (.08%)	0/10 (0%)	1/16 (.06%)	0.42 moderate
P2*	**7**	**0.68**	**6**	**13**	**46 (11.40)**	**12.5 (10.61)**	**10/10 (100%)**	**6/16 (38%)**	**4/10 (40%)**	**2/10 (20%)**	**0.85 very large**
P3	2	−0.05	1	8	18.33 (5.77)	5 (0)	0/15 (0%)	0/18 (0%)	0/15 (0%)	0/0 (0%)	−0.38 no effect
P4*	**5**	**0.75**	**2**	**12**	**38.33 (14.43)**	**37.5 (10.61)**	**6/10 (60%)**	**8/10 (80%)**	**8/10 (80%)**	**4/4 (100%)**	**1.00 very large**
P5*	**8**	**0.85**	**4**	**14**	**43.33 (7.64)**	**35 (35.36)**	**4/5 (80%)**	**6/7 (86%)**	**4/5 (80%)**	**0/1 (0%)**	**1.04 very large**
P6	0	0.05	0	0	11.25 (4.79)	0 (0)	10/17 (59%)	0/20 (0%)	0/17 (0%)	0/0 (0%)	0.84 very large
P7*	**2**	**0.83**	**2**	**4**	**13.75 (6.29)**	**20 (0)**	**5/10 (50%)**	**6/12 (50%)**	**4/10 (40%)**	**2/5 (40%)**	**1.00 very large**

* identifies participants who generalized from picture naming to definition naming.

1. Wechsler Memory Scale Spatial Span Backward subtest (Wechsler, 1997). Max score = 16.
2. The 1-back used symbols from Gauthier et al. (2003). Scores were calculate using a discrimination index, Pr = Hit rate (probability of selecting a target) – False alarm rate (probability of selecting a non-target) (Evans et al., 2011).
3. Picture Span (Dede et al., 2014) Max score = 14.
4. Listening Span (Tompkins et al., 1994) Max score = 42.
5. Tau-U effect size interpretation was based on Vannest and Ninci (2015). Effect sizes > 0.2 = Small, 0.2–0.6 = Moderate, 0.6–0.8 = Large, > 0.8 = Very Large.

between the forward and backward conditions (Kessels et al., 2008), indicating that perhaps both conditions use nonverbal working memory (e.g. visuospatial sketchpad, Baddeley & Hitch, 1974), without recruiting the central executive. One proposed difference between the backward condition of the digit and spatial span tasks to explain why the former may recruit additional attentional oversight is that the digit span relies on a transient auditory stimulus that is slowly decaying (Cowan, 2008), whereas the spatial span blocks are always present, and only the path between blocks needs to be recalled (Smyth & Scholey, 1992). Individuals with aphasia tend to perform better on forward conditions of the digit and spatial span than they do on backward conditions (Laures-Gore, Marshall, & Verner, 2011; Potagas et al., 2011; Ween, Varfaellie & Alexander, 1996), but it is still unclear the degree to which performance on forward and backward span tasks differentially relies on working memory.

In the present study, nonverbal working memory abilities were assessed using the backward condition of the spatial span subtest of the *Wechsler Memory Scales* (Wechsler, 1997) and a 1-back task. The backward condition of the spatial span was selected, as opposed to the forward condition, because our prior work showed a stronger relationship between performance on the backward spatial span and treatment outcomes in individuals with aphasia than did the forward condition (Harnish & Lundine, 2015). Administration and scoring of the spatial span occurred according to protocol. The administrator touched a series of blocks at a rate of 1 s per block. Participants were instructed to touch the blocks in the *reverse* order of the administrator. All correct trials received a score of one and incorrect trials were given a score of zero. Each level included two trials, beginning with level 2 (two blocks) and continuing until the participant made errors on both trials of a level.

The n-back is widely used to investigate nonverbal working memory (Wright & Fergadiotis, 2012), however, due to differences in response type, modality of stimuli (verbal/nonverbal) and inter-stimulus interval, it is difficult to determine reliability across different n-back tasks (Mayer & Murray, 2012). The 2-back task was shown to be much less stable than the 1-back in a study of younger and older neurologically intact adults (Dede, Ricca, Knilans, & Trubl, 2014). Therefore, we used a 1-back task, which was modeled on Christensen and Wright (2010). The picture stimuli were novel 2D shapes (Gauthier, James, Curby, & Tarr, 2003) presented in the clockwise picture plane. The task included one practice block with 10 images and two targets, followed by four blocks with 25 images and eight targets in each block (i.e. 100 total images and 32 total targets). The percentage of tokens was 32%, which is consistent with the n-back tasks reported in the literature, but also a reasonable length for persons with aphasia to complete (Christensen & Wright, 2010). Interstimulus interval was 2750 ms, and stimuli duration was 750 ms. Participants responded via button press on a Serial Response Box (Psychology software tools, model 200a). The 2nd button from the left (of 5 total buttons) was identified with a laminated green checkmark taped to the top of it. There were no other identifying markers on the box. Participants were instructed to rest a finger on the key with the checkmark and press it each time they saw a shape that matched the one presented immediately before it. Feedback for responses was provided during the practice block. 1-back performance was analyzed for accuracy using a signal detection statistic, Pr (Evans, Selinger, & Pollak, 2011), which took into consideration hits, misses, false alarms, and correct rejections.

Verbal working memory was assessed using a picture span task (Dede et al., 2014) and a listening span protocol (Tompkins, Bloise, Timko, & Baumgaertner, 1994). During the picture span, participants heard a sequence of one-syllable concrete nouns at a rate of one word per second. They were presented with a 3 × 3 grid of colored photographs. Participants were asked to point to the pictures of the spoken words in the *reverse* order. Dede et al. (2014) found that for the picture span, test-retest reliability was excellent and construct validity and internal consistency were acceptable in individuals with aphasia.

The listening span protocol (Tompkins et al., 1994) has been used in previous studies to assess auditory-verbal working memory in individuals with aphasia (Murray, Timberlake, & Eberle, 2007; Sung et al., 2009). Individuals were asked to remember the last word in each of a series of simple statements. At the end of each set, they were asked to recall these final words. Simultaneously, while listening to each sentence in the set, individuals were asked to indicate if the statements were true or false via a button press on the Serial Response Box (Psychology Software Tools, model 200a). The button for *true* responses (*2* on the response box) was marked with a green check, and the button for *false* responses (*4* on the response box) was marked with a red *X*. Participants were instructed to respond with only one finger and to rest their finger on the neutral button (*3* on the response box) in between trials. Validity and reliability were demonstrated in healthy adults and individuals with focal right hemisphere damage as well as individuals with left hemisphere damage with and without resulting aphasia (Lehman & Tompkins, 1998; Tompkins et al., 1994). Scores on working memory measures are included in Table 2.

The Western Aphasia Battery-Revised (WAB; Kertesz, 2007) and Boston Naming Test (Kaplan et al., 2001) were given to each participant as a measure of overall language abilities and naming abilities, respectively. Total scores and WAB subtest scores are included in Table 1.

