HHS Public Access
Author manuscript
J Neurolinguistics. Author manuscript; available in PMC 2019 November 01.

Published in final edited form as:
J Neurolinguistics. 2018 November ; 48: 199–225. doi:10.1016/j.jneuroling.2018.02.006.

Assessment of linguistic and verbal short-term memory components of language abilities in aphasia

Nadine Martin, **Irene Minkina**, **Francine P. Kohen**, and **Michelene Kalinyak-Fliszar**
Department of Communication Sciences and Disorders, Temple University, Philadelphia, USA

Abstract

Some current models of aphasia emphasize a role of short-term memory in the processing of language and propose that the language impairment in aphasia involves impairment to cognitive processes that activate and maintain representations of words over the time-period needed to support single word and multiple word tasks, including verbal span tasks. This paper reports normative data from 39 people with aphasia and 16 age-matched neurotypical controls on a test battery for aphasia that assesses effects of increased short-term/working memory load on word and sentence processing as well as effects of linguistic variations on verbal short-term memory abilities Two concepts are discussed that capture the unique potential of this test battery for research and clinical practice: specificity of diagnosis and sensitivity to all degrees of aphasia severity, including mild aphasia. An analysis is included that shows how the performance of individuals with mild aphasia who achieve normal level of performance on the Western Aphasia Battery (Kertesz, 2006) show a decline in a temporal delay condition that is greater than performance of control participants. We also report preliminary data showing differential effects of adding a time interval before a response or between items to be compared: reduced accuracy for some individuals with aphasia and improved accuracy for others. The theoretical and clinical importance of this finding is discussed, as well as the overall potential for this test battery to be used in research and as a clinical tool. Finally, we discuss the relevance of this test battery to investigate functional communication abilities in aphasia.

Introduction

Approaches to assessment of language impairment in aphasia have evolved in accordance with changing views of the nature of aphasia and the level of description used to diagnose a language impairment. Neuroanatomical models motivated classification of aphasia impairments in terms of symptom complexes associated with the regions of neurological impairment, with the symptoms described at the 'task' level (e.g., naming, repetition or comprehension). Psycholinguistic models provided a more microscopic perspective with

Address Correspondence: Nadine Martin, Ph.D., Department of Communication Sciences and Disorders, Aphasia Rehabilitation Research Laboratory Eleanor M. Saffran Center for Cognitive Neuroscience, Temple University, 1701 N. 13th St. Philadelphia, PA 19122, nmartin@temple.edu.

their description of language impairments in aphasia in terms of the linguistic representations (e.g., semantics) and processes (access, retrieval) involved in carrying out language tasks. For example, in this type of model, a naming impairment could be attributed to poor access of word representations from semantics or poor phonological encoding of words. These models reflected an emerging view of aphasia as a disorder that affects processing of language representations (e.g., McNeill, 1982; McNeill & Pratt, 2001), not the loss of linguistic representations. As this characterization of the nature of aphasia has increased in prominence, research has focused more on identifying the component operations of language processing, including mechanisms of accessing and retrieving words (e.g., Dell & O'Seaghdha, 1992; Dell, Schwartz, Martin, Saffran & Gagnon, 1997) and the temporal aspects of processing (e.g., Martin & Gupta, 2004; Martin, Saffran & Dell, 1996; R. Martin, Shelton & Yaffee, 1994). In this paper, we introduce a test battery for aphasia, the Temple Assessment of Language and Short-term Memory in Aphasia (TALSA), which builds on this research, specifying processes that support access and retrieval of language representations and adding a unique focus on the short-term memory (STM)/temporal component of that processing.

Language processing involves timely access to and retrieval of language representations. These operations are supported by cognitive abilities such as STM, working memory (WM) and executive functions. The hypothesized role of executive processes is to 'control' access to and retrieval of language representations via fundamental functions such as attention and inhibition (Conway & Engle, 1994; Engle, 2002; Hula & McNeil, 2008; Kane & Engle, 2003; McNeil, Odell, & Tseng, 1991; Wright & Fergadiotis, 2012). In contrast, STM's role is thought to be more integral to the processing of word representations, serving to maintain activation of semantic, lexical and phonological representations over the time course of processing single and multiple word utterances (Martin & Saffran, 1992; 1997; Martin et al., 1996).

STM is related to WM and sometimes the two abilities are not sufficiently distinguished from each other. Cowan (2008) describes their relationship as follows: (1) STM is a mental ability that maintains a limited amount of information in a temporarily accessible state and (2) WM includes STM along with other cognitive mechanisms (e.g., attention) that "make use of short-term memory" (p. 325). We agree that verbal WM tasks are supported by a STM capacity (measured minimally by a forward digit or word span), but WM and STM demands also vary in the degree of "work" entailed based on the linguistic (e.g., abstract words are harder to recall than concrete words (e.g., Walker & Hulme, 1999) and/or attentional and executive requirements of the task.

Evidence shows that individuals with aphasia almost ubiquitously exhibit reduced verbal STM capacity as measured on verbal span tasks. How does STM support language processing and how is it implicated in aphasia? The interactive activation model of word processing (Dell & O' Seaghdha, 1992), which has been used to account for word production impairments in aphasia (e.g., Dell et al., 1997; Schwartz, Dell, Martin, Gahl & Sobel, 2006), holds that access to and retrieval of words depend on stable activation of those representations. That stability depends on two processing parameters, connection strength (strength of activation spread) and decay rate (how quickly activation dissipates). Both parameters contribute to the likelihood that a representation will be able to compete with other semantically and phonologically related word representations that are primed by spreading activation. Connection strength needs to be sufficiently strong for the target word's activation level to be greater than competing representations that are primed by spreading activation. At the same time, activation decay rate needs to be slow enough to ensure that the activation level will remain competitive relative to other words in the lexicon until the word is comprehended or retrieved for production or repetition. The latter function, sustaining a strong activation level of the target word that will be uttered, can be viewed as a form of verbal STM that supports access to and retrieval of words in single and multiple word processing tasks, including verbal span tasks used as measures of verbal STM capacity.

The Temple Assessment of Language and (Verbal) Short-term Memory in Aphasia (TALSA): Purpose and aims of this study

The TALSA test battery is designed to assess language and verbal STM abilities in aphasia. Information gained from the TALSA battery can be used to identify the following:

1. The linguistic characteristics (semantic, phonological) of language/STM impairment in aphasia at all levels of severity.
2. The processing nature of the language/STM impairment (weak activation or too-rapid decay of activated semantic and phonological representations),
3. The ability to activate and maintain activation of language representations in the contexts of delayed response time, increased memory load and/or verbal interference.

The TALSA includes three groups of subtests:

1. Language tasks with filled and unfilled intervals between a) two stimuli to be compared in some way or b) stimulus and response.
2. Judgments of semantic and phonological similarity that vary working memory load (comparing meanings or sounds of two vs. three words).
3. Verbal span tasks that vary characteristics of stimuli in the span to probe semantic or phonological levels of word processing.

In what follows, we provide the theoretical and clinical motivation for development of this test, as well as empirical support for its assumptions and content. Second, we provide details of the tasks in the test battery, their rationale, normative data from individuals with aphasia as well as age-matched controls, and reliability and validity measures. Third, we present data from the TALSA that demonstrate the positive or negative effects of an increase in response time or WM load on language performance. In the discussion, we focus on some research and clinical applications of the test battery and its value in providing an assessment of aphasia that is highly specific in its detail of linguistic and processing impairment and highly sensitive to all levels of impairment severity. We also discuss the relevance of the data from this test battery to functional communication in aphasia.

Theoretical motivation for developing the TALSA battery—Definitions of aphasia quite naturally focus on the linguistic characteristics of the impairment, but, as noted above, more recent models acknowledge the involvement of STM and other cognitive processes (Darley, 1982; McNeil, 1982; McNeil & Pratt, 2001; Murray, 2004). Even as early as the 19th century, the neurologist Hubert Grashey (1885) attributed anomia to a rapid decay of word representations. A resurgence of interest in this topic in the late 1960s grew from observations of a nearly ubiquitous co-occurrence of verbal STM and word processing impairments in aphasia, which motivated investigations to determine the role of STM in language processing and aphasia. Two viewpoints emerged at this time, each linked to influential cognitive models of verbal processing and memory. Some researchers hypothesized that verbal STM and verbal processing were dissociable systems. (e.g., Vallar & Baddeley, 1984; Warrington & Shallice, 1969). This line of investigation was greatly influenced by Baddeley's WM model (Baddeley, 1986; Baddeley & Hitch, 1974), which hypothesizes three components, a central executive and two subsidiary systems that store (1) speech representations (the phonological loop) and (2) visual-spatial representations (the visuo-spatial sketchpad). More recent versions of Baddeley's working memory model (Baddeley, 2000) include an episodic buffer which is a multimodal temporary store of information that connects the phonological loop and visual-spatial sketch pad with information from long-term memory.

At about the same time, other researchers invoked language-based models of verbal STM such as the 'levels-of-processing' model advanced by Craik & Lockhart (1972) as a framework to understand the co-occurrence of language and STM impairments in aphasia (e.g., Berndt & Mitchum, 1990; Friedrich, 1990; Saffran, 1990; Saffran & Martin, 1990; R. Martin et al., 1994). The Levels-of-processing framework resulted in the emergence of two important assumptions of multi-store models: (1) temporary storage of language representations is a property of the language system rather than a separate memory function (see Monsell, 1984; Saffran, 1990), and (2) verbal STM includes temporary storage of phonological as well as lexical and semantic representations (e.g., Martin & Gupta, 2004; Martin, et al., 1994), and even conceptual representations (Potter, 1993). Below, we describe the Interactive Activation (IA) model's account of the multi-store, language-based view of verbal STM and provide empirical support for this model.

Our hypothesis of the relationship between language processing and verbal STM derives from the theoretical framework of Dell's Interactive Activation (IA) model of word production (e.g., Dell, et al., 1997; Dell & O'Seaghdha, 1992; Schwartz, et al., 2006), which is depicted in the central part of Figure 1. This localist connectionist model of word processing is similar to psycholinguistic and information processing models developed in the 1980s and 1990s (e.g., Morton, 1980; Patterson & Shewell, 1987) in that it includes levels of a word's component representations (e.g., semantic, lexical, phonological). However, it expands on these earlier models with greater detail of the connections (depicted by arrows) between the levels of a word's representation and the role of those connecting processes in language performance.

Word retrieval is mediated by interactive spreading activation (depicted as the arrows) that primes the target word (e.g., cat) and semantically or phonologically related words. When retrieval occurs, it is the most highly active word node that is selected via a competitive activation process. Spreading activation throughout the semantic-lexical-phonological network takes place over time and is controlled by two parameters, connection strength (rate of activation spread) and decay rate (rate of activation decay). These parameters work together to maintain activation of an intended utterance's semantic, lexical and phonological representations over the time course of a language task. Word processing impairment in aphasia is characterized in terms of the linguistic representations that are affected (e.g., semantic or phonological) as well as the processing impairment, expressed in terms of connection weight and decay rate parameters. If connection strength is weak, activation spreads too slowly. If decay rate is too rapid, representations are activated but decay too quickly. The outcome of each impairment is seen in differing patterns of word retrieval errors (e.g., greater proportions of semantic errors than phonological). The TALSA battery is designed to identify error patterns reflecting both types of processing impairments. Evidence of both patterns will be discussed when presenting the data from the TALSA test battery.

Temporal component of single word processing and extension of the model to repetition, multiple-word processing and verbal STM—The stability of connection weight and decay rate parameters over the temporal course of word processing is central to accuracy of word retrieval. Martin and Saffran (1992) extended the IA model to repetition and outlined the temporal course of repeating a single word in a case study of the repetition and verbal STM impairment in deep dysphasia (Martin, 1996; Martin, Dell, Saffran & Schwartz, 1994; Martin et al., 1996). On either side of the IA model in Figure 1, there are depictions of the pathways through the levels of representation for production (left side) and repetition (right side). In naming, activation begins at the conceptual level and advances forward to semantic, lexical and phonological levels of representation. In repetition, activation begins with the phonological level of representation and then spreads to the lexical and semantic networks with feedback activation spreading to preceding levels of activated representations at each step. This distinction is important clinically, especially when evaluating how an impairment at semantic or phonological levels of representation will affect access to words in tasks that involve input processing (e.g., rhyming judgments), output processes (e.g., naming), or both (e.g., repetition).

Martin & Gupta (2004) extended the IA repetition model further (Figure 2) to depict the hypothesized role of word processing in verbal STM. In a verbal span task, a sequence of words is heard and subsequently recalled (repeated) in the same order in which the words were heard. They hypothesized that spreading activation in the phonological-lexical-semantic network of each word in that sequence is held in a temporary store until a response to reproduce the sequence is required. Thus, connection weight and decay rate parameters support the activation and maintenance of each single word in the sequence, and impairment to either or both parameters affects performance in single word and multiple word tasks, including verbal span tasks. Although other factors such as rehearsal also support maintenance of the contents of verbal STM (e.g., Baddeley & Hitch, 1974), the evidence from aphasia indicates that activation and maintenance processes that enable access to and retrieval of words also support retention of verbal representations in verbal STM.