2.2.1.1. Picture-naming probes. Picture-naming probes provided a measure of skill acquisition for naming pictures for trained items and were delivered by computer throughout the four phases of the study: baseline, treatment (prior to the treatment sessions each day), post-treatment (after all treatment sessions were completed), and at a three-month follow-up. Picture-naming probe items were delivered via Eprime 2.0 Professional (Psychology Tools, www.pstnet.com) on a 14 or 15.6 inch Dell Latitude E6540 laptop computer screen and were scored as correct or incorrect by the therapist. Responses that differed by one phoneme or more were counted as incorrect unless the variation was dialectical. Minor phoneme distortions were counted as correct if the target was still intelligible. Similarly, plural/singular variations (e.g. tacos for taco) were counted as correct if the change did not result in a different root word (e.g., checker for checkers) or impact the intelligibility of the target.

Responses were compared to a list of acceptable synonyms, and ongoing consensus meetings occurred to discuss potentially acceptable responses that were different from the target word. Participants were given 10 s to name each black and white picture while it remained on the computer screen. The clinician indicated on a keyboard and a written record form whether the participant gave a correct response, incorrect response, or no response. Self-corrections were scored as correct. That is, if the participant arrived at the correct production anytime within the 10-s time limit, then the trial was scored as correct. After the 10-s time limit elapsed, a white screen held the place in the program until the therapist advanced to the next picture. No cueing or feedback occurred during probing.

Three picture-naming probe lists were delivered. List 1 included items trained in therapy. List 2 included untrained items that were assessed with the same frequency as trained items from List 1 to control for a probing effect. Picture-naming probe Lists 1 and 2 were delivered at baseline, prior to treatment each day, at post-treatment, and at the three-month follow-up session. List 3 included untrained items that were only delivered once at baseline, post-treatment, and follow-up, and were included to assess generalization of trained to untrained items. Probe lists were matched according to living/nonliving items, word frequency, number of syllables, number of semantic categories, and non-overlapping semantic categories.

Participants achieved a stable baseline on List 1 trained items, as determined by no ascending trends for the last three probe sessions prior to beginning therapy. Three baseline sessions were completed for each participant, with additional baselines occurring in the event of an ascending trend. Inter-rater and intra-rater reliability of response scoring were calculated for 20% of picture-naming probe sessions. Inter-rater reliability was calculated at 96.4% and intra-rater reliability was calculated at 97.8%.

2.2.1.2. Definition-naming probes. Definition-naming probes provided a measure of generalization from visual picture naming to auditory definition naming. Definition naming was assessed for all items trained (List 1) and untrained (List 2 and List 3) in therapy. The same words were targeted for untrained definition-naming probes and untrained picture-naming probes. Definition probes for List 1 and List 2 were delivered twice at baseline, once at post-treatment, and once at follow-up. Definition probes for List 3 were delivered only once at baseline, once at post-treatment, and once at follow-up. Definitions did not include the target word, the root of the target word, or a derivative of the target word, nor did they include any words used in semantic cues provided during cued picture-naming treatment (CPNT). When there were naturally overlapping words for semantic cues and definitions, either semantic cues were changed to include alternate words or synonyms, or that particular cue was discarded and replaced by another cue without overlapping words. Definition-naming probes were piloted on 10 healthy control participants to ensure that the definitions elicited the desired name for the target item with at least 90% accuracy. Definitions were presented in the auditory modality alone by the same female speaker from a Dell Latitude E6540 laptop computer. Participants heard the definition twice with a 5-s delay between presentations and were given 20 s to name the definition after the second presentation.[1] The clinician indicated whether the

---

[1] The exception to this was for the first two participants (P1 and P2). They heard the definition once, then were asked to indicate if they wanted the definition repeated or not. If they decided they wanted the definition repeated, then it was played again and they were given 20 s to name the definition. If they did not request a repetition of the stimulus, they were given 20 s after the initial presentation to name the definition. The research team altered the protocol after the first two participants to control for the total amount of time available for future participants to name definitions. For all remaining participants, definitions were provided twice with 5 s between presentations.

participant gave no response, an incorrect response, or a correct response by pressing a key on the keyboard. All responses were also recorded on a written record form. Self-corrections were scored as correct. Inter-rater and intra-rater reliability of response scoring were calculated for 20% of definition-naming probe sessions. Inter-rater reliability was calculated at 95% and intra-rater reliability was calculated at 99.2%.

2.3. Therapy

Prior to beginning therapy, all participants named a corpus of 575 pictures on two separate occasions. Participants were given 10 s to name each picture. Self-corrections within the time limit were counted as correct. Therapy stimuli and stimuli for untrained probes and infrequently-administered probes were selected from items that were named incorrectly for both presentations.

Cued picture-naming treatment (CPNT) (Harnish & Lundine, 2015; Harnish et al., 2014) adapted from Kendall, Raymer, Rose, Gilbert, and Rothi (2014), was delivered on four days per week for two weeks. Participants attempted to name eight consecutive presentations of the same black and white picture, with cueing from the administrator. Trials included (1) independent naming, (2) orthographic cueing (i.e. the written word was beneath the picture), (3) repeating, (4) naming after a short delay (i.e. approximately 3 s), (5) semantic cueing (i.e. three cues providing semantic information about the target were spoken to the participant by the clinician), (6) phonological cueing (i.e. the first sound and letter were spoken by the clinician), (7) repeating, and (8) naming after short delay. The administrator provided the correct response for each incorrect participant response and asked the participant to repeat it. Dosage remained consistent across participants. Once eight presentations of each of the twenty pictures were completed, the therapy session was terminated regardless of session duration. Sessions were video- and audio-recorded. An individual who was unaffiliated with the treatment sessions watched 12.5% of sessions (i.e. one randomly selected treatment session per participant) and indicated via a record form checklist whether the key elements of the treatment were administered by the clinician for each trial. Treatment fidelity was 94%.

2.4. Analysis

Items that were incorrectly named at baseline (i.e. twice to definition and on a majority of attempts to picture) were compared with correctly-named definition probes delivered post-treatment to determine which items generalized from picture-naming therapy (CPNT) to definition naming. A generalization percentage was calculated for each participant for items generalized to definition naming out of total possible items that could have been generalized. The same process was used to determine generalization for untrained items and infrequently probed items (See Tables 3 and 4). Spearman's rank-order correlations were used to assess the relationship between measures of working memory and generalization percentage. Scatterplots for working memory measures and percent generalization are presented in Fig. 1.

In order to determine if *acquisition* of picture names in CPNT was required for generalization, we also compared within participant, on an item-by-item basis, whether pictures trained during CPNT but not acquired (i.e. incorrectly named to picture post-therapy), may have generalized to definition naming. These data are presented in Table 3.

Scores on picture-naming probes were plotted for each participant at baseline, during treatment, and post-treatment (Fig. 2). Individual and group effect sizes for CPNT (Table 3) were calculated using Tau-U (Parker, Vannest, Davis, & Sauber, 2011; Vannest, Parker, Gonen, & Adiguzel, 2016), a nonoverlap technique that controls for baseline trend and is robust enough for small datasets (Vannest & Ninci, 2015). Interpretation of Tau-U, based on general benchmarks outlined in Vannest and Ninci (2015) was a small effect for improvement of .20, a moderate effect for .20-.60, a large effect for .60-.80, and a very large effect for above .80.