Empirical support for a common source of word processing and verbal STM impairments in aphasia—In typical speakers, it is well established that verbal span varies depending on characteristics of items to be recalled, digit span > word span (Brener, 1940) and word span > nonword span (Hulme, Maughan & Brown, 1991). Span capacity is also influenced by phonological similarity (Conrad & Hull, 1964), word frequency (Hulme, Roodenrys, Schweickert, Brown, Martin & Stuart, 1997; Watkins & Watkins, 1977) and semantic similarity (Shulman, 1971). In the case of speakers with aphasia, studies show that lexical and semantic factors (e.g., imageability) influence verbal span quantitatively (i.e., number of items recalled (Martin & Saffran, 1997) and qualitatively (i.e., serial order effects (Martin, et al., 1996; Martin & Saffran, 1997). Additional evidence for temporary storage of semantic information comes from case studies documenting production of semantic errors in repetition of single words (Howard & Franklin, 1988; Martin et al, 1996) and paraphrases of sentences in immediate repetition (Saffran & Marin, 1975).

There is considerable evidence from studies of aphasia to support a close relationship between verbal STM and word processing. As studies of verbal STM in typical speakers have shown, span capacity is fluid in that it varies depending on the nature of the items being recalled. This is true in aphasia, as well. Studies indicate that span size varies depending on items to be recalled (e.g., word span > nonword span (e.g., Kalinyak-Fliszar, Kohen & Martin, 2011; Martin, Kohen & Kalinyak-Fliszar, 2010), high image > than low image word span, and high frequency > low frequency word span (Martin, et al., 1996; Martin & Ayala, 2004; Martin & Saffran, 1990; 1992; 1997).

Additional evidence comes from observed associations between verbal STM capacity and recovery from aphasia. Martin, Saffran and Dell (1996) studied the recovery patterns of word processing and verbal STM in a case of deep dysphasia and documented increases in verbal STM span size that coincided with improvement in word processing ability. Martin and Gupta (2004) demonstrated that verbal span size in individuals with aphasia was directly correlated with their performance on naming and lexical decision tasks, which assess the integrity of word processing ability. A relationship of severity of aphasia and verbal span was also observed by Potagas, Kasselmis, and Evdokimidis (2011).

Task differences also influence span performance in aphasia. Using a word-to-picture pointing span task and a word repetition span task, Martin and Ayala (2004) demonstrated that span capacity varies depending on the pathways through the lexical network that are used in each of these tasks and how those tasks interact with a person’s language impairment. Pointing span, which engages input phonological, lexical and semantic processes, correlates with lexical and semantic ability. Repetition span, which needs only to engage input and output phonological processes, correlates positively with measures of phonological ability, but not semantic ability. These findings support the hypothesis that the verbal span capacity of someone with aphasia is systematically influenced by the primary source of their word processing impairment (semantic or phonological) and the type of task that is used to assess verbal span capacity.

Finally, there are studies of serial position effects in verbal span recall in typical speakers showing that semantic factors influence retention of initial items in a sequence (i.e., *primacy* *effect*, Poirier & Saint-Aubin, 1995; Poirier, Saint-Aubin, Mair, Tehan, & Tolan, 2015; Shulman, 1971), and phonological factors influence retention of later items in a sequence (i.e., *recency effect*, Brooks & Watkins, 1990). In aphasia, reduction of primacy effects in serial recall is associated with semantic impairment and reduction in recency effects with phonological impairment (Martin & Bunta 2007; Martin & Saffran, 1990; 1997; Minkina & Martin, 2016). These studies show that verbal span capacity is not fixed, but depends on items to be recalled, paradigm used to assess span, and in aphasia specifically, the source of language impairment (semantic vs. phonological).

Taken together, these findings are consistent with the hypothesis that the activation parameters, connection weight and decay rate, which support stability (e.g., activation strength and maintenance) of semantic and phonological representations in single word tasks, also affect their stability in multiple word tasks, including verbal span tasks. On this account, the co-occurrence of language and verbal STM impairments in aphasia is a consequence of a single impairment to processing mechanisms that determine the strength at which words are activated and the short-term retention of that activation and short-term maintenance of that activation (e.g., Martin et al., 1996; Martin & Ayala, 2004; Martin & Saffran, 1997). In terms of clinical diagnosis, the primary distinction between word processing and verbal STM impairments in aphasia is the severity of that impairment. When it is severe enough to compromise maintenance of the activation of a single word's semantic and/or phonological representations, performance on single word tasks (e.g., repetition, naming, word-to-picture matching) could be affected. When it is milder, the activation-maintenance ability may be sufficient to support single word processing, but falters when called upon to maintain activation of multiple words. In this case, the deficit will affect performance on multi-word tasks such as verbal span (pointing or repetition) and sentence repetition. Note, however, that this continuum of impairment severity applies to erroneous retrieval of semantic and phonological representations of words and not to production errors that are attributable to apraxia, dysarthria or articulatory disturbances.

The present study

The TALSA battery was developed with the clinical implications of this research in mind. The test expands on accomplishments of earlier test batteries for aphasia which profile language impairments at the task level (e.g., naming or comprehension, as in Western Aphasia Battery-Revised (WAB-R, Kertesz, 2006) or Boston Diagnostic Aphasia Examination (BDAE, Goodglass, Kaplan & Barresi, 2001) and the level of psycholinguistic sources of that impairment (e.g., semantic, phonological as in the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA; Kay, Lesser & Coltheart, 1992) and the Comprehensive Aphasia Test (CAT; Swinburn, Porter & Howard, 2005). The TALSA introduces an additional focus on the temporal aspect of word processing and the effects of increased STM/WM load on language function. Four assumptions, based on the research findings discussed above, guided the development of the TALSA battery.

1. Word processing occurs over time and therefore has a temporal component.

2. Aphasia, in part, is an impairment of the activation and short-term maintenance of word representations that are accessed and retrieved in language comprehension, repetition and production.

3. Impairment can affect either or both of two processing parameters: (1) connection strength, which affects rate/strength of activation (slowed or weak activation) and (2) decay rate, which affects short-term maintenance of linguistic representations (too-fast decay of activation).

4. Adverse effects of the impairment on activation parameters that support word processing become apparent in language performance when time available to access or retrieve words is mismatched with rate of access (slowed connection weight) or rate of activation decay (too-fast decay).

In what follows, we report our initial findings with the TALSA Battery. First, we describe the subtests within the battery, including their linguistic and memory load characteristics, and provide normative data from 39 individuals with aphasia and 16 age-matched control participants without aphasia or brain damage. The norms are provided individually for each subtest as the number of participants varied for each subtest. We then provide some assessment of reliability (internal consistency) for subtests in Part 1 of the battery. We also demonstrate how integrating a 5-second time interval into a subset of language tasks in Part 1 results in two distinct patterns of performance (increase or decrease in accuracy). Finally, in the discussion, we review some current and potential applications of the TALSA to research and clinical activities, strengths and limitations of the TALSA and this study, as well as future directions.

Method

Participants

Thirty-nine individuals with aphasia resulting from a left cerebral vascular accident (LCVA) participated in this study. Based on diagnoses reported in medical records and/or classification on the Western Aphasia Battery-Revised (WAB-R, Kertesz, 2006) the participants with aphasia included four with Wernicke's, seven with Broca's, nineteen with anomic, six with conduction, and two with transcortical motor aphasia. One participant's aphasia classification was not available. Age range of the participants with aphasia was 32–78 (mean = 56.03, SD = 11.12). Time post-onset before enrollment in this study was at minimum, six months, but ranged upward to 301 months. Table 1 provides demographic information on the 39 people who participated in this study including education, sex, age at time of testing, time post onset (in months) etiology, WAB AQ and Handedness. Sixteen age-matched control participants, nine females and seven males, were enrolled in this study. The age range was 31 to 78 years (mean = 56.33, SD= 15.33) and their education ranged from 11 to 16 years. All participants voluntarily enrolled in this research program and signed a consent form approved by the Internal Review Board at Temple University.

Overview of the development of the TALSA Battery

The TALSA battery includes tests that are commonly used in assessment of language abilities in aphasia, but also has several unique features that allow for assessment of verbal

STM capacity and effects of increased memory load on language processing. There are three parts to this battery; Parts 1 and 2 include word processing tasks with STM/WM variations and Part 3 includes span tasks with linguistic variations.

The word processing tasks in Part 1 probe semantic and phonological abilities, vary in difficulty to be sensitive to a full range of impairment severity, and incorporate a time interval between stimulus and response or between two stimuli to be compared to assess the effects of time delay on performance. There are three interval conditions: 1-second unfilled (1-sec UF), 5-second unfilled (5-sec UF) and 5-second filled (5-sec F). Adding a 5-second response delay (e.g., in the word repetition subtest) or five second interval between presentation of two stimuli to be compared (e.g., in the phoneme discrimination task) is one means of increasing the STM load of the language task. Thus, comparison of the 1-sec UF and 5-sec UF conditions allows for assessment of the ability to passively maintain activation of verbal representations in STM. A third response condition is a 5-second 'filled' interval that includes a distracter task during the interval in which the participant attempts to name numbers (single digits 1–9) that appear in random order on the computer screen, one at a time. This dual-task situation (rehearsing the stimulus and counting numbers on a screen) requires maintaining activation of linguistic representations in the context of verbal interference, which adds STM load and draws on executive processing abilities.

Part 2 includes two word-level similarity judgment tasks (rhyming and synonymy) that vary the working memory load of these tasks by increasing the number of items that need to be held in verbal STM while making the similarity judgment.

Part 3 includes span tasks that vary phonological, lexical and semantic characteristics of the stimuli (e.g., frequency, imageability, lexicality). These are intended especially for individuals with mild aphasia who have little difficulty with single word processing tasks, but falter with multiple word processing tasks. The verbal span tasks are sensitive to spared and impaired semantic and phonological abilities and are ideal for diagnosing the linguistic source of a language impairment of those individuals whose main difficulty is in comprehending and producing multiple word utterances.

Development of test stimuli—Details of stimuli for each subtest in the battery will be described below, but some general characteristics are noted here. All subtests were administered on the computer using E-prime software (Schneider, Eschman & Zuccolotto, 2012). Stimuli included pictures and recordings of words, nonwords, sentences, and sequences of words, nonwords, and digits.

Pictures: For those subtests using picture stimuli, we used black and white line drawings from several sources, including therapy materials (Abbate & La Chappelle, 1984a, b) as well as artist renderings.

Auditory stimuli: All verbal stimuli were recorded in a female voice in a sound booth and sound files were incorporated into the E-Prime programs for each subtest.

Test content: The procedures and some of the stimuli from the subtests in this battery were drawn from tests that have been developed for previous studies (e.g., Martin & Saffran, 1992; 1997, 1999). For each subtest, stimuli were controlled for or varied for variables that were most critical to the level of linguistic representation that we aimed to probe with a particular task. For example, for picture naming, we used highly imageable words, but varied the frequency values and syllable length of the items to be named. For the Lexical Comprehension subtest (word-to-picture matching with semantically related distractors), category relationships were the dominant characteristic that was controlled. Word length and frequency were varied, but not systematically.

Because some of the subtests in this battery were adapted from longer versions that were used in the first author's lab, the databases used to rate their stimuli characteristics such as frequency and imageability were from sources available in the 1990's when they were first created (e.g., frequency: Kucera & Francis, 1982; concreteness: Kroll & Merves, 1986; imageabilty: Paivio, Yuille & Madigan, 1968). Where possible, we have updated the ratings of norms using more current databases (e.g., Clark & Paivio, 2004), but this could not be done for most of the subtests. However, the clinical version of the TALSA test battery that is under development will use the most current sources of frequency, imageability, concreteness and other stimulus variables and will use only original, artist-created drawings.

TALSA Battery Part 1. Word and Sentence Processing Tasks That Incorporate Unfilled and Filled Temporal Intervals

Part 1 includes language tests that incorporate the three interval conditions described above which are intended to assess effects of increasing the STM demands of the task (5-sec UF interval) and increasing the working memory load (5-sec F interval) on response accuracy. These are:

1. input phonological processing (phoneme discrimination, rhyming judgments), input lexical-semantic processing (lexical comprehension, category judgments),
2. input and output phonological and lexical-semantic processing (word and nonword repetition),
3. output semantic and phonological processing (picture naming), (4) sentence comprehension and repetition subtests.

Below, we describe the purpose, stimuli, procedure and outcome measure (s) for each test in the battery. Note that for outcome measures that involve verbal output (e.g., naming nor repetition) responses were scored as correct if they were (1) completely accurate or (2) included articulatory distortions that could be attributed to apraxia or dysarthria or (3) included sound productions typical of African American Vernacular English (e.g., reduction of final consonant clusters) or regional pronunciations (e.g., water, /watʊ/, typically pronounced as /wʊdɚ/, wooder, in the Philadelphia region).

Phoneme Discrimination

Purpose: This subtest examines early input phonological processing using an identity judgment task requiring the participant to decide whether two words or two nonwords sound the same or different. The effects of increased STM and WM load on accuracy of phoneme discrimination are also tested.

Stimuli: There were 20 word and 20 nonword pairs. Each of these sets had 10 pairs with the same items and 10 pairs with different items. The words were concrete (ratings from Kroll and Merves, 1986), > 5.50 on a scale of 1 to 7 with 7 being most concrete), and low in frequency (Kucera & Francis, 1967), < 25 occurrences per million). Nonwords were derived from real words by changing one to three phonemes from the initial, medial or final positions. Those pairs with different items were created by altering one (consonant and/or stressed vowel), sampling from initial, medial, and final positions. The time interval between the two words varied depending upon the interval condition.

Procedure: Word or nonword pairs were presented auditorily via E-Prime software in two sets of mixed lists. Nonword pairs were never presented in the same set with the word pairs from which they were derived. Participants listened to each word or nonword pair and determined whether the two items in the pair were the same or different. The word and nonword pairs were presented under the three interval conditions that vary the time interval between the two items.

Outcome measure: The primary measure of interest was the proportion of correct judgments of same and different pairs at each interval condition. Word and nonword items were scored separately.