3. Results

Spearman correlations were conducted to investigate the relationship between working memory and stimulus generalization.

Table 4
Participant generalization ratios and percentages.

Subject	List 1[b]	List 2[b]	List 3[b]
P1	0/10 (0%)	3/11 (27%)	2/4 (50%)
P2[a]	**4/10 (40%)**	**1/10 (10%)**	**1/11 (9%)**
P3	0/15 (0%)	0/20 (0%)	0/20 (0%)
P4[a]	**8/10 (80%)**	**4/14 (29%)**	**1/14 (7%)**
P5[a]	**4/5 (80%)**	**3/7 (43%)**	**3/14 (21%)**
P6	0/17 (0%)	1/16 (6%)	0/16 (0%)
P7[a]	**4/10 (40%)**	**2/15 (13%)**	**0/17 (0%)**

Note. List 1 are CPNT acquired/trained items. List 2 are untrained items that were probed with the same frequency as List 1. List 3 are untrained items that were only probed at baseline and post-treatment. Lists 2 and 3 were included as control measures.

[a] identifies participants who generalized from picture naming to definition naming.

[b] Denominator is total number of incorrectly produced at baseline, defined as incorrectly named definition twice at baseline and incorrectly named picture on majority of baseline attempts. Numerator is number of items generalized from picture naming to definition naming, defined as incorrectly produced at baseline and correctly named to definition at post-treatment.

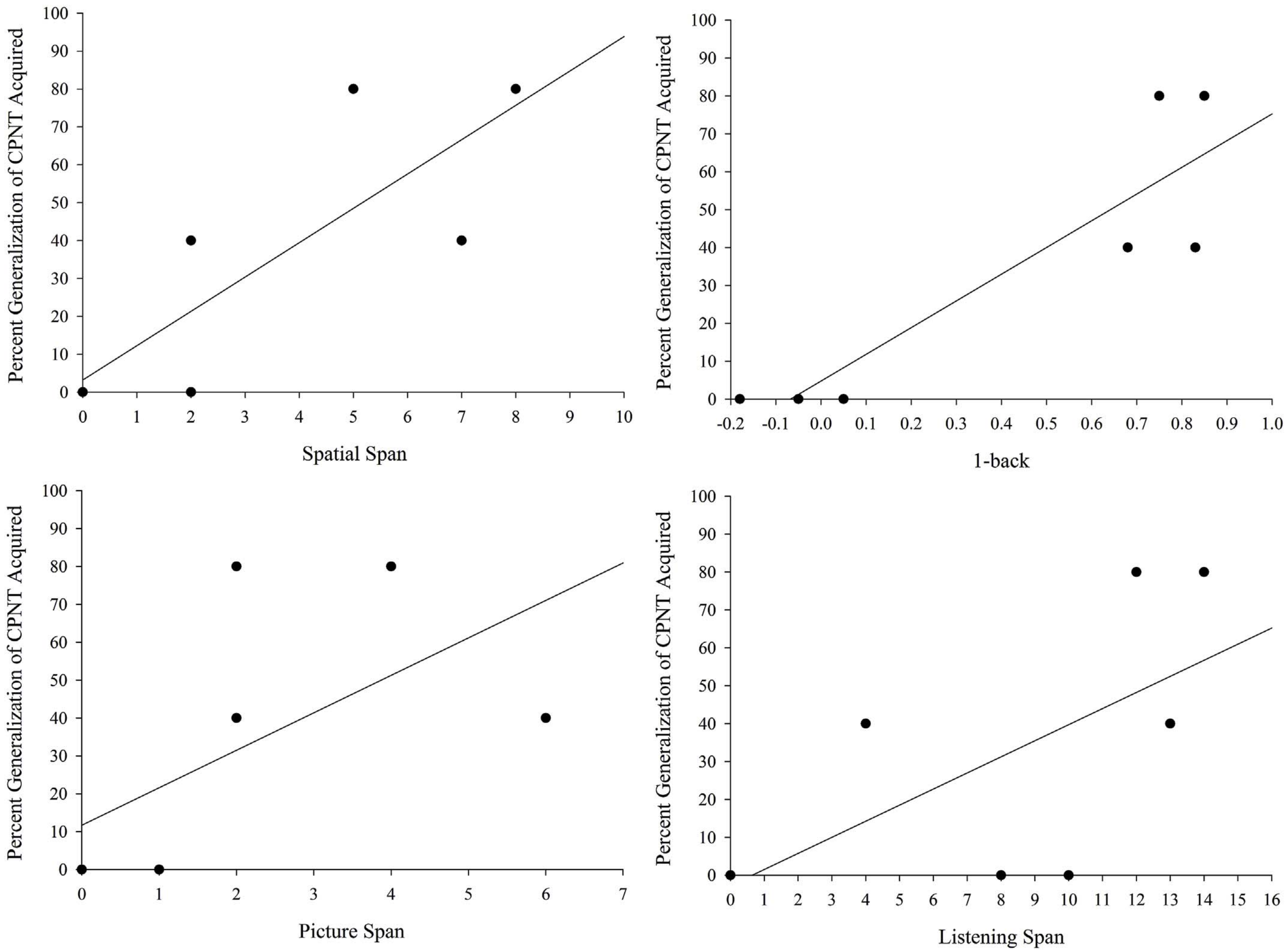
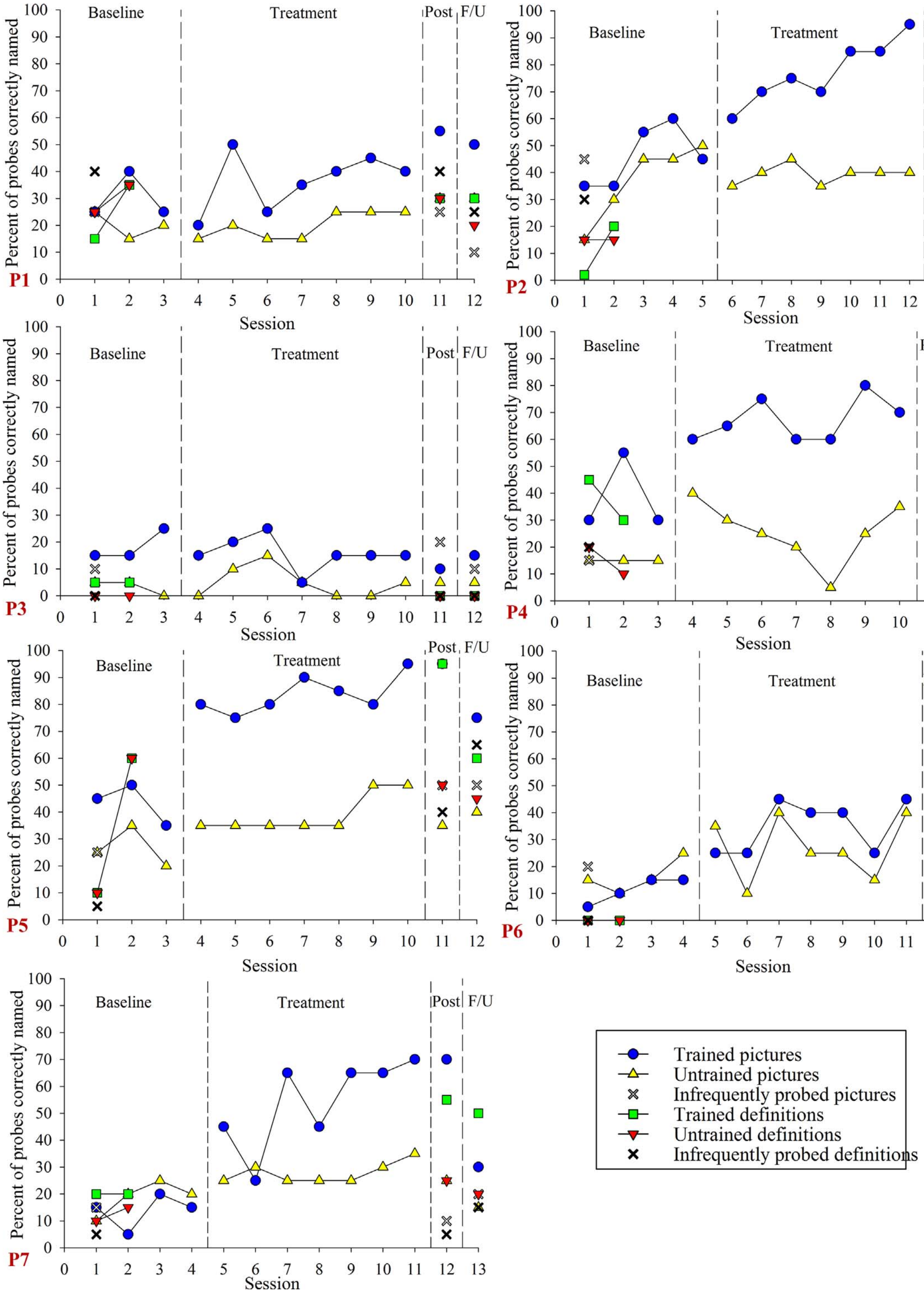


Fig. 1. Working Memory Measures and Generalization Percentage.

Scores for the working memory measures and generalization percentage are graphed for each participant to illustrate the relationship between working memory measures and the ability to generalize from a picture-naming therapy to definition naming.

Results showed that generalization percentage was correlated with the two nonverbal working memory measures, spatial span [$r_s(5) = .837$, $p = .019$] and 1-back [$r_s(5) = .869$, $p = .011$], and one of the verbal working memory measures, picture span [$r_s(5) = .770$, $p = .043$], were significantly correlated with generalization percentage at the $p < .05$ level (Table 5). The listening span was also moderately correlated with generalization percentage [$r_s(5) = .680$, $p = .093$], but was not statistically significant. To visualize these associations, scatterplots demonstrating the relationship between each working memory measure and generalization



(caption on next page)

Fig. 2. Probe Data for Picture Naming and Definition Naming for each Participant.

Probes for picture naming and definition naming are presented across baseline, treatment, post-treatment and follow-up phases for P1. Infrequently probed pictures and definitions were untrained. P1 overlapping data points include session 1 trained pictures, untrained pictures, infrequently probed pictures, and untrained definitions; session 2 trained definitions and untrained definitions; session 11 untrained pictures, trained definitions and untrained definitions; session 12 untrained pictures and trained definitions. P2 overlapping data points include session 1 untrained pictures and untrained definitions; session 13 untrained definitions and infrequently probed definitions; session 14 untrained pictures and trained definitions. P3 overlapping data points include session 1 untrained definitions and infrequently probed definitions; session 1 untrained pictures and trained definitions; session 2 untrained pictures and trained definitions; session 11 trained definitions, untrained definitions and infrequently probed definitions; session 12 trained definitions, untrained definitions and infrequently probed definitions. P4 overlapping data points include session 1 untrained definitions and infrequently probed definitions; session 1 untrained pictures and infrequently probed pictures; session 11 infrequently probed pictures and infrequently probed definitions; session 12 infrequently probed pictures and infrequently probed definitions. P5 overlapping data points include session 1 trained definitions and untrained definitions; session 1 infrequently probed pictures and untrained pictures; session 2 trained definitions and untrained definitions; session 11 trained pictures and trained definitions; session 11 untrained definitions and infrequently probed pictures. P6 overlapping data points include session 1 trained definitions, untrained definitions, and infrequently probed definitions; session 2 trained definitions and untrained definitions; session 2 trained pictures and untrained pictures; session 3 trained pictures and untrained pictures; session 12 trained definitions and infrequently probed definitions; session 13 trained pictures and untrained definitions. P7 overlapping data points include session 1 trained pictures and infrequently probed pictures; session 1 untrained pictures and untrained definitions; session 2 untrained pictures and trained definitions; session 12 untrained pictures and untrained definitions; session 13 infrequently probed pictures and untrained definitions; session 13 untrained pictures and infrequently probed definitions.

Table 5

Group correlations for generalization percentage and working memory measures.

	Percent generalization	Spatial Span	1-back	Picture Span	Listening Span
Percent generalization	1				
Spatial Span	.837*	1			
	.019				
1-back	.869*	.727	1		
	.011	.064			
Picture Span	.770*	.898*	.655	1	
	.043	.006	.111		
Listening Span	.680	.837*	.393	.800*	1
	.093	.019	.383	.031	

*Significant at $p < 0.05$.

Significance values are below correlations.

percentage are presented in Fig. 1. Although limited by the sample size, scatterplots revealed primarily positive associations between each working memory measure and generalization percentage, as indicated by the positively sloped trend lines.

Spearman correlations were also conducted to examine the relationship between working memory and treatment gains during CPNT. Results showed moderate but not statistically significant correlations, at the .05 significance level for any of the working memory measures [spatial span r_s (5) = .727, p = .064; 1-back r_s = (5) = .607, p = .148; picture span r_s (5) = .746, p = .054; listening span r_s (5) = .643, p = .119].

Comparison of scores for generalization of CPNT-acquired words and words not acquired during CPNT (Table 3) showed that three participants (P2, P4, P7) who generalized to definition naming for CPNT-acquired words, also generalized some words that were trained in CPNT, but not acquired. Conversely, three participants (P1, P3, P6) who did not generalize for CPNT-acquired, showed very little to no generalization for items not acquired during therapy.

Comparison of patterns of generalization (Table 4) for items acquired in CPNT (list 1), untrained, but probed with the same frequency as trained items (list 2), and untrained and infrequently-probed items (list 3) shows that the individuals who generalized for CPNT acquired items also showed some improved definition naming for items that weren't trained in CPNT, albeit to a lesser extent. Individuals who did not generalize to CPNT-acquired items demonstrated varied performance patterns. P1 named no CPNT-acquired items to definition, but did name some untrained items. P3 and P6 named no CPNT-acquired items, and performed similarly on untrained items with the exception of P6 naming 1/16 (6%).