Rhyming Recognition

Purpose: This subtest also examines input phonological processing using a similarity judgment task that requires the participant to decide whether two words or two nonwords rhyme. This task draws on the ability to segment the onset and rhyme portions of the words. The effects of increased STM and WM load on accuracy of rhyme judgments are also tested.

Stimuli: There were 20 word pairs (10 rhyming and 10 non-rhyming) and 20 nonword pairs (10 rhyming and 10 non-rhyming). Stimuli were one syllable in length. They were not controlled for frequency or imageability. Although the stimuli were presented auditorily, we varied the orthography to include rhyming pairs with same or different orthography (e.g., break-steak, bare-pear). This was done to avoid any possible bias that might arise from the participant mentally visualizing the orthography of the words in the pair. The nonwords were derived from the real words by altering one or two phonemes, sampling equally from initial, medial and final positions of the word.

Procedure: The participant listened to pairs of words or nonwords and decided whether they rhymed or not by pushing keys on a keyboard marked YES or NO. Word pairs and nonword pairs were randomly presented in two sets. Nonword pairs were never included in the same set as the word pairs from which they were derived. The word and nonword pairs were presented under three conditions that varied the time interval between the two items, 1 sec-UF, 5-sec-UF, and 5-sec F.

Outcome measure: The primary measure of interest was the proportion of correct judgments of rhyming and non-rhyming pairs at each interval condition. Scoring was separate for words and nonwords.

Lexical Comprehension

Purpose: This task assesses the integrity of the input lexical-semantic pathway by examining the ability to access a word's semantics from its spoken word form in the context of a spoken word-to-picture matching task. The effects of increased STM and WM load on accuracy of word-to-picture matching are also tested.

Stimuli: There were three sets of imageable words, 16 items in each set. Across all three sets, there were 20 different semantic categories represented. Words and pictures representing each of the 48 total items are grouped with three semantically related distractors (words and pictures). Word frequency varied across all test items, as the primary stimulus variable of interest was semantic category membership. However, all test items had frequency values less than 57 and all but four items had frequency values less than 20 (Pastizzo & Carbone, 2007).

Procedure: This subtest required matching a spoken word to one of four pictures that were members of the same semantic category. Participants heard a word, then four pictures appeared on the screen after one of three interval conditions, 1-sec UF, 5-sec UF and 5-sec F. The participant pointed to the picture that he/she thought was a match to the spoken word.

Outcome measure: The primary measure of interest was the proportion of correct word to picture matches at each interval condition.

Category Judgments

Purpose: This subtest measures the ability to access knowledge of category membership through words (verbal semantics) and through pictures (conceptual semantics).

Stimuli: There were three sets of word pairs, 20 in each set, and three sets of picture pairs which correspond with the words. The 20 item pairs reflect five different semantic categories (animals, transportation, vegetables, furniture, musical instruments). Picture and word stimuli were presented in separate subtests. Each set of pictures or words was assigned to an interval condition, 1-sec UF, 5-sec UF or 5-sec F.

Procedure: Two items (words or pictures) were presented in succession. The first was on the screen for 3 seconds and went off before the second item appeared. The second item appeared after one of the three intervals, and remained on the screen for 3 seconds. The task was to determine whether the two items belonged to the same category (e.g., apple, banana) (e.g., apple, hammer) by pressing keys on a keyboard labeled YES or NO. In the written word condition, words in the pair were presented auditorily and visually on the screen. In the picture condition, stimuli were presented on the screen, with no names provided.

Outcome measure: The primary dependent outcome was the proportion of correct responses (same or different category) for the spoken/written word pairs and picture pairs at each interval condition.

Word and Nonword Repetition

Purpose: This subtest assesses the ability to repeat stimuli with lexical representations (known words) and without (phonotactically legal nonwords). The effects of increased STM and WM load on accuracy of word and nonword repetition are also tested.

Word repetition involves access to phonological, lexical and/or semantic levels of representation. Nonword repetition is mediated primarily by the phonological route without significant support from lexical or semantic levels. Activation of phonemes in nonwords could spread to the lexicon and prime words that are phonologically related to the nonword, but this is not necessary.

Impaired repetition of words can be indicative of impaired input processing (access to the lexicon from input phonological representations), impaired output processing, or access to output phonological representations from the lexicon.

Impaired repetition of nonwords indicates impairment of the phonological input and/or output pathways. Error types can provide some clues to what part of the input-output pathway is compromised. For example, lexicalization errors (producing a phonologically related word instead of the target nonword) indicate that a direct pathway between input and output phonological networks is impaired, and point to a reliance on lexical representations to repeat the nonword.

Stimuli: There were 15 words high in imageability (based on Paivio et al., 1968, > 4.97, the mean of ratings on a 1–7 scale) and high in frequency (based on Kucera & Francis, 1982, ratings > 40 occurrences per million). There are five each of 1, 2, and 3 syllables. There were also fifteen nonwords matched for length and CVC structure. They were created by altering one or two phonemes from initial, medial or final positions.

Procedure: Word and nonword conditions were presented separately. Stimuli were presented one at a time auditorily via E-Prime software. The task was to repeat the stimulus after an auditory cue (i.e., ‘beep’ sound) that occurs after 1 sec UF, 5-sec UF or 5-sec F interval, depending on the condition.

Outcome measures: The primary outcome measure was the proportion of correct repetitions in each of the response interval conditions. A second outcome measure of interest was the types of errors made in repetition.

Picture Naming

Purpose: This subtest examines the integrity of the output semantic-lexical-phonological pathways to production through a picture naming task. The effects of increased STM and WM load on accuracy of word retrieval in this picture naming task are also tested.

Stimuli: There were three sets of 30 pictures, consisting of one-, two-, and three-syllable names. The sets were matched for syllable length and frequency (Pastizzo & Carbone, 2007). The words were divided by frequency ratings, high (>25, range 27 to 673) and low (< 25, range: 1 to 9 for those listed, n = 32, plus 13 not listed, assuming rating of < 1), and were distributed equally across the three sets. Each set was assigned to one of three interval conditions, 1-sec UF, 5-sec UF and 5-sec F. Syllable length and word frequency were balanced across all three sets.

Procedure: A picture appeared on the screen for 2000 milliseconds, followed by a cue to name (beep) after one of the 3 interval conditions. Scoring of this subtest was modeled after the scoring protocol of the *Philadelphia Naming Test* (Roach et al., 1996).

Outcome measures: The primary outcome measure was the proportion of correctly named pictures in each of the response interval conditions. A second outcome measure of interest was the types of errors made in naming (e.g., semantic or phonological substitutions).

Sentence Repetition

Purpose: This subtest examines the effects of increased memory load on accuracy of sentence repetition. The effects of increased STM and WM load on accuracy of sentence comprehension in a sentence-to-picture matching task are also tested.

Stimuli: There were three sets of 20 transitive sentences. Ten in each set included a noun phrase, verb phrase, and prepositional phrase (e.g., The boy walked the dog in the park). These had four content words and across the three sets, an average of 10.03 total words each (range 7–13). The other 10 sentences also were transitive with a noun phrase, verb phase and prepositional phrase, but were ‘padded’, with two modifiers preceding the subject, object of the verb or object of the preposition (e.g., The *tall* boy walked the dog in the *public* park). Thus, there were seven content words in each sentence and across the three sets, an average of 14.7 syllables in each sentence (range 12–17). Content words had concreteness ratings > 5.50, based on Kroll and Merves (1986). Sentences were balanced across the three sets with equal numbers of present progressive and past tense verbs, number of syllables (based on sentence type (e.g., present, past)). No irregular tense verbs were used.

Procedure: Each sentence was presented auditorily and the participant was asked to repeat the sentence when a cue is provided after one of the three intervals.

Outcome measures: The proportions of content words repeated correctly and in serial order were calculated for the unpadded and padded sentences.

Sentence Comprehension

Purpose: This subtest is intended to assess comprehension of sentences, and transitive sentences with semantically reversible arguments. The effects of increased STM and WM load on accuracy of sentence comprehension in a spoken sentence-to-picture matching task are also tested.

Stimuli: This spoken sentence-picture matching task was modeled after the *Philadelphia Comprehension Battery* (Saffran et al., 1988) and used stimulus pictures and test items from this unpublished test. There were 20 sentences that represented five syntactic structure types: Simple active declarative, Passive, Subject relative clause, Object relative clause, Locatives. All sentences had an agent and a patient that were semantically reversible. Each sentence was paired with two distracter type sentences (with corresponding pictures). Lexical distracter sentence/pictures replaced the agent or patient with another object or being (e.g., *The policeman shoots the robber* vs. *The policeman shoots the dog*). Reversible distracter sentence/pictures reversed the objects or beings in the agent and patient roles (*The policeman shoots the robber* vs. *The robber shoots the policeman*). There were 4 exemplars of each syntactic structure. The two sets of sentence pairs (20 semantically reversible argument distracter and 20 with lexical distracter) were randomized and divided into two sets of 20 such that a sentence token never appeared with its reversible and lexical distracter in the same set.

Procedure: The sentences were presented auditorily one at a time. The participant listened to the sentence and then two pictures appeared on the screen, one depicting the spoken sentence and the other picturing a distracter sentence. The task was to point to the picture that matches the sentence. Each set was tested in the three interval conditions, 1-sec UF, 5-sec UF and 5-sec F.

Outcome measures: The primary dependent measure was the proportion of sentences matched correctly to pictures for each of the sentence conditions (reversible semantic roles and nonreversible semantic roles) and each sentence structure type (e.g., active, passive etc.) in each of the three interval conditions.

TALSA Battery Part 2. Judgments of Semantic and Phonological Similarity Under High and Low Working Memory Load Conditions

This section of the TALSA battery varies working memory load by presenting words that need to be compared for rhyming or meaning similarity in two formats that vary the number of pairs to be compared in working memory. These triplet judgment tasks are described below.

Rhyming triplet judgments

Purpose: This is a verbal working memory task that primarily taps into phonological working memory. It assesses the ability to make rhyming judgments under high and low working memory conditions.

Stimuli: Stimuli were spoken 1-syllable, picturable nouns with consonant-vowel structures (CVC, CCVC, CCVCC, CVCC) and digitized pictures corresponding to the words. There were 30 triplets tested in each of two presentation conditions that varied the working memory load of the task (see below). Although we did not rate words in this subtest for imageability, we used words that were pictureable. Word frequency values were not considered in selection of the stimuli, as the phonological composition and picturability of the items was the primary concern. However, retrospective analysis of the word frequency values of the stimuli using the Pastizzo and Carbone (2007) database indicated that of the 75 words listed in the database, 52 were rated < 25, range 1–23 (low frequency) and 23 were rated > 25, range 27–365 (with one outlier, “well” rated at 4917).

Procedure: Sequences of three pictures were presented on the screen, diagonally from top-left to the bottom right of the screen. Two of the picture names rhymed and the non-rhyming foil overlapped phonologically with one or two of the rhyming words in one of three ways: same initial phoneme (e.g., *fan, pan, pail*), same stressed vowel, (e.g., *bag, rag, cat*) or same final phoneme (e.g., *mouse, house, dice*).

The task was to identify two of three words that rhyme, but under two context conditions that varied the number of words or word pairs to compare in working memory. In the low memory load condition, the picture in the center of the screen was highlighted. The spoken name of the center picture was presented first (e.g., *bag*) followed by the names of the other two pictures (e.g., *rag, cat*). The task was to decide which of those two picture names rhymed with the name of the picture in the center. For this version of the task, then, there were two possible rhyme pairings (e.g., *bag-rag, bag-cat*). In the high memory load condition, the same three pictures were presented in the same visual format, and names were presented auditorily in the same sequence as the picture display (from top-left to bottom-right). The task in this condition was to decide which two of the three words rhymed. Thus, in this version of the task, there were three possible rhyme pairings (e.g., *bag-rag, bag-cat, cat-rag*).

Participants were instructed in each memory load condition to listen to all three words before making their decision. The strategies used in each condition to determine which words rhymed might have involved holding two vs. three words in working memory or two vs. three word pairs. The results from the present study do not distinguish between these two possibilities.

The high and low memory load conditions were split into two sets (15 items each). All four sets were tested in an ABBA design (high-set one, low-set two, low-set one, high-set two). Set one of the high and low memory load conditions included the same items as set two of the high and low memory load conditions, but item order was randomized within each subset.

Outcome measure: The proportion of correct judgments of rhyming words was calculated under each memory load condition.

Synonymy triplet judgments

Purpose: This is a verbal working memory task that primarily taps into lexical-semantic working memory. Two presentation formats allow assessment of the ability to judge similarity of word meanings under high and low working memory conditions.

Stimuli: There were 40 word triplets altogether, 20 noun triplets and 20 verb triplets. Within noun and verb sets, there were 10 with concrete words and 10 with abstract words (based on Kroll & Merves, 1986). Concrete and abstract word triplets were blocked in presentation.

Because the test was concerned primarily with comparison of word meanings, the stimuli were not controlled for word frequency within or across triplets. However, using the Pastizzo & Carbone (2007) spoken frequency word norms, it was determined that the mean and range of frequencies of words in each concreteness-word class condition were as follows: concrete nouns - Mean: 48.2, Range: 1–80; abstract nouns – Mean: 49.9, Range: 1–274; concrete verbs – Mean: 20.7, Range: 1–125; abstract verbs – Mean: 389.6, Range: 1–6557 (Note: The highest rating was an outlier and when removed, mean was 229.4 and range was 1–862). Although the abstract verbs were generally higher in frequency, this did not add any apparent advantage in making the judgments. As is typically shown in many studies (reviewed earlier), performance on many tasks using abstract or low imageability words is lower than those same tasks using concrete or high imageability words. This was the case with performance on this subtest (see Martin et al., 2012).