Post hoc Spearman correlations were conducted in order to investigate if auditory comprehension deficits, aphasia severity, or overall naming impairment were related to definition naming abilities in our sample. Results showed some, but not statistically significant, correlations between percent generalization and the auditory verbal comprehension subtest of the WAB [r_s (5) = .61, p = .150], aphasia severity as measured by the WAB AQ [r_s (5) = .68, p = .093], and naming abilities as measured by the BNT [r_s (5) = .34, p = .455).

We also investigated the relationship between aphasia severity and working memory abilities. Results showed that overall aphasia severity, as measured by the WAB AQ was significantly correlated with the spatial span [r_s (5) = .837, p = .019], picture span [r_s (5) = .800, p = .031] and listening span [r_s (5) = 1.0, p = .000]. Naming severity, as measured by the BNT, was not significantly correlated with any of the working memory measures [spatial span r_s (5) = .436, p = .328; 1-back r_s = (5) = .250, p = .589; picture span r_s (5) = .491, p = .263; listening span r_s (5) = .679, p = .094)].

4. Discussion

The present study aimed to determine the relationship between working memory abilities and generalization from picture-naming therapy to definition naming, a skill that may begin to bridge the gap from picture naming to word retrieval in conversation. Results indicated that the two nonverbal measures, spatial span and 1-back, and one of the verbal measures, picture span, were each positively correlated with generalization percentage. The only measure that did not significantly correlate with generalization percentage was the verbal listening span task.

These results suggest that stimulus generalization, or the ability to use a trained item in another context, may be related to working memory skills. Patients who scored higher on the spatial span, 1-back, and picture-span also tended to have better generalization from a picture-naming therapy to definition naming.

4.1. "Non-generalizers"

Of the three participants who did not generalize to definition naming (P1, P3, and P6), two acquired some of the picture names during therapy (P1 and P6). P1 acquired 40% (4/10) of words during therapy, which indicated a moderate Tau-U effect size. P1 also demonstrated some ability to name definitions, as illustrated by baseline definition naming ability at 25% for items to be trained in CPNT and accurate definition naming for some untrained probe items for probe lists 2 and 3. However, none of the items acquired in CPNT generalized to definitions.

P1 was able to name 25% of definitions at baseline, but did not generalize CPNT-acquired words to definition naming. P6 demonstrated a very large effect size for picture naming, acquiring 59% (10/17) of picture names, but was unable to name to definition at baseline and after therapy. P6 also scored at floor for spatial span, picture span and listening span, and scored relatively low for the auditory verbal comprehension subtest of the WAB (6.1 out of 10). We suspect the combination of working memory and auditory comprehension deficits contributed to the inability to name to definition in this individual.

P3 neither responded to the picture naming therapy, nor generalized to definition naming. Overall, the individuals who did not generalize performed poorly on the 1-back, spatial span and picture span. However, responsiveness to therapy was varied in individuals who did not generalize. Two individuals showed at least modest gains after CPNT, whereas one person did not respond to therapy.

4.2. "Generalizers"

Individuals who generalized CPNT gains to definition naming (P2, P4, P5, and P7) showed variable patterns of performance on the cognitive measures. When compared with the non-generalizers, all of the generalizers performed better on the 1-back. P2, P4, and P5 also performed better than the non-generalizers on the spatial span. Thus, scores on the nonverbal working memory measures were higher for generalizers than for non-generalizers, with the exception of the spatial span score for P7. Scores on the picture span, a verbal working memory measure, were also higher for all generalizers than for non-generalizers. This pattern was not seen for the listening span. For example, P7, an individual who did generalize, scored very low on the listening span compared to the three others who generalized (P2, P4, and P5). In addition, P1, who demonstrated little effect of therapy and did not generalize, scored just below three of the generalizers. Thus, the listening span was the only measure that did not show a relationship with generalization in this sample.

Functional neuroanatomy provides a theoretical basis linking language and working memory. Brain regions that support naming and other language processes have also been implicated in working memory. There is recent evidence that damage to areas that support semantic storage, such as the lateral temporal lobes and the temporoparietal cortex (Binder, Desai, Graves, & Conant, 2009), results in reduced verbal working memory performance (Bormann, Seyboth, Umarova, & Weiller, 2015). Additional evidence comes from research showing no differences in laterality for some language tasks in younger and older adults (Obler, Woodward, & Albert, 1984), even though older adults who show better confrontation naming tend to use right hemisphere regions contralateral to left hemisphere language areas (Obler et al., 2010). This may suggest that language does not shift with aging, rather, older adults with preserved confrontation naming may call upon additional cognitive processes housed in the right hemisphere that provide supplementary contributions to core language areas. Previous work showing a relationship between working memory scores and gains on confrontation naming therapy in individuals with aphasia (Harnish & Lundine, 2015; Lambon Ralph et al., 2010; Seniow et al., 2009) provide further support for this relationship. Moreover, subcortical structures, such as the basal ganglia, have been implicated in lexical retrieval by assisting with selection of a word among competitors (Crosson, McGregor et al., 2007), as well as in working memory tasks, by controlling when prefrontal cortex representations should be maintained or updated (O'Reilly, 2006). Thus, brain regions that support language, and when damaged cause or impact aphasia, may also underlie working memory performance.

We consider several possible explanations for the finding that generalizers tended to outperform non-generalizers on three of the working memory measures. First, as reviewed above, there is a body of work showing that verbal and nonverbal working memory are related to aphasia therapy gains (Harnish & Lundine, 2015; Lambon Ralph et al., 2010; Seniow et al., 2009). It is possible that working memory skills are also a necessary foundation for using items acquired in therapy in other contexts. Working memory, by definition, includes mental manipulation of information. In accordance with Baddeley and Hitch's model (1974), as phonological information about how a word sounds is held in verbal working memory, and visuospatial information about how a picture looks is held in nonverbal working memory, perhaps the central executive system and the episodic buffer work together to integrate, store, and apply this knowledge to novel stimuli, such as spoken definitions. The episodic buffer accounts for cross-domain associations, such as those between names and faces (Cowan, 2008). We propose that it may also chunk and store verbal and nonverbal information about newly acquired words, while the central executive system may provide attentional oversight for using the integrated information in novel contexts.

An alternative explanation is that for some participants, definition naming difficulties occurred due to task demands and not lack of generalization, per se. Individuals who had verbal working memory deficits may have also had difficulty maintaining a definition in working memory long enough to access the label. If working memory resources were compromised, then definition naming may have overtaxed an already impaired system because of task-specific working memory requirements associated with definition naming. That is, there may not have been enough working memory resources to allocate to the task of definition naming that required maintaining a fading verbal stimulus in memory while searching for the word. Although we propose that working memory is a skill associated with stimulus generalization, better generalization may have occurred in contexts with a continuously present stimulus, so as not to overtax the working memory system. Nevertheless, working memory demands on processing verbal stimuli are also present in conversation, and thus, still capture relevant weaknesses.