Procedure: In the low memory load condition, three printed words with similar meanings (e.g., *jail prison, cage*) were presented diagonally on the screen from the top-left corner to the bottom-right corner. They were also presented auditorily via E-Prime. The word in the center (*prison*) was designated as the target word and the task was to determine which of the other two words (*jail, cage*) was closest in meaning to the target word (*prison*). Participants could hear the words up to five times to help insure that they were certain of the words that were being compared. This version required consideration of two potentially synonymous pairs (*jail-prison, jail-cage*). In the 3-pair condition, the same three words (e.g., *jail, prison, cage*) were presented in the same way (auditorily with words diagonally on screen from top left to bottom right), but the task was to decide which two of the three words were most similar in meaning. This format required participants to consider three potential synonymous word pairs.

As in the rhyming triplet judgment task, participants were instructed in each memory load condition to listen to all three words before making their decision. The strategies used in each condition to determine which words had the same meaning might have involved holding two vs. three words in working memory or two vs. three word pairs. The results from the present study do not distinguish between these two possibilities.

The high and low memory load conditions were split into two sets (20 items each). All four sets were tested in an ABBA design (high-set one, low-set two, low-set one, high-set two). Set one of the high and low memory load conditions included the same items as set two of the high and low memory load conditions, but item order was randomized within each subset.

Outcome measure: The proportion correct in sets of concrete nouns and verbs and abstract nouns and verbs was calculated under each memory load condition. Although not reported in this paper, other outcome measures could include a comparison of accuracy on concrete vs. abstract triplets (see Martin et al., 2012) in the two memory load conditions or differences based on word class.

TALSA Battery Part 3. Span Measures with Language Variations

In Part 3 of the TALSA Battery, we adapted the standard immediate serial recall span task and varied characteristics of the items to be recalled in ways that would make the span measure sensitive to levels of representation (e.g., semantic, lexical, phonological). In this way, the span tasks can serve as a diagnostic for individuals whose aphasia is mild and who frequently have no difficulty processing tests that are limited to single words (e.g., the subtests in Part 1). For example, they may perform well on word and nonword repetition of single words and yet start to show a disparity between these two stimulus types when asked to repeat more than a single item. Three types of items were used in these span tasks: digits, words, and nonwords. Except for the probe memory tasks, stimuli were limited to 1 or 2 syllables. Auditory stimuli for all the span measures were digitally recorded. Spans for "serial order" and "in any order" were calculated for the first three measures (but only serial order results are reported here). Additionally, although we do not report these data here, we also administered the Corsi Block Span Task to all participants as a measure of nonverbal span (Corsi, 1972).

Digit and word spans

Purpose: Verbal short-term memory span for digits and words is assessed by means of a pointing task and a repetition task to determine if task differences affect verbal span capacity in aphasia (e.g., Martin & Ayala, 2004). Also, digit and word span are increasingly used in studies of aphasia as standard measures of verbal STM span.

Stimuli: Digit and word spans were assessed up to seven items. In each string length condition, there were 10 lists of digit sequences or word sequences. Number sequences were generated from a finite set of nine digits (1–9) but with no repeats of digits within a sequence. Distribution of digits was balanced across all digit strings such that each digit occurred 31 times in the test except for the number 4 which occurred 32 times. The word span subtest consisted of a finite set of 9 words, all of which were high in imageability (Clark & Paivio, 2004), high in frequency (Pastizzo & Carbone, 2007), and matched in syllable length to the nine digits (eight 1-syllable words and one 2-syllable word). The four span conditions, digit-repeat, word-repeat, digit-point, and word-point, were administered separately. Words and digits were digitally recorded and presented in sequences programmed with E-Prime software.

Procedure: The procedures for the digit and word span subtests were identical. In repetition span tasks, the participant heard a sequence of words or digits and was instructed to repeat the sequence in the same order after hearing it in its entirety. In the pointing span tasks, the participant heard a sequence of words or digits, and this was followed by presentation of a 3 × 3 block visual array of digits or pictures corresponding to those they heard on the screen. The participant was asked to point to the digits or object pictures corresponding to those in the spoken sequences and in the same order in which they were presented. The array of digits or objects was randomly rearranged on each trial.

Outcome variable: Two spans were calculated, one based on recall of words in serial order and the other based on recall of words in any order. This distinction is important theoretically, as there is some evidence that these two can be dissociated in aphasia (Majerus, 2012). There is clinical relevance as well, as it has been shown that some individuals with aphasia who have difficulty accessing semantics tend to repeat only the final items of a sequence of digits or words that are supported by recent phonological activation (e.g., Martin & Saffran, 1997). This pattern of preserved access to recent items (recency) and loss of access to initial items is diagnostically significant, indicating impaired access to semantics and relatively preserved access to phonological representations. Span size was calculated using Shelton, Martin and Yaffee's (1992) formula: Sequence length at which at least 50% of the sequences are recalled + .50 (proportion of sequences recalled in the next sequence length).

Repetition span for words varied for imageability and frequency

Purpose: Word frequency reflects how often a word's form occurs in a language (in this test, English). Imageability reflects how easy it is to conjure an image of a word's meaning. Words can be associated with highly imageable concepts (e.g., dog or dress), or they can be associated with less imageable concepts (e.g., justice or logistics). These two variables are closely linked to concrete and abstract variables, respectively. In this subtest, both frequency and imageability were varied systematically, but imageability is the variable of greater interest diagnostically. In any word repetition task, the words to be repeated are supported by activation of both their phonological and semantic representations; however, the degree to which each level of representation supports repetition varies depending on the word's imageability. As imageability reflects semantic processing, highly imageable words are strongly supported by this level of processing in repetition tasks. In contrast, words of low imageability draw less support from semantic representations and relatively more from phonological representations of the words. The key outcome of this repetition span subtest is the presence of an imageabilty effect, better repetition of high image than low image words. If this difference is large, it indicates phonological processing of words is impaired and not providing the support needed to repeat or hold low image words in STM. At the same time, it indicates that semantic processing is strong enough to support repetition of high image words. If there is no strong imageabilty effect, it suggests that phonological ability is strong enough to support repetition and span for words of both high and low imageability.

Stimuli: This subtest was adapted from a stimulus set used in Martin and Saffran (1997). It consisted of sequences of 1–6 words in length. Word strings were made up of one and two-syllable words varied by lexical-semantic characteristics of frequency (F) and imageability (I) in four ways: HF-HI, LF-HI, HF-LI, and LF-LI. There were five sequences for each frequency-imageability type at each sequence length. The Kucera & Francis (1982) database was used to classify frequency: High frequency = > 25 occurrences per million and Low frequency = < 25 occurrences per million. We used the Paivio, Yuille, & Madigan (1968) database to classify the imageability of words. Those with ratings higher than 4.97 (the mean rating on a scale of 1–7) were classified as high imageability and those less than 4.97 were labeled as low imageability.

Procedure: Participants heard the sequence of words presented in E-Prime and attempted to repeat it immediately in serial order. Words were digitally recorded and presented in sequences programmed with E-Prime software. The five word strings for each frequency-imageabilty type were presented in random order within each span length condition.

Outcome variable: Two spans were calculated, one based on recall of words in serial order and the other based on recall of words in any order. Span size was calculated using Shelton, Martin and Yaffee's (1992) formula: Sequence length at which at least 50% of the sequences were recalled + .50 (proportion of sequences recalled in the next sequence length) for each frequency-imageability word type. Because the frequency-imageability word string types were presented in random order within a string length condition, it was necessary to proceed through increasing string lengths until criterion for span limit was reached for all four frequency-imageability types. If criterion had been reached for one type (e.g., low frequency-low imageability) before another type (e.g., high frequency-high imageability), those string types were still administered at increasing string lengths, but span for a frequency-imageability type was calculated based on the point where criterion had been reached.

Word-nonword span

Purpose: This subtest assesses the ability to repeat single words and single nonwords. Depending on the accuracy and types of errors, it is sensitive to deficits of input and output processing.

Stimuli: The word span measure included 10 lists of high imageability-high frequency words in each of four string length conditions (two to five words). All words were 1–2 syllables long. Nonword span stimuli (string lengths 1–5 nonwords) were derived from the items in the word span subtest by changing 2–3 phonemes, sampling equally from initial, medial and final positions of the word. All nonwords were phonotactically legal. Presentation of word and nonword span conditions was blocked. The sequence of items within a test string was the same for word and matching nonword strings. The order of test strings within each span length condition was randomized. Word and nonword stimuli were digitally recorded and presented in sequences programmed with E-Prime software.

Procedure: A participant listened to each word or nonword sequence presented via E-Prime software and is instructed to repeat the items immediately in the order in which they heard them.

Outcome variable: Two spans were calculated, one based on recall of words or nonwords in serial order and the other based on recall of words or nonwords in any order. Span size was calculated using Shelton, Martin and Yaffee's (1992) formula: Sequence length at which at least 50% of the sequences are recalled + .50 (proportion of sequences recalled in the next sequence length).

Probe span tests of identity STM, semantic STM and phonological STM

Purpose: This subtest is designed to probe the integrity of semantic STM and phonological STM via a probe-memory paradigm developed by Sternberg (1969; 1975). This task assesses short-term verbal recognition memory and depending on the condition, is sensitive to, phonological level recognition (rhyming), and/or semantic-level recognition (category membership).

Stimuli: There were two probe span subtests: Rhyming Probe to assess receptive phonological STM and Category Membership Probe to assess receptive semantic STM. Half of the probes were unrelated to any word in the string. The other half of the probe words were related in some way to one of the words in the string, depending on the condition: (1) Semantic probe - one word in the string was from the same category as the probe word, (2) Phonologic probe - one word in the string rhymed with the probe word. Sequences of words ranged from 1–7 for the Category Membership Probe and 1–7 for the Rhyming Probe. The number of syllables within sequence trials for a sequence length set was controlled and varied by no more than two syllables among test strings within a string length. Probe words were not used as distracter probes in any of the sequences. The distracter probes were repeated only minimally throughout the test and when they were repeated, it was not within the same sequence length set. Probe matches were sampled five times from each position in the string. Testing began at the one-item string length and proceeded until criteria for span were met (see below).

Procedure: The participant heard a sequence of words followed by a brief pause and a spoken probe word. The task was to judge whether the probe was related to one of the words in the string. Depending on the condition, participant's judgments were based on whether a probe was categorically related to or rhymed with a word in the string.

Outcome measure: Criteria for calculation of span on this probe memory task took into consideration that this was a recognition memory task rather than a recall task (as in the other span tasks). The chance of getting a correct response in this probe task was 50% (yes or no). Thus, criterion for reaching a span level was set at 50% above the chance of making a correct response (25%), thus increasing it to 75%. This criterion was set for both yes and no responses. If that level was achieved, testing continued to the next higher string length. If these criteria were not met, testing was stopped. Span was then calculated as the string length at which criteria were met plus the proportion of correct YES responses at the next higher string length divided by .75. As an example, if.75 of the YES and .75 of the NO responses were correct at string length four, and only .50 of the Yes responses were correct at string length five, span was equal to 4 plus .50/.75, or 4.67.

Analyses and Results

Descriptive statistics of the data from the TALSA subtests are provided below. These include the means, standard deviations, ranges, and medians of performance on the word processing and verbal STM measures of the TALSA tasks. There are several important features of these data, including the effects of the memory load manipulations (temporal interval conditions and item load) in Part 1 and the patterns of response to the verbal span measures in Part 2. These features are noted below.

Part 1: Word Processing Subtests with Unfilled and Filled Temporal Intervals between Stimulus and Response or Between Items Being Compared

1-sec UF interval

Participants with aphasia: At the 1-sec UF temporal interval accuracy levels of the subtests of single word and sentence level processing showed a gradual progression in task difficulty. Accuracy was near ceiling for Phoneme Discrimination for words (Table 2) with a median score of 1.00, but the range indicated that some people scored as low as .80 on this subtest suggesting that this subtest was sensitive to individuals with more severe input processing difficulties, as in, for example, Wernicke's aphasia. Rhyming Judgments were a little more challenging, with the median at .90 and a range of .60 – 1.00. Both input phonological processing subtests were more challenging when the stimuli to be compared were nonwords rather than words.

A similar pattern of easy and hard assessments of input lexical-semantic processing can be observed in Table 3. Accuracy on the Lexical Comprehension word-to-picture matching subtest, which measures access to semantics from the lexical form level, was overall quite high (mean .97, median 1.00), and yet, the range of scores indicates that for some individuals, this subtest was challenging. Performance is high on the Category Judgment task but less accurate than the Lexical Comprehension subtest overall. Whereas the Lexical Comprehension subtest requires confirming a match between a spoken word and one of four pictures, the Category Judgment task requires holding two words in working memory long enough to compare their semantic categories.

The subtests in Part 1 gradually become more difficult with tasks that involve both input and output processing (word and nonword repetition, Table 4), output processing (picture naming, Table 5) and sentence level processing (repetition, Table 6 and comprehension, Table 7).

Control participants: This group performed at a higher level overall and were at or near ceiling on phoneme discrimination, rhyming judgments, lexical comprehension, category judgments and word repetition. Lower levels of accuracy were observed at the 1-sec UF interval for subtests that also were more difficult for participants with aphasia including nonword repetition (mean = .88, range .67–1.00), comprehension of sentences with reversible distracters (mean = .93, range 0.66–1.00) and repetition of 'padded" sentences (mean = .94, range .83–1.00).

5-sec UF interval

Participants with aphasia: Tables 2–7 also include data from the 5-sec UF conditions for each of the language subtests in Part 1. For many of the subtests, proportions correct in the 5-sec UF interval suggested that there was little effect of this delay between stimuli to be compared or before a response was elicited on a task. Our statistical analysis confirmed this.

Control participants: As shown in Tables 2–7, control participants also showed minimal change in performance in the 1-sec UF and 5-sec UF interval conditions.

For the data from participants with aphasia, Wilcoxon Signed-Rank Tests were used to determine whether there were significant effects of the addition of a delay in the 8 TALSA subtests with interval manipulations. One second unfilled (1-sec UF) interval conditions were compared to five second unfilled (5-sec UF) interval conditions. The alpha level was determined by dividing .05 by the number of TALSA subtests with interval manipulations ($p = .05/8$), yielding an alpha level of .006. Full results are reported in Table 8. Median changes in accuracy are listed below for significant contrasts (or those that trend towards significance). Comparisons of accuracy in the 1-second filled delay and 5-second unfilled delay condition yielded the following statistically significant (or trends towards significant) changes in accuracy:

1. Nonword Repetition: decrease in accuracy ($Mdn = -0.07$) in the Nonword portion of the Word-Nonword Repetition subtest in the 5-sec UF interval condition compared to the 1-sec UF interval condition.

2. Sentence Repetition (Unpadded): trend towards a significant decrease in accuracy ($Mdn = -0.04$) in the Unpadded portion of the Sentence Repetition subtest in the 5-sec UF interval condition compared to the 1-sec UF interval condition.

3. Sentence Repetition (Padded): decrease in accuracy ($Mdn = -0.03$) in the Padded portion of the Sentence Repetition subtest in the 5-sec UF interval condition compared to the 1-sec UF interval condition.

All other contrasts were nonsignificant.

A closer look at the difference in performance in the 5-sec UF interval condition: Although the differences in 1-sec UF and 5-sec UF conditions were not significant for most of the tests in the first part of this battery, a closer look at the data from individual participants with aphasia indicated three patterns of change in proportion correct between the two conditions: an increase in accuracy, decrease in accuracy and no change. In Table 9, we show the data for those individuals whose accuracy increased or decreased in the 5-sec UF conditions including the mean changes, standard deviations and range of change.

The important feature of these results is the range of change in each direction. We would expect some fluctuation in performance that would not be significant, but some of the ranges suggest that increases and decreases in accuracy for some participants are significant and that there are two possible effects of time passage on language performance. Martin and Dell (2017) demonstrated this in an analysis of performance on the naming and word repetition subtests of the TALSA reported here. They first identified those individuals whose accuracy (measured as proportion correct) increased or decreased after the 5-sec UF interval and then determined whether these changes reflected a true difference using an estimate of the critical difference between the 1-sec and 5-sec conditions as twice the standard error of the difference of two proportions. They further argued that regardless of direction of change, the rate of individuals showing this degree of change should be greater than .05. The analysis showed significant changes for .46 of the participants on the naming task and .20 on the repetition task. In a second analysis, Martin and Dell (2017) observed that the magnitude of change in accuracy following a response delay in repetition was correlated with severity of semantic and phonological deficits and in naming, with severity of phonological deficit. Their analysis provides some evidence, at least for picture naming and word repetition, that a subset of individuals with aphasia in this study were sensitive to a response delay, with some showing improved accuracy and others showing decreased accuracy. We will offer some potential implications of this finding in the discussion.

5-sec F interval

Participants with aphasia: Tables 2–7 also show the proportions correct on these language measures after a 5-sec F interval. In this condition, the participant heard a stimulus which was followed by a five second interval in which numbers (1 to 9) appeared on the computer screen. The participant was asked to name the numbers. After five seconds, a second stimulus appeared (to be compared to the first stimulus before the interval) or a cue to respond appeared (e.g., in the naming or word and nonword repetition tests). As the data in Tables 2–7 show, average performance was worse on this condition than the 1-sec UF or 5-sec UF condition for participants with aphasia.

For the data from the participants with aphasia, we used Wilcoxon Signed-Rank Tests (Table 10) to determine whether there were significant effects of the addition of a filled interval in the eight TALSA subtests with interval manipulations for the participants with aphasia. Five-second unfilled delay conditions were compared to 5-second filled delay conditions. The alpha level was determined by dividing .05 by the number of TALSA subtests with interval manipulations (p = .05/8), yielding an alpha level of .006.

Full results of the Wilcoxon Signed-Rank Tests are listed in Table 10. Median changes in accuracy are listed below for significant contrasts (or those that trend towards significance). Comparisons of accuracy in the 5-sec F interval and 5-sec UF interval condition yielded the following statistically significant (or trends towards significant) changes in accuracy:

1. Phoneme discrimination Word: decrease in accuracy (*Mdn* = −0.05) in the 5-sec F delay condition compared to the 5-sec UF delay condition.
2. Phoneme discrimination Nonword: decrease in accuracy (*Mdn* = −0.10) in the 5-second filled delay condition compared to the 5-second unfilled delay condition.
3. Rhyme Judgment Word: decrease in accuracy (*Mdn* = −.05) in the 5-sec F delay condition compared to the 5-sec UF delay condition.
4. Rhyme Judgment Nonword: Trend towards a significant decrease in accuracy (*Mdn* = −0.05) in the 5 second filled delay condition compared to the 5-sec UF delay condition.
5. Lexical Comprehension: decrease in accuracy (*Mdn* = 0.00) in the 5-sec F delay condition compared to the 5-second unfilled delay condition.
6. Category Judgments (Picture): decrease in accuracy (*Mdn* = −0.05) in the 5-sec F delay condition compared to the 5-sec UF delay condition.
7. Category Judgments (Word): decrease in accuracy (*Mdn* = −.15) in the 5-sec F delay condition compared to the 5-sec UF delay condition.

8. Word Repetition: decrease in accuracy (*Mdn* = −0.13) in the 5-sec F delay condition as compared to the 5-sec UF delay condition.
9. Nonword Repetition: decrease in accuracy (*Mdn* = −0.20) in the 5-sec F delay condition as compared to the 5-sec UF delay condition.
10. Sentence Repetition (Unpadded): decrease in accuracy (*Mdn* = −0.10) in the in the 5-sec F delay condition compared to the 5-sec UF delay condition.
11. Sentence Repetition (Padded): decrease in accuracy (*Mdn* = −0.07) in the 5-sec F delay condition compared to the 5-sec UF delay condition.

The remaining contrasts within Rhyme Judgment, Picture Naming, and Sentence Comprehension subtests were not significant.

Control participants: For the control group, decrements in performance in the 5-sec F interval condition are present, but for most tests, the decline is less than .04. Two exceptions are the padded condition of the sentence repetition task (.89 correct in the 5-sec UF interval condition dropping to .80 in the 5-sec F condition and nonword repetition (.83 in the 5-sec UF interval condition dropping to .63 in the 5-sec F condition). Wilcoxin signed rank tests for these two sets of data showed notable decreases in accuracy on these two tests: Sentence Repetition (Padded), a trend of a significant decrease in accuracy (*Mdn* = −0.08) in the 5-sec F interval condition compared to the 5 sec UF interval condition, $z = -2.61$, $p = .009$; Nonword Repetition, significant decrease in accuracy (*Mdn* = −0.13) in the 5-sec F interval condition compared to the 5-sec UF interval condition, $z = -2.77$, $p < .006$.

These results indicate that 5-sec F interval condition was clearly more difficult than the other two, especially for the participants with aphasia. The reasons for this difficulty are yet to be determined. The filled interval condition created a dual-task situation when performing these language tasks. The participant had to maintain activation of a word, nonword, or sentence for five seconds and at the same time produce names of numbers appearing on the computer screen. Although accuracy of that naming task was not considered, the participant was still engaged in two tasks. In the General Discussion, we will discuss current efforts to determine whether executive processes are engaged in this condition in addition to the increased memory load imposed by the time interval.

Part 2. Tests of Lexical-Semantic and Phonological Processing Under High and Low Memory Load Conditions

The Rhyming Triplet Judgments and Synonymy Triplet Judgments assessed semantic and phonological processing ability under high and low working memory load conditions. These conditions were created by requiring the identification of two of three words that rhyme or two of three words that have similar meanings. As Table 11 shows, performance on both subtests became worse in the high memory load conditions when the choice of rhyming or synonymous words needed to be made among three words rather than determining which of two words rhymed or was synonymous with a third word.

Control participants: This group performed near ceiling in both of these tests, but also showed a mild decrement in accuracy (−.04 in each test) in the high memory load condition.

Martin, Kohen, Kalinyak-Fliszar, Soveri and Laine (2012) showed that the difference in accuracy was significant, even when controlling for chance performance. We will discuss the diagnostic significance of this outcome further in the General Discussion.

Part 3. STM Span tasks with linguistic variations

The results from four verbal span tasks are shown in Table 12, including means, standard deviations, ranges and medians for participants with aphasia and control participants. Outcomes for each span measure are reported below.

Digit and word spans—Data from 36 individuals with aphasia indicate that digit span is greater than word span, but span for both stimulus types fall between three and four items, on average. Repetition span is somewhat greater than pointing span for both digits and words, although again, spans using each task fall between three and four items.

Control participants: For this group, digit span is greater (between six and seven) than word span (between five and six) regardless of task type, and there is no overall difference for repetition vs. pointing span.

Repetition span for words varied for frequency and imageability—For participants with aphasia, span ranges between two and three items. It is highest when words are both high in frequency and imageability and lowest when words are low in frequency and imageability. That pattern is observed in the data from nine control participants as well, although their overall spans are greater, ranging between four and five items.

Wilcoxon Signed Rank Tests were used to determine whether there were significant effects of frequency and imageability on word span length on this subtest for individuals with aphasia. To obtain each participant's high imageability span, high frequency/high imageability and low frequency/high imageability spans were averaged together. To obtain each participant's low imageability span, high frequency/low imageability and low frequency/low imageability spans were averaged together. High frequency spans were also compared to low frequency spans in the same manner (by collapsing accuracy scores across the imageability variable). The alpha level was determined by dividing .05 by 4 (as four contrasts were performed for this subtest, 2 for individuals with aphasia and 2 for controls), yielding an alpha level of .0125. There was a statistically significant decrease in span length ($Mdn = -0.20$) in low imageability span ($Mdn = 2.25$) compared to high imageability span ($Mdn = 2.58$), $z = -3.58$, $p < .001$. Additionally, there was a statistically significant decrease in span length ($Mdn = -0.20$) in low frequency span ($Mdn = 2.20$) compared to high frequency span ($Mdn = 2.68$), $z = -3.43$, $p < .001$.

Control participants: In contrast to the individuals with aphasia, controls did not show significant differences between low imageability and high imageability span length, $z = -1.43$, $p = 0.153$, or between low frequency and high frequency span length, $z = -1.83$, $p = 0.068$.

Word and nonword repetition span—The data from this subtest indicate a clear difference in repetition span for words and nonwords for participants with aphasia (mean word span = 2.79, mean nonword span = 1.55) as well as for control participants (mean word span = 4.84, mean nonword span = 2.92). Wilcoxon Signed Rank Tests were used to determine whether there was a significant difference between word and nonword span length in the subtest. The alpha level was determined by dividing .05 by 2 (as two contrasts were performed for this subtest, one for individuals with aphasia and one for controls), yielding an alpha level of .025. There was a statistically significant decrease in span length ($Mdn = -1.20$) in nonword span ($Mdn = 1.40$) compared to word span length ($Mdn = 2.70$), $z = -4.49$, $p < .001$.

Control participants: Like the individuals with aphasia, controls showed a significant decrease in span length ($Mdn = -1.90$) in nonword span ($Mdn = 2.90$) compared to word span length ($Mdn = 5.00$), $z = -3.52$, $p < .001$.

Probe memory span—Participants with aphasia showed a wide range of ability on this task with average spans of 2.97 on the semantic probe span task and 3.22 on the phonological probe span task, but with spans as great as 7 items. Although not shown in the table, there were examples of individuals who showed contrasting semantic and phonological probe spans. For example, participant TUEL5 demonstrated a span of 2.89 on the category probe task and 6.99 on the rhyme probe span task. In contrast, VA6UT's span on the category probe task was 6.80 and 1.80 on the rhyme probe task. Across participants, there were many who showed modest to moderate differences between the two, in one direction or the other.

Control participants: The spans of control participants were considerably higher than those of the participants with aphasia (semantic probe, 6.10, and phonological probe, 6.70).

Measures of Reliability and Validity

Below we provide some limited evidence of reliability and validity of some of the subtests in this battery. Our reliability measure is limited to the internal consistency of items in a subset of subtests in Part 1 and Part 2. We also include some discussion of face validity, concurrent validity, and criterion validity. As we develop the clinical version of this battery, we will be obtaining these and other forms of reliability (e.g., test-retest) and validity.

Internal Consistency—To assess internal consistency, Chronbach's alpha was computed for the following subtests of Parts 1 and 2 of the TALSA: Phoneme Discrimination (1s unfilled, 5s unfilled, and 5s filled conditions), Rhyming Judgments (1s unfilled, 5s unfilled, and 5s filled conditions), Synonymy Triplet Judgments (2-choice and 3-choice), and Rhyming Triplet Judgments (2-choice and 3-choice). These subtests had two forms, A and B, with equal numbers of items, allowing us to compare accuracy proportions between the two forms to assess internal consistency. Chronbach's alpha values are reported in Table 13, and interpreted based on benchmarks from George and Mallery (2003). All but one subtest (Phoneme Discrimination in the 3 interval conditions) showed Acceptable, Good, or Excellent internal consistency. These outcomes will be considered in development of the clinical version of the TALSA battery, for which we will aim to achieve internal consistency for these subtests that is consistently in the good or excellent range.