Another consideration in the present investigation is whether definition-naming difficulties may be due to auditory comprehension deficits or overall aphasia severity and not working memory. Definition naming relies on auditory comprehension, whereas picture naming does not. Correlations between percent generalization and the auditory verbal comprehension subtest of the WAB, WAB AQ, and the BNT were small-to-moderate, but not statistically significant. In addition, inclusion criteria insured that individuals with severe auditory comprehension deficits were excluded. Therefore, auditory comprehension abilities alone may not explain poor generalization to definition naming; however, this should be verified in future studies.

Finally, are working memory abilities related to generalization specifically, or simply to treatment gains (i.e. acquisition) and by extension to generalization? All participants who generalized picture-naming therapy gains to definition naming demonstrated a very large effect size during CPNT. Potentially due to the small sample size, a positive trend was present in the correlations between CPNT gains and working memory measures, but they did not reach significance. One participant (P6) had a very large effect size, but was unable to name to definition at any point in the study. To investigate this further, we looked at whether items that were not acquired in therapy also showed improvement in definition naming. We considered items to not be acquired if at baseline they were named incorrectly to definition twice and on a majority of picture naming attempts, and named incorrectly to pictures at post-therapy. We then determined how many of them were named correctly to definition (Table 3). We found similar patterns in that generalizers (P2, P4, P7) tended to name items not acquired during CPNT to definition and non-generalizers (P1, P3, P6) tended not to name items not acquired during CPNT to definition. In other words, the act of definition naming improved for people with higher working memory scores, even for items they did not successfully acquire during picture naming; whereas people who scored lower on the working memory measures showed no increases in definition naming, even for items acquired during picture naming. These data are difficult to interpret because we generally think of generalization as occurring after acquisition; however, the exposure to pictured stimuli during CPNT may have been enough for generalization to definitions in some individuals, even if the picture name was incorrect. We propose that this may illustrate a modality or task (i.e. definition naming) that is more responsive to therapy than picture naming in some participants. The present findings that generalizers tended to generalize even for items they did not acquire during picture naming, and non-generalizers did not show improvements for definition naming, even for acquired pictures, lends support to the idea that working memory abilities are not simply related to generalization by extension of treatment gains.

This study also highlights the notion that although age, lesion size, and aphasia severity may indicate poor prognosis for some individuals, those patients may still benefit from therapy. For example, P6 and P7 were both older adults in their 70s, achieved low scores on the BNT and WAB, and had lesions with a relatively wide distribution. However, they both responded to the aphasia therapy, acquiring 59% and 50% of trained items, respectively.

4.3. Comparison of working memory measures

The nonverbal spatial span and the verbal picture span scores correlated more highly with each other than with scores on any other tasks. This leads us to make two tentative conclusions based on the small dataset in the present study. First, the verbal and nonverbal span tasks utilized in this study are either measuring the same construct or measuring two constructs that are driven by a shared processing mechanism. Although one is verbal and one is nonverbal, these tasks may be isolating recall memory – either domain general working memory or domain specific working memory with a central processing component, such as attentional oversight. At least two prominent models of working memory (Baddeley, 2000; Baddeley & Hitch, 1974; Cowan, 1988, 1997) include a central processing mechanism for verbal and nonverbal stimuli. In contrast to our results, Potagas et al. (2011) found that the nonverbal spatial span and the verbal digit span did not significantly correlate with each other in 58 individuals with aphasia; however, performance on the digit span may have been impacted by verbal production deficits. It is noteworthy that the authors implemented exclusion criteria, such as severe aphasia and the inability to repeat at least one two-syllable word, to limit the contribution of confounding speech output deficits,. However, it is possible that some individuals who could repeat a two-syllable word may have still had motor deficits that would affect the ease with which verbal responses were made, and thus, may have made a verbal output task more challenging than a task that did not require a verbal response. In the present study, task demands may have been more closely matched between the picture span and the spatial span than they are with the digit span because both included a visual component that allowed for nonverbal responses in individuals with aphasia.

Although the 1-back task was positively correlated with generalization percentage in our sample, it showed low correlation with other measures of working memory, which has also been found in other studies (Jaeggi et al., 2010; Kane, Conway, Miura, & Colflesh, 2007; Oberauer, 2005; Roberts & Gibson, 2002). N-back tasks are thought to have strong face validity, but construct validity is not well established (Wright & Fergadiotis, 2012). Since the 1-back measures speeded recognition with interference (necessitating recognition memory and inhibition of competing stimuli) as opposed to serial retrieval measured in other working memory span tasks, it may be assessing a different construct (Kane et al., 2007). As Dede et al. (2014) pointed out, recognition tasks employ external cues, such as pictures, whereas recall tasks rely on internal cues. It has been suggested that there may be different memory processes activated for remembering with interference (Kane et al., 2007). In addition, the n-back requires continuous updating of working memory to eliminate items that are no longer needed in a sequence. It requires a person to remember the previous item, but not the items presented before the previous item (i.e. competitors with the target that must be inhibited). Thus, the cognitive processes required for successful performance on the n-back may go beyond those required for other working memory tasks (Wright & Fergadiotis, 2012).

Skills required to complete the 1-back task may also be an important contribution to the overall cognitive skills required for using auditory definitions to retrieve and produce names of items that were acquired in a picture-naming therapy. This would explain the positive correlation between the 1-back and generalization percentage. The 1-back requires recognition of nonlinguistic figures at a rate that is forced on the user. The program runs and the participant only responds by button press to identify a match between a current stimulus and the stimulus immediately preceding it. It taps into working memory by requiring the participant to maintain a running log of items to compare a presented item with the item immediately preceding it. It also requires continuously removing items from the running log so that they do not interfere. Similarly, definition naming requires rapid processing of auditory information at a rate forced on the listener, and maintenance of phonological information (as opposed to visuospatial for the n-back) in working memory while searching for the label. Definition naming may also require matching a stimulus to previously encountered information (e.g., the target word) as occurs in the n-back task. Production of labels was practiced using picture naming, and thus, there may have been a matching component whereby the participant matched semantic features presented in the auditory modality to a previously rehearsed label. Based on these theoretical comparisons that highlight similarities in required processing for n-back completion and naming to definition, it is not surprising that individuals who did not show improvements in generalization also performed poorly on the n-back.

The listening span was the only working memory measure that did not significantly correlate with generalization percentage. Dede et al. (2014) found that for individuals with aphasia, the listening span and the 1-back were less sensitive to differences in working memory than other measures (e.g. picture span, 2-back). Unlike the other working memory measures, the listening span has a verbal production component. In the present study, participants who performed most poorly on the listening span (P3, P6, P7) also demonstrated lower BNT (Kaplan et al., 2001) scores (5, 10, 6, respectively). It is unclear the extent to which aphasia impacted performance on the listening span, and whether it was a reliable measure of working memory in our participants.

5. Summary and conclusions

It is well known that individuals with aphasia demonstrate verbal and nonverbal working memory deficits, and there is recent evidence that working memory may serve as a prognostic indicator of how well an individual with aphasia will respond to therapy. The current study found a relationship between generalization percentage and three measures of working memory, two nonverbal and one verbal. We suggest that definition naming is one step closer to lexical retrieval in conversation than picture naming due to the absence of a picture stimulus and the inherent requirement to rapidly process and respond to an incoming verbal stimulus. The relationship between working memory measures and generalization to definition naming provide additional support that working memory may be associated with degree of behavioral gains from aphasia therapy, including stimulus generalization to other contexts.