Face Validity: The TALSA battery was designed to follow cognitive models of language processing (e.g., Patterson & Shewell, 1987) which specify component stages of word processing (e.g., semantic or phonological). This test battery offers an additional focus on processes that enable access, retrieval and short-term maintenance of activated linguistic representations which are emphasized in current processing-oriented models of word processing (e.g., Dell, et al., 1997; Hickok & Poeppel, 2000; Martin & Gupta, 2004).

Concurrent Validity: The TALSA battery's focus on effects of STM load on language processing is unique, making it difficult to find evidence of validity based on its similarity to other measures. We will seek such evidence in development of the clinical version for a limited set of subtests in the battery that do have counterparts in published tests, (e.g., digit span, (Wechsler, 2009) and lexical comprehension (Kay, Lesser & Coltheart,1992; Dunn & Dunn, 2007).

Criterion Validity: Subtests of the TALSA battery have been used in several studies as outcome measures following treatment. For example, Kalinyak et al. (2011) demonstrated improvement in verbal span abilities following a treatment to improve word and nonword repetition with increasing STM load. Additionally, the outcomes of subtests have been systematically related to working memory load. Martin et al. (2012) demonstrated that performance on synonymy and rhyming judgment tasks was adversely affected by manipulating the task format in a way that increased the working memory load.

General Discussion

This paper reports data from a new test battery to evaluate language and verbal short-term memory abilities and impairments in aphasia. The assessment battery is motivated by current thinking about the nature of aphasia as a linguistic processing disorder (as opposed to an impairment of linguistic representations). The data provided here are from the original version of this test battery, developed in the Aphasia Rehabilitation Research Laboratory, and this version represents the first stage in development of a clinical version. The outcomes of this study have relevance to both theoretical and clinical issues, and these are discussed below.

Theoretical Implications of This Study

Although the eventual goal of this project is to develop a clinical version of this test, the current version has provided some important data relevant to theories of the nature of aphasia. In a broad sense, the outcomes of these TALSA subtests serve as a test of theories that claim short-term retention processes are an integral part of word processing and that language impairment in aphasia is rooted in an impairment of the strength and/or maintenance of activation processes that support access to and retrieval of words. Evidence first comes from those subtests that add a 5 second interval to the language task either before a response is elicited (e.g., word and nonword repetition) or between two stimuli to be compared in a task (e.g., phoneme discrimination in word and nonword pairs).

Our early investigations of the effects of adding time intervals to a language task focused on the more common occurrence pattern of worse performance after a delay (e.g., Martin et al., 1996). Continued investigation of this effect, however, revealed evidence of two effects of time passage on language performance of some individuals with aphasia (e.g., Martin, 2006). The data from this test battery showed three patterns of performance in response to a time delay within a task or before a response: increased accuracy, decreased accuracy or no change. The analyses of the significance of these changes by Martin and Dell (2017) indicated that in a subset of cases, these changes were significantly greater than chance. This evidence justifies moving forward with further tests of the hypothesis that the two effects of a time delay on language performance could reflect impairments of two components of processes that support access and retrieval of words, activation strength and short-term maintenance of that activation.

We are only at the initial stages of understanding these two processing deficits in aphasia and how they affect language function. Although not all individuals with aphasia experience change in language performance in response to a time delay, some do, especially in more severe cases of semantic or phonological impairment. The same can be said for occurrence of different patterns of speech errors that are observed across individuals with aphasia. Some people exhibit high rates of semantic or phonological errors in naming or repetition and others do not. Research has led to a greater understanding of why there are different error profiles observed across cases of aphasia and it is likely that with a greater understanding of the nature of the processing deficits in aphasia, we will also gain an understanding of the factors that affect timing and maintenance of activation processes that support production, comprehension, repetition and verbal STM.

Some Clinical Implications of This Study

The clinical applications of this version of the TALSA are somewhat limited because the test is quite long. We are developing a clinical version which will be considerably shorter, but will include all the features that are unique to the current version, including the temporal and working memory load manipulations and the verbal span tasks. Nonetheless, the current version still has clinical value, as it is not necessary to administer the test in its entirety. Individual subtests can be administered as in-depth probes of areas of language function for mild and severe aphasia as well as assessment of verbal STM span capacity. Or, the effects of increased memory load can be assessed informally by adding a response delay when testing a few items from each of the subtests that vary memory load. Thus, in its current form, the TALSA is useful as a secondary test to be used following an initial diagnosis when there is some evidence of an impairment in an area of language function. Additionally, as will be described in more detail below, the verbal span tasks and test conditions with added memory load are especially useful in diagnosing the source of a language impairment in mild aphasia.

Specificity and Sensitivity: Two Features that Define the Diagnostic Value of the TALSA

The subtests that comprise the TALSA battery are designed to probe levels of linguistic representation (e.g., phonological and semantic representations of words) as in some other test batteries for aphasia, such as the Psycholinguistic Analysis of Language Processing in

Aphasia (PALPA) and the Comprehensive Aphasia Test. And yet the TALSA offers some unique features related to temporal aspects of language processing and effects of memory load on production and comprehension of words and sentences. With these additions, the TALSA battery can provide an assessment of aphasia that is more precisely in line with current views of aphasia as a linguistic processing disorder and guide development of treatment approaches that directly address the linguistic and processing aspects of the communication disorder.

There are two concepts that capture the potential clinical and research uses of this test battery. The first is *specificity* of its assessment of aphasia. The subtests of this test battery probe phonological and lexical-semantic levels of linguistic representation in input as well as output processing. Part 1 subtests, which incorporate a 5-second unfilled interval into language tasks, provide information on the nature and severity of the processing deficit that affects comprehension and/or production of words and sentences (slow processing or too-fast decay of activation). Additionally, the 5-second filled interval in Part 1 subtests introduces a dual task component into the language tasks, and consequently may reveal evidence of an executive processing disorder in addition to language impairment. This is also true of the synonymy and rhyming triplets in Part 2, which vary the working memory load in the judgment tasks. Although these possibilities require further investigation, the data from this test battery reported here (also Martin et al., 2012) suggest that sensitivity to dual task or increased working memory load conditions may serve as diagnostic markers of executive function difficulty.

The second concept that guided the development of this assessment battery for aphasia is *sensitivity* to language impairments at all levels of severity. The subtests in Part 1 include somewhat standard tasks of phonological and semantic input and output processing that vary in difficulty. The Phoneme Discrimination and Lexical Comprehension subtests are relatively easy and many of our participants with aphasia were near ceiling in completing these tasks. However, the exceptions were those individuals with more severe input processing difficulties (e.g., those with Wernicke’s aphasia). Thus, these tasks were easy for most of the participants, but were sensitive to the more severe language impairments and provided evidence regarding a participant’s ability to access to semantic and phonological representations. Other subtests in Part 1, which by nature are more challenging (e.g., nonword repetition), were difficult for most participants, even those with mild aphasia. Also, the addition of a 5 second interval to the language tasks reduced accuracy for some participants (but increased accuracy for others).

Parts 2 and 3 of the TALSA battery are especially sensitive to language impairments in milder aphasia. The Synonymy Triplets and Rhyming Triplets, which vary the working memory load, reveal impairments of semantic or phonological processing that are only evident when working memory load is increased (Martin et al., 2012). This finding is clinically significant, as it indicates that the tasks we use to assess language ability can play a critical role in revealing the presence of a semantic or phonological impairment.

The subtests of Part 3 of the TALSA battery provide a unique assessment of verbal span in aphasia that provides valuable diagnostic information beyond assessment of span capacity.

By varying stimulus characteristics in each task to be sensitive to specific levels of linguistic representation, the test is sensitive enough to identify the type of language impairment even in individuals whose language impairment is so mild, that it is not detected in standard language tests (e.g., the Western Aphasia Battery, WAB). To further explore the sensitivity of the TALSA to language impairment in mild aphasia, we identified two participants with aphasia (TUEL5 and TUFD26) who scored above the cut-off score of the WAB AQ of 93.8 and examined their scores relative to control data from the TALSA. Table 14 includes scores from all subtests (those subtests that impose delays are reported in the 5-sec UF condition), and highlights the subtests in which participants were ≥ 2 standard deviations below the control mean. We determined this criterion based on the logic of Brookshire, Wilson, Nadeau, Gonzales Rothi, and Kendall (2014), who used this criterion to determine the prevalence of alexia in their group of individuals with aphasia based on its previous use in a related study (Rapscak et al., 2009, as cited by Brookshire et al., 2014) and the Comprehension Aphasia Test, a theoretically-driven language processing test (Swinburn et al., 2005, as cited by Brookshire et al., 2014).

Participant TUEL5 (AQ = 94.3) demonstrated impaired performance on a subset of TALSA subtests in Parts 1 and 2, and all but one subtest in Part 3 (Probe Memory Span – Phonological), showing that the span test measures are particularly sensitive to milder language and short-term memory impairments. Participant TUFD26 (AQ = 95) demonstrated performance comparable to controls on all but two subtests of Parts 2 and 3 (Category Judgments – Pictures and Rhyming Triplets – 2-Choice Format), but broke down on most the span tasks in Part 3, again revealing the sensitivity of the TALSA span tasks to milder impairments.

Relevance of the TALSA Battery to Functional Communication Disorders

The TALSA battery is based on a model of language processing that acknowledges the role of STM in processing language. Does the assessment of the verbal STM component of language processing or effects of increased memory load on language performance have relevance to functional communication? Everyday conversations with others are filled with moments when one must ‘hold onto’ the utterances they want to contribute to a conversation until other speakers are finished. Moreover, there are frequent interruptions in the flow of a conversation. These delays and interruptions of our narrative contribution to a conversation require maintaining activation of the words we are ready to use when the moment to speak comes about. If we can improve a basic capacity of maintaining activation of word representations for longer periods as well as the ability to tolerate increases in memory load within language processing, these abilities are likely to transfer to language function more generally.

There is a genuine need for a means to assess non-linguistic influences such as STM on language function, as communication in everyday situations involves variations in processing time and fluctuations in memory load. Thus, the outcomes of this project should have a substantial impact on clinical practice and aphasia rehabilitation research by providing a tool to evaluate language impairment in aphasia from a processing perspective.

Limitations of the Present Study

We have presented a large set of data on an initial version of a test battery designed to assess word and sentence processing abilities of individuals with aphasia under conditions of added memory load. Although great effort was made to control for many variables when developing the stimuli for this test, we chose to focus primarily on variables that were critical to the level of linguistic representation and modality (input or output processing). Thus, for tasks such as rhyming, we focused less on controlling for word frequency and more on identifying pairs of words that varied orthography. We identified shortcomings such as these as we developed this initial version and are in the process of modifying these aspects in a clinical version of this test.

One change that is worth mentioning is being made for the Probe Memory Span measures of semantic STM and phonological STM. A limitation of that test is that the relative performances on the semantic and phonological span tasks can be compared between participants, but comparisons of the two conditions within a participant are confounded by possible differences in difficulty of the two tests. We are currently expanding this probe span test to include three semantic STM probe spans and three phonological STM probe spans that vary in difficulty. From these, we will establish composite probe span scores for semantic and phonological STM. We hypothesize that these composite span scores will be better estimates of phonological and semantic STM because they will be based on recall tasks with items that vary in difficulty.

Availability of the TALSA Battery

Although we are in the process of developing a clinical version of the TALSA, the original version described in this paper is available for use. Administration of the TALSA requires E-prime software (E-Prime 2.0, 2012, *Psychology Software Tools, Pittsburgh, PA. Requests to obtain the TALSA battery should be sent to saffranlab@temple.edu or the first author,* nmartin@temple.edu.

Conclusions and Future Directions

The TALSA battery is proving to be effective in determining how language processing impairments in aphasia affect access to and retrieval of semantic, lexical and phonological representations of words. Results of the memory load variations indicate that language performance can be affected in ways that can improve or impair accuracy of language performance in specific language tasks. The test battery has proved useful to our research on theoretical issues (e.g., Martin et al., 2012) but also as a pre- and post-testing instrument for our treatment studies. As noted, we are currently developing a clinical version of this test and plan to test its efficacy within a clinical setting. One of our aims in developing that version, will be to formalize the features of diagnostic specificity and sensitivity to severity of impairment that we aimed to incorporate in the present version of the TALSA battery.

Another important direction that we will follow relates to the relevance of the data reported here to functional communication abilities. We are currently implementing a treatment protocol that addresses the effects of adding a time delay to language tasks. One future goal is to further define the connection between the effects of temporal delay on language performance observed in the subtests of the TALSA and difficulties that individuals with aphasia experience in the context of functional communication situations. This understanding would guide development of treatment approaches that focus on problems of needing more time to retrieve words and sentences or maintain activation of representations long enough to effectively participate in functional communication settings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this paper was supported by National Institute on Deafness and other Communication Disorders Center of the National Institutes of Health under award number R21DC 008782 and R01DC013196. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We are very grateful to the participants who cheerfully and patiently participated in this study. Special thanks go to Jamie Reilly who provided invaluable advice in the early stages of development of this test battery. We also thank the research personnel who diligently helped with data collection and technical implementation of the test battery including Melissa Correa, Nick Gruberg, Amanda Concha, Samantha Rosenberg, Rebecca Afman, Drew Tipton, Beth Levitan, Kate Schmitt, Rebecca Berkowitz, Meghan McCluskey, Dana Roberts, Ariel Hegedus, Rachel Kamen, Andrew Demarco, Ferenc Bunta, Thomas Seminack, Bridget Parkes, Amelia Wisniewski-Barker, Julie Schlesinger, Kevin McCaffery, Sarah Meehan as well as Jessica Obermeyer for her assistance in manuscript preparation.

References

Abbate MS, La Chappelle NB. Pictures, please! A language supplement. Tucson, AZ: Communication Skill Builders; 1984a.

Abbate MS, La Chappelle NB. Pictures, please! An articulation supplement. Tucson, AZ: Communication Skill Builders; 1984b.

Baddeley AD. Working memory. New York, NY: Oxford University Press;

Baddeley A. The episodic buffer: a new component of working memory? Trends in Cognitive Science. 2000; 4(11):417–423.

Baddeley AD, Hitch G. Working memory. Psychology of learning and motivation. 1974; 8:47–89.

Bemdt RS, Mitchum CC. Auditory and lexical information sources in immediate recall: Evidence from a patient with deficit to the phonological short-term store. In: Vallar G, Shallice T, editorsNeuropsychological impairments of short-term memory. Cambridge, UK: Cambridge University Press; 1990. 115–144.

Brener R. An experimental investigation of memory span. Journal of Experimental Psychology. 1940; 26(5):467–492.

Brooks JO III, Watkins MJ. Further evidence of the intricacy of memory span. Journal of Experimental Psychology: Learning, Memory and Cognition. 1990; 16(6):1134–1141.

Brookshire CE, Wilson JP, Nadeau SE, Gonzalez Rothi LJ, Kendall DL. Frequency, nature, and predictors of alexia in a convenience sample of individuals with chronic aphasia. Aphasiology. 2014; 28:1464–1480. DOI: 10.1080/02687038.2014.945389

Clark JM, Paivio A. Extensions of the Paivio, Yuille, and Madigan (1968) norms. Behavior Research Methods. 2004; 36(3):371–383.

Conrad R, Hull AJ. Information, acoustic confusion and memory span. British Journal of Psychology. 1964; 55(4):429–432. [PubMed: 14237884]

Conway ARA, Engle RW. Working memory and retrieval: A resource-dependent inhibition model. Journal of Experimental Psychology: General. 1994; 123:354–373. [PubMed: 7996121]

Corsi PM. PhD. McGill University; 1972. Human memory and the medial temporal region of the brain.

Cowan N. What are the differences between long-term, short-term, and working memory? Progress in Brain Research. 2008; 169:323–338. [PubMed: 18394484]

Craik FI, Lockhart RS. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior. 1972; 11(6):671–684.

Darley FL. Aphasia. Philadelphia, PA: W.B. Saunders; 1982.

Dell GS, O'Seaghdha PG. Stages of lexical access in language production. Cognition. 1992; 42(1): 287–314. [PubMed: 1582160]

Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. Psychological Review. 1997; 104(4):801. [PubMed: 9337631]

Dunn LM, Dunn DM. Peabody picture vocabulary test. 4. Minneapolis, MN: Pearson Assessments; 2007. Manual

Engle RW. Working memory capacity as executive attention. Current Directions in Psychological Science. 2002; 11(1):19–23.

Francis W, Kučera H. Frequency analysis of English usage: Lexicon and Grammar. Boston, MA: Houghton Mifflin; 1982.

Friedrich FJ. Multiple phonological representations and verbal short-term memory. In: Shallice T, Vallar G, editorsNeuropsychological impairments of short-term memory. Cambridge, UK: Cambridge University Press; 1990. 74–93.

George D, Mallery P. SPSS for Windows step by step: A simple guide and reference. 11.0 update. 4. Boston: Allyn & Bacon; 2003.

Goodglass H, Kaplan E, Barresi B. Boston Diagnostic Aphasia Examination. 3. Austin, TX: Pro-Ed; 2001.

Grashey H. Über Aphasie und ihre Beziehung zur Wahrnehmung. Archiv fur Psychiatrie und Nervenkrankheiten. 1885; 16:654–688. (translation: R.de Bleser. On aphasia and its relation to perception, *Cognitive Neuropsychology, 6*,515–546, 1985).

Hickok G, Poeppel D. Towards a Functional Anatomy of Speech Perception. Trends in Cognitive Sciences. 2000; 4:131–138. [PubMed: 10740277]

Howard D, Franklin S. Missing the meaning? A cognitive neuropsychological study of the processing of words by an aphasic patient. Cambridge, MA: MIT Press; 1988.

Hula WD, McNeil MR. Models of attention and dual-task performance as explanatory constructs in aphasia. Seminars in Speech and Language. 2008; 29(3):169–187. [PubMed: 18720315]

Hulme C, Maughan S, Brown GD. Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. Journal of Memory and Language. 1991; 30(6):685–701.

Hulme C, Roodenrys S, Schweickert R, Brown GD, Martin S, Stuart G. Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. Journal of Experimental Psychology-Learning Memory and Cognition. 1997; 23(5):1217–1232.

Kalinyak-Fliszar M, Kohen F, Martin N. Remediation of language processing in aphasia: Improving activation and maintenance of linguistic representations in (verbal) short-term memory. Aphasiology. 2011; 25(10):1095–1131. [PubMed: 22791930]

Kane MJ, Engle RW. Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. Journal of Experimental Psychology: General. 2003; 132(1):47. [PubMed: 12656297]

Kay J, Lesser R, Coltheart M. Psycholinguistic Assessments of Language Processing in Aphasia. Hove, UK: Lawrence Erlbaum Associates Ltd; 1992.

Kertesz A. Western Aphasia Battery-Revised (WAB-R). Austin, TX: Pro-ed; 2006.

Kroll JF, Merves JS. Lexical access for concrete and abstract words. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1986; 12:92–107.

Kučera H, Francis W. Computational analysis of present-day English. Providence, RI: Brown University Press; 1967.

Martin N, Gagnon D, Schwartz MF, Dell GS, Saffran EM. Phonological facilitation of semantic errors in normal and aphasic speakers. Language and Cognitive Processes. 1996; 11(3):257–282.

Martin N. Repetition in aphasia. Theoretical and clinical implications. In: Berndt R, editorHandbook of Neuropsychology 2nd ed., Language and its disorders. Vol. 3. Amsterdam: Elsevier; 2001. 136–156.

Martin N. Two effects of time passage on performance on phoneme discrimination judgments: Sometimes it helps sometimes it hurts. Brain and Language. 2006; 99:174–175.

Martin N, Ayala J. Measurements of auditory-verbal STM span in aphasia: Effects of item, task, and lexical impairment. Brain and Language. 2004; 89(3):464–483. [PubMed: 15120538]

Martin N, Bunta F. Effects of lexical processing on primacy effects in repetition of words and nonwords: Evidence from aphasia. Brain and Language. 2007; 103(1):183–184.

Martin N, Dell GS, Saffran EM, Schwartz MF. Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. Brain and Language. 1994; 47(4):609–660. [PubMed: 7859057]

Martin N, Gupta P. Exploring the relationship between word processing and verbal short-term memory: Evidence from associations and dissociations. Cognitive Neuropsychology. 2004; 21(2–4):213–228. [PubMed: 21038201]

Martin N, Kohen F, Kalinyak-Fliszar M. The processing approach to the assessment of language and verbal short-term memory abilities in aphasia. Poster presented at the annual meeting of Clinical Aphasiology; Isle of Palms, SC. 2010 May.

Martin N, Kohen F, Kalinyak-Fliszar M, Soveri A, Laine M. Effects of working memory load on processing of sounds and meanings of words in aphasia. Aphasiology. 2012; 26(3–4):462–493. [PubMed: 22544993]

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

Martin N, Saffran EM. A computational account of deep dysphasia: Evidence from a single case study. Brain and Language. 1992; 43(2):240–274. [PubMed: 1393522]

Martin N, Saffran EM. Repetition and verbal STM in transcortical sensory aphasia: A case study. Brain and Language. 1990; 39(2):254–288. [PubMed: 2224495]

Martin N, Saffran EM, Dell GS. Recovery in deep dysphasia: Evidence for a relation between auditory–verbal STM capacity and lexical errors in repetition. Brain and Language. 1996; 52(1): 83–113. [PubMed: 8741977]

Martin RC, Shelton JR, Yaffee LS. Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. Journal of Memory and Language. 1994; 33(1):83–111.

McNeil MR. The nature of aphasia in adults. In: Lass NJ, McReynolds L, Northern F, Yoder D, editorsSpeech, language and hearing. Vol. 2. Philadelphia, PA: W.B. Saunders; 1982. 692–740.

McNeil MR, Odell K, Tseng CH. Toward the integration of resource allocation into a general theory of aphasia. Clinical Aphasiology. 1991; 20:21–39.

McNeil MR, Pratt SR. Defining aphasia: Some theoretical and clinical implications of operating from a formal definition. Aphasiology. 2001; 15(10–11):901–911.

Minkina I, Martin N. Verbal short-term memory and receptive language impairment in aphasia. Poster presented at Science of Aphasia; Venice Lido. Sept 27, 2016;

Monsell S. Components of working memory underlying verbal skills: A “distributed capacities” view. In: Bouma H, Bouwhuis DG, editorsAttention and performance X: Control of language processes. Hillsdale, NJ: Erlbaum; 1984. 327–350.

Morton J. The logogen model and orthographic structure. In: Firth U, editorCognitive processes in spelling. London, UK: Academic Press; 1980. 117–133.

Murray LL. Cognitive treatments for aphasia: Should we and can we help attention and working memory problems? Medical Journal of Speech-Language Pathology. 2004; 12:xxi–xxxviii.

Paivio A, Yuille JC, Madigan SA. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology. 1968; 76(1–2):1–25.

Pastizzo MJ, Carbone RF. Spoken word frequency counts based on 1.6 million words in American English. Behavior Research Methods. 2007; 39(4):1025–1028. [PubMed: 18183922]

Patterson KE, Shewell C. Speak and spell: Dissociations and word-class effects. In: Coltheart M, Sartori S, Job R, editorsThe Cognitive Neuropsychology of Language. Hove, England: Lawrence Erlbaum Associates Ltd; 1987. 273–295.

Poirier M, Saint-Aubin J. Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. The Quarterly Journal of Experimental Psychology. 1995; 48(2):384–404. [PubMed: 7610273]

Poirier M, Saint-Aubin J, Mair A, Tehan G, Tolan A. Order recall in verbal short-term memory: The role of semantic networks. Memory & Cognition. 2015; 43(3):489–499. [PubMed: 25301565]

Potagas C, Kasselimis D, Evdokimidis I. Short-term and working memory impairments in aphasia. Neuropsychologia. 2011; 49(10):2874–2878. [PubMed: 21704643]

Potter MC. Very short-term conceptual memory. Memory & Cognition. 1993; 21(2):156–161. [PubMed: 8469123]

Psychology Software Tools, Inc. E-Prime 2.0. 2012. Retrieved from http://www.pstnet.com

Rapcsak SZ, Beeson PM, Henry ML, Leyden A, Kim E, Rising K, … Cho H. Phonological dyslexia and dysgraphia: Cognitive mechanisms and neural substrates. Cortex. 2009; 45:575–591. DOI: 10.1016/j.cortex.2008.04.006 [PubMed: 18625494]

Roach A, Schwartz MF, Martin N, Grewal RS, Brecher A. The Philadelphia naming test: scoring and rationale. Clinical Aphasiology. 1996; 24:121–133.

Saffran EM. Short-term memory impairment and language processing. In: Caramazza A, editorCognitive neuropsychology and psycholinguistic. Hillsdale, NJ: Erlbaum; 1990. 137–168.

Saffran EM, Marin OS. Immediate memory for word lists and sentences in a patient with deficient auditory short-term memory. Brain and Language. 1975; 2:420–433. [PubMed: 1218376]

Saffran EM, Martin N. Neuropsychological evidence for lexical involvement in short-term memory. In: Vallar G, Shallice T, editorsNeuropsychological impairments of short-term memory. New York, NY: Cambridge University Press; 1990. 145–166.

Saffran EM, Schwartz MF, Linebarger MC, Martin N, Bochetto P. Unpublished test battery. 1988. The Philadelphia comprehension battery.

Schneider W, Eschman A, Zuccolotto. E-Prime User's Guide. Psychology Software Tools, Inc; 2002.

Schwartz MF, Dell GS, Martin N, Gahl S, Sobel P. A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. Journal of Memory and language. 2006; 54(2): 228–264.

Shallice T, Vallar G. The impairment of auditory-verbal short-term storage. In: Vallar G, Shallice T, editorsNeuropsychological impairments of short-term memory. New York, NY: Cambridge University Press; 1990. 11–53.

Shelton JR, Martin RC, Yaffee LS. Investigating a verbal short-term memory deficit and its consequences for language processing. In: Margolin D, editorCognitive neuropsychology in clinical practice. New York, NY: Oxford University Press; 1992. 131–167.

Shulman HG. Similarity effects in short-term memory. Psychological Bulletin. 1971; 75(6):399–415.

Sternberg S. Memory scanning: New findings and current controversies. The Quarterly journal of experimental psychology. 1975; 27(1):1–32.

Sternberg S. Memory-scanning: Mental processes revealed by reaction-time experiments. American scientist. 1969; 57(4):421–457. [PubMed: 5360276]

Sternberg S. High-speed scanning in human memory. Science. 1966; 153(3736):652–654. [PubMed: 5939936]

Swinburn K, Porter G, Howard D. The Comprehensive Aphasia Test. Hove, UK: Psychology Press; 2005.