Researchers have begun to advocate for assessment of working memory and/or short-term memory in individuals with aphasia to improve predictions of therapy response and treatment planning (Lambon Ralph et al., 2010; Murray, 2012; Salis et al., 2015; van de Sandt-Koenderman et al., 2008). As generalization is the goal of anomia therapy, future studies should both investigate the impact of working memory and other cognitive abilities on the ability to generalize in larger samples of individuals with aphasia and investigate the influence of working memory training on improving generalization. The present investigation is one step in a continuum of research necessary to determine the cognitive predictors of generalization capacities in persons with aphasia. The addition of cognitive measures in anomia therapy studies may help to shed light on the causes of inadequate generalization during treatment and guide preliminary trials to target these causes with intervention.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We would like to acknowledge Jennifer Brello, Dr. Christy Ray, Nadine Whiteman, and Angie Dubis-Bohn who assisted with screening procedures for inclusion of participants.

References

Armstrong, E. (2000). Aphasic discourse: The story so far. *Aphasiology, 14*, 875–892.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Vol. Ed.), *The psychology of learning and motivation: 8*, (pp. 47–89). New York: Academic Press.

Beck, A. T., Steer, R. A., & Brown, G. (1996). *Manual for the beck depression inventory-II.* San Antonio, TX: Psychology Corporation.

Best, W., Grassly, J., Greenwood, A., Herbert, R., Hickin, J., & Howard, D. (2011). A controlled study of changes in conversation following aphasia therapy for anomia. *Disability & Rehabilitation, 33*(3), 229–242.

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.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*, 2767–2796.

Boo, M., & Rose, M. L. (2011). The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. *Aphasiology, 25*, 154–175.

Bormann, T., Seyboth, M., Umarova, R., & Weiller, C. (2015). "I know your name, but not your number"- Patients with verbal short-term memory deficits are impaired in learning sequences of digits. *Neuropsychologia, 72*, 80–86.

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-language Pathology, 13*, 236–249.

Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-language Pathology, 4*, 94–98.

Burgio, F., & Basso, A. (1997). Memory and aphasia. *Neuropsychologia, 35*(6), 759–766.

Carragher, M., Conroy, P., Sage, K., & Wilkinson, R. (2012). Can impairment-focused therapy change the everyday conversations of people with aphasia? A review of the literature and future directions. *Aphasiology, 26*(7), 895–916.

Caspari, I., Parkinson, S. R., LaPointe, L. L., & Katz, R. C. (1998). Working memory and aphasia. *Brain and Cognition, 37*(2), 205–223.

Cherney, L. R., Patterson, J. P., & Raymer, A. M. (2011). Intensity of aphasia therapy: Evidence and efficacy. *Current Neurology and Neuroscience Reports, 11*(6), 560–569.

Christensen, S. C., & Wright, H. H. (2010). Verbal and non-verbal working memory in aphasia: What three n-back tasks reveal. *Aphasiology, 24*(6–8), 752–762.

Conroy, P., Sage, K., & Lambon Ralph, M. (2009). Improved vocabulary production after naming therapy in aphasia: Can gains in picture naming generalize to connected speech? *International Journal of Language & Communication Disorders, 44*(6), 1036–1062.

Coppens, P., & Patterson, J. (2018). *Aphasia Rehabilitation: Clinical challenges.* Burlington, MA: Jones & Bartlett.

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.

Cowan, N. (1997). *Attention and memory.* Oxford, NY: Oxford University Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*, 323–388.

Crosson, B., McGregor, K., Gopinath, K., Conway, T., Benjamin, M., Chang, Y. L., … White, K. D. (2007). Functional MRI of language in aphasia: A review of the literature and the methodological challenges. *Neuropsychology Review, 17*(2), 157–177.

Dabul, B. L. (2000). *Apraxia Battery for adults* (2nd ed.). Austin, Tex: Pro-Ed.

Dede, G., Ricca, M., Knilans, J., & Trubl, B. (2014). Construct validity and reliability of working memory tasks for people with aphasia. *Aphasiology, 28*(6), 692–712.

Dignam, J. K., Rodriguez, A. D., & Copland, D. A. (2016). Evidence for intensive aphasia Therapy: Consideration of theories from neuroscience and cognitive psychology. *PM&R, 8*(3), 254–267.

Evans, J. L., Selinger, C., & Pollak, S. D. (2011). P300 as a measure of processing capacity in auditory and visual domains in Specific Language Impairment. *Brain Research, 1389*, 93–102.

Faroqi-Shah, Y., & Virion, C. R. (2009). Constraint-induced language therapy for agrammatism: Role of grammaticality constraints. *Aphasiology, 23*(7/8), 977–988.

Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology, 20*(3), 507–523.

Glosser, G., & Goodglass, H. (1990). Disorders in executive control functions among aphasic and other brain-damaged patients. *Journal of Clinical and Experimental Neuropsychology, 12*(4), 485–501.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston diagnostic aphasia examination* (3rd ed.). Philadelphia: Lippincott Williams & Wilkins.

Grande, M., Hussmann, K., Bay, E., Christoph, S., Piefke, M., Willmes, K., et al. (2008). Basic parameters of spontaneous speech as a sensitive method for measuring change during the course of aphasia. *International Journal of Language & Communication Disorders, 43*(4), 408–426.

Greenwood, A., Grassly, J., Hickin, J., & Best, W. (2010). Phonological and orthographic cueing therapy: A case of generalised improvement. *Aphasiology, 24*, 991–1016.

Harnish, S. M., & Lundine, J. P. (2015). Nonverbal working memory as a predictor of anomia treatment success. *American Journal of Speech-language Pathology, 24*(4), S880–S894.

Harnish, S. M., Morgan, J., Lundine, J. P., Bauer, A., Singletary, F., Benjamin, M. L., … Crosson, B. (2014). Dosing of a cued picture-naming treatment for anomia. *American Journal of Speech-language Pathology, 23*(2), 285–299.

Hashimoto, N., & Frome, A. (2011). The use of a modified semantic features analysis approach in aphasia. *Journal of Communication Disorders, 44*, 459–469.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtis, G. (1993). *Wisconsin card sorting test (WCST) Manual: Revised and expanded.* Odessa: Psychological Assessment Resources Inc.

Hickin, J., Mehta, B., & Dipper, L. (2015). To the sentence and beyond: A single case therapy report for mild aphasia. *Aphasiology, 29*(9), 1038–1061.

Howard, D., & Patterson, K. (1992). *Pyramids and Palm Trees: A test of semantic access from pictures and words.* Bury St. edmunds, England: Thames Valley Test Company.

Hula, W. D., & McNeil, M. R. (2008). Models of attention and dual-task performance as explanatory constructs in aphasia. *Seminars in Speech and Language, 29*(3), 169–187.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning: Implications for training and transfer. *Intelligence, 38*, 625–635.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 615–622.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston naming test.* Philadelphia, PA: Lea and Febiger.