Vallar G, Baddeley AD. Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. Journal of Verbal Learning and Verbal Behavior. 1984; 23(2):151–161.

Warrington EK, Shallice T. The selective impairment of auditory verbal short-term memory. Brain. 1969; 92(4):885–896. [PubMed: 5364015]

Watkins OC, Watkins MJ. Serial recall and the modality effect: Effects of word frequency. Journal of Experimental Psychology: Human Learning and Memory. 1977; 3(6):712–718.

Wright HH, Fergadiotis G. Conceptualising and measuring working memory and its relationship to aphasia. Aphasiology. 2012; 26(3–4):258–278. [PubMed: 22639480]

Walker I, Hulme C. Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25(5):1256–1271.

Wechsler D. Wechsler memory scale - IV. San Antonio, TX: Pearson; 2009.

Highlights

1. A new approach to assessment of language and verbal short-term memory (STM) abilities in aphasia.
2. Normative data on measures of language and verbal STM abilities in aphasia.
3. Focus on aphasia as a language processing impairment.
4. Evidence for two processing impairments that affect language and verbal STM, slowed processing and too-fast decay of activation of linguistic representations.
5. Assessment effects of increased memory load on language performance.
6. Assesses language and verbal STM abilities in mild, moderate and severe aphasia.

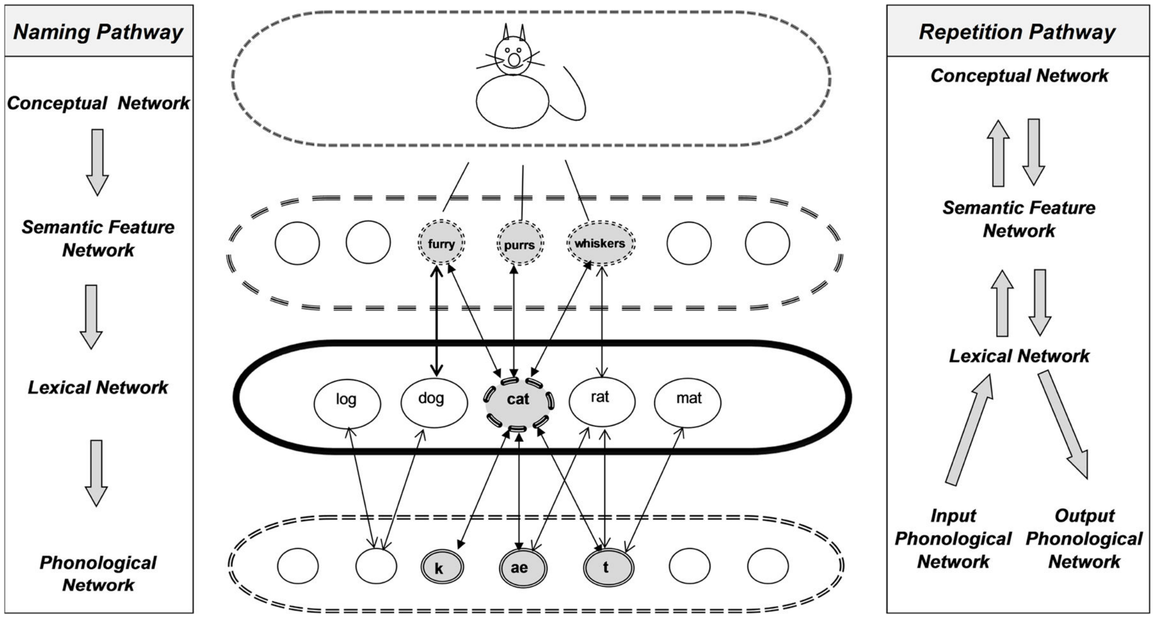

Figure 1.

Depiction of an interactive spreading activation model of word processing (center of figure) and depiction of the stages of input processing of words (left side of figure) and output processing of words (right side of figure).

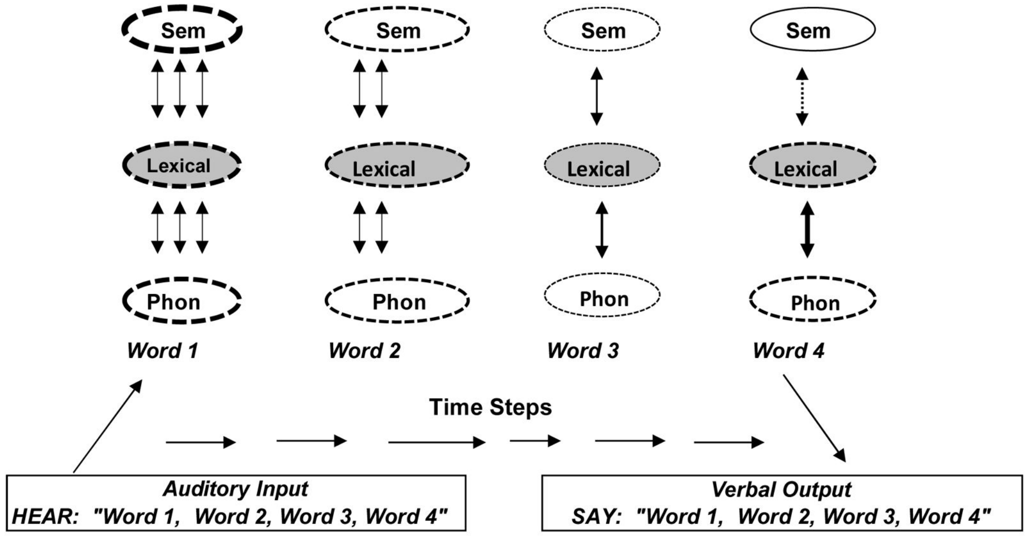

Figure 2.
Depiction of hypothesized activation of semantic, lexical and phonological representations of four words in a sequence in serial recall, at the point the words are retrieved for production. Dashed borders of semantic, lexical and phonological representations indicate their activation. The thickness of the dashed lines is intended to reflect the relative strength of that activation, which is hypothesized to increase as more time passes and interactive activation process have had more time to build the activation strength of a representation. The number and weight of arrows between the semantic, phonological and lexical representations is also related to the relative strength of each word's activation level as time passes. Words that are activated early in the sequence receive more activation through feedforward-feedback activation as time passes and the sequence is encoded in short-term memory.

Table 1

Demographic information on participants with aphasia

Participant ID	Education	Sex	Age (Years + Months)	Time Post Onset (Months)	Etiology	WAB AQ	WAB Aphasia Classification	Handedness
TU-FS 1	12	F	53:8	11	LCVA[1]	70.6	Wernicke's	R
TUSX3	14	M	47:7	188	LCVA	92.8	Anomic	L
TUEH4	13	F	46: 1	64	LCVA	81.4	Broca's	R
TUEL5	13	F	49: 0	181	LCVA	94.3	Anomic	R
TU-DD6	16	M	60: 4	78	Left Frontal Parietal CVA	55.6	Broca's	R
TU-CT7	10	F	45: 8	133	Left temporoparietal	62.4	Conduction	R
TU-HN8	16	M	57: 6	108	Left thalamic CVA	93.4	Anomic	R
TU-XD9	16	M	58: 7	195	LCVA	47.2	Broca's	N/A
TU-MI10	19	F	55: 10	86	LCVA	71.5	Transcortical Sensory	N/A
TU-KX11	16	F	68: 0	73	LCVA	47.3	Wernicke's	R
TU-KL12	14	M	59 :1	29	Left Thalamic CVA	92.4	Anomic	R
TUCB14	14	F	52: 0	12	LCVA	N/A		N/A
TUEC15	17	F	54: 4	103	LCVA	83.5	Anomic	N/A
TU-TB16	12	F	40: 0	69	Left frontal parietal basal ganglia infarct	92.2	Anomic	R
TUXX17	10	M	65: 1	62	LCVA	65.6	Conduction	R
TU-IU19	17	M	64: 4	12	LCVA	82	Anomic	R
TU-SX20	16	M	73: 0	22	LCVA (MCA[2])	34.9	Conduction	N/A
TU-SL21	12	M	55: 3	109	Left CVA (parietal aneurysm)	89	Anomic	R
TU-QH22	18	M	57: 2	22	LCVA	84.9	Anomic	L Ambidextrous
TU-NH23	12	F	67: 2	46	Left MCA infarct	49.4	Broca's	R
TU-GI24	12	M	47: 2	93	LCVA	70	Anomic	R
TUEC25	18	F	61: 8	301	LCVA	62.5	Wernicke's	R
TUFD26	18	M	71: 8	15	LCVA	95	Anomic	R
TUKL27	14	F	34: 4	10	LCVA	93.3	Anomic	R
TUHI28	13	M	53: 4	21	LCVA	65.3	Conduction	R
TUUN29	17	M	72: 7	12	LCVA	33.8	Broca's	R
TUSC32	13	M	78:4	14	LCVA	N/A		R

Participant ID	Education	Sex	Age (Years + Months)	Time Post Onset (Months)	Etiology	WAB AQ	WAB Aphasia Classification	Handedness
TUKU33	15	M	67: 3	6	LCVA	90.5	Anomic	R
TU-LT34	16	F	32: 3	12	LCVA	88.6.	Conduction	R
TUUP35	14	M	48: 1	37	LCVA	88.4	Anomic	R
TUDC37	14	M	49: 5	23	LCVA	33.6	Broca’s	N/A
TUKM38	18	M	67: 8	193	LCVA	80.3	Transcortical Motor	R
TUCN39	10	M	49: 3	10	LCVA	76.3	Broca’s	R
VA1-FL	12	M	61: 0	102	LCVA	58.1	Broca’s	R
VA2-BI	12	M	78: 4	14	LCVA	57	Conduction	R
VA3-KC	14	M	44:6	82	LCVA	64.1	Wernicke’s	R
VA4-TB	12	M	52: 0	36	LCVA	66.7	Conduction	R
VA5-CM	10	M	46: 11	6	LCVA	89.3	Anomic	R
VA6-UT	14	M	52: 11	17	LCVA (MCA)	91	Anomic	N/A

[1] Left cerebrovascular accident

[2] Middle cerebral artery

Table 2

Tests of Input Phonological Processing: Mean Proportion Correct, with Standard Deviation (SD), Range and Median (Mdn) at Three Interval Conditions

			Participants with Aphasia (n=38)			Control Participants (n=16)		
			Interval Condition			Interval Condition		
Subtest			***1-sec UF***	***5-sec UF***	***5-sec F***	***1-sec UF***	***5-sec UF***	***5-sec F***
Phoneme Discrimination	*Words*	Mean	0.96	0.94	0.87	0.99	1.00	0.96
		SD	0.06	0.09	0.10	0.02	0.01	0.05
		Range	0.80–1.00	0.55–1.00	0.70–1.00	0.95–1.00	0.95–1.00	0.85–1.00
		Mdn	1	1	0.88	1.00	1.00	1.00
	Nonwords	Mean	0.93	0.91	0.81	1.00	0.98	0.95
		SD	0.08	0.08	0.12	0.01	0.03	0.05
		Range	0.70–1.00	0.75–1.00	0.55–0.95	0.95–1.00	0.90–1.00	0.85–1.00
		Mdn	0.95	0.95	0.85	1.00	1.00	0.95
Rhyme Judgments	*Words*	Mean	0.86	0.87	0.79	0.98	0.98	0.95
		SD	0.12	0.15	0.14	0.04	0.04	0.06
		Range	0.60–1.00	0.45–1.00	0.30–0.95	0.85–1.00	0.90–1.00	0.85–1.00
		Mdn	0.90	0.90	0.83	1.00	1.00	0.95
	Nonwords	Mean	0.84	0.82	0.76	0.96	0.95	0.96
		SD	0.15	0.15	0.16	0.05	0.05	0.05
		Range	0.35–1.00	0.45–1.00	0.40–1.00	0.85–1.00	0.85–1.00	0.85–1.00
		Mdn	0.90	0.90	0.80	0.98	0.95	0.98

Table 3

Tests of Input Lexical-Semantic Processing: Mean Proportion Correct, with Standard Deviation (SD), Range and Median (Mdn) at Three Interval Conditions

		Participants with Aphasia (LC n=39, CJ n=38)			Control Participants (n=17)		
		Interval Condition			Interval Condition		
Subtest		***1-sec UF***	***5-sec UF***	***5-sec F***	***1-sec UF***	***5-sec UF***	***5-sec F***
Lexical Comprehension (LC)	Mean	0.97	0.98	0.91	1.00	1.00	1.00
	SD	0.06	0.04	0.12	0.00	0.00	0.02
	Range	0.81–1.00	0.88–1.00	0.56–1.00	1.00–1.00	1.00–1.00	0.94–1.00
	Mdn	1.00	1.00	0.94	1.00	1.00	1.00
Category Judgments (CJ) Pictures	Mean	0.94	0.92	0.85	0.98	0.98	0.95
	SD	0.07	0.08	0.12	0.06	0.02	0.05
	Range	0.70–1.00	0.60–1.00	0.55–1.00	0.75–1.00	0.95–1.00	0.85–1.00
	Mdn	0.95	0.95	0.85	1.00	1.00	0.95
Category Judgments (CJ) Words	Mean	0.93	0.93	0.81	0.99	0.98	0.96
	SD	0.09	0.08	0.13	0.02	0.04	0.05
	Range	0.70–1.00	0.70–1.00	0.50–1.00	0.95–1.00	0.85–1.00	0.85–1.00
	Mdn	0.95	0.95	0.80	1.00	1.00	1.00

Table 4

Word and Nonword Repetition: Mean Proportion Correct, with Standard Deviation (SD), Range and Median