Kendall, D., Raymer, A., Rose, M., Gilbert, J., & G. Rothi, L. J. (2014). Anomia treatment platform as behavioral engine for use in research on physiological adjuvants to neurorehabilitation. *Journal of Rehabilitation Research and Development, 51*(3), 391–400.

Kertesz, A. (2007). *Western aphasia Battery revised.* San Antonio, TX: Harcourt Assessment, Inc.

Kessels, R. P., van den Berg, E., Ruis, C., & Brands, A. M. (2008). The backward span of the corsi block-tapping task and its association with the WAIS-III digit span. *Assessment, 15*, 426–434.

Kiran, S., Sandberg, C., & Sebastian, R. (2011). Treatment of category generation and retrieval in aphasia: Effect of typicality of category items. *Journal of Speech, Language, and Hearing Research, 54*(4), 1101–1117.

Kiran, S., & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language, and Hearing Research, 46*(3), 608–622.

Kleim, J. A., & Jones, T. A. (2008). Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research, 51*(1), S225–S239.

Korda, R. J., & Douglas, J. M. (1997). Attention deficits in stroke patients with aphasia. *Journal of Clinical and Experimental Neuropsychology, 19*(4), 525–542.

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.

Lang, C. J., & Quitz, A. (2012). Verbal and nonverbal memory impairment in aphasia. *Journal of Neurology, 259*(8), 1655–1661.

Laures-Gore, J., Marshall, R. S., & Verner, E. (2011). Performance of individuals with left-hemisphere stroke and aphasia and individuals with right brain damage on forward and backward digit span tasks. *Aphasiology, 25*, 43–56.

Lehman, M. T., & Tompkins, C. A. (1998). Reliability and validity of an auditory working memory measure: Data from elderly and right-hemisphere damaged adults. *Aphasiology, 12*(7/8), 771–785.

Maher, L. M., Kendall, D. L., Swearengin, J. A., Rodriguez, A., Leon, S., Pingel, K., … Rothi, L. J. (2006). A pilot study of use-dependent learning in the context of Constraint Induced Language Therapy. *Journal of the International Neuropsychological Society, 12*(6), 843–852.

Martin, N., & Ayala, J. (2004). Measurements of auditory-verbal STM span in aphasia: Effects of item, task, and lexical impairment. *Brain and Language, 89*, 464–483.

Mayer, J. F., & Murray, L. L. (2012). Measuring working memory deficits in aphasia. *Journal of Communication Disorders, 45*(5), 325–339.

Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex figure test and recognition trial.* Odessa, FL: Psychological Assessment Resources, Inc.

Murray, L. L. (2012). Direct and indirect treatment approaches for addressing short-term or working memory deficits in aphasia. *Aphasiology, 26*(3–4), 317–337.

Murray, L., Timberlake, A., & Eberle, R. (2007). Treatment of underlying forms in a discourse context. *Aphasiology, 21*(2), 139–163.

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journl of Experimental Psychology, 134*(3), 368–387.

Obler, L. K., Rykhlevskaia, E., Schnyer, D., Clark-Cotton, M. R., Spiro, A., Hyun, J., ... Albert, M. L. (2010). Bilateral brain regions associated with naming in older adults. *Brain and Language, 113*(3), 113–123.

Obler, L. K., Woodward, S., & Albert, M. L. (1984). Changes in cerebral lateralization in aging. *Neuropsychologia, 22*, 235–240.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.

O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science, 314*, 91–94.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*, 284–299.

Potagas, C., Kasselimis, D., & Evdokimidis, I. (2011). Short-term and working memory impairments in aphasia. *Neuropsychologia, 49*(10), 2874–2878.

Rider, J. D., Wright, H. H., Marshall, R. C., & Page, J. L. (2008). Using semantic feature analysis to improve contextual discourse in adults with aphasia. *American Journal of Speech-language Pathology, 17*(2), 161–172.

Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research, 31*(6), 573–698.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The test of Everyday attention.* Bury St. Edmonds, England: Thames Valley Test Company.

Rose, M., Douglas, J., & Matyas, T. (2002). The comparative effectiveness of gesture and verbal treatments for a specific phonologic naming impairment. *Aphasiology, 16*, 1001–1030.

Rose, M., Mok, Z., Carragher, M., Katthagen, S., & Attard, M. (2016). Comparing multi-modality and constraint-induced treatment for aphasia: A preliminary investigation of generalisation to discourse. *Aphasiology, 30*(6), 678–698.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50*, 696–735.

Salis, C., Kelley, H., & Code, C. (2015). Assessment and treatment of short-term and working memory impairments in stroke aphasia: A practical tutorial. *International Journal of Language & Communication Disorders, 50*(6), 721–736.

van de Sandt-Koenderman, W. M., van Harskamp, F., Duivenvoorden, H. J., Remerie, S. C., van der Voort-Klees, Y. A., Wielaert, S. M., ... Visch-Brink, E. G. (2008). MAAS (Multi-axial aphasia system): Realistic goal setting in aphasia rehabilitation. *International Journal of Rehabilitation Research, 31*(4), 314–320.

Seniow, J., Litwin, M., & Lesniak, M. (2009). The relationship between non-linguistic cognitive deficits and language recovery in patients with aphasia. *Journal of the Neurological Sciences, 283*, 91–94.

Sivan, A. B. (1992). *Benton visual retention test* (5th ed.). San Antonio, TX: Psycological Corporation.

Smyth, M. M., & Scholey, K. A. (1992). Determining spatial span: The role of movement time and articulation rate. *Quarterly Journal of Experimental Psychology, 45*, 479–501.

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

Sung, J. E., McNeil, M. R., Pratt, S. R., Dickey, M. W., Hula, W. D., Szuminsky, N. J., et al. (2009). Verbal working memory and its relationship to sentence-level reading and listening comprehension in persons with aphasia. *Aphasiology, 23*(7–8), 1040–1052.

Tompkins, C. A., Bloise, C. G. R., Timko, M. L., & Baumgaertner, A. (1994). Working memory and inference revision in brain-damaged and normally aging adults. *Journal of Speech & Hearing Research, 37*, 896–912.

del Toro, C. M., Altmann, L. J. P., Raymer, A. M., Leon, S., Blonder, L. X., et al. (2008). Changes in aphasic discourse after contrasting treatments for anomia. *Aphasiology, 22*, 881–892.

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling and Development, 93*, 403–411.

Vannest, K. J., Parker, R. I., Gonen, O., & Adiguzel, T. (2016). *Single case research: Web based calculators for SCR analysis* (Version 2.0) [Web-based application] College Station, TX: Texas A&M University. Available from: singlecaseresearch.org.

Wambaugh, J. L., Mauszycki, S., & Wright, S. (2014). Semantic feature analysis: Application to confrontation naming of actions in aphasia. *Aphasiology, 28*(1), 1–24.

Wechsler, D. (1997). *WMS-III administration and scoring manual.* San Antonio, TX: The Psychological Corporation, Harcourt Brace Jovanovich.

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

Wright, H. H., & Fergadiotis, G. (2012). Conceptualizing and measuring working memory and its relationship to aphasia. *Aphasiology, 26*(3–4), 258–278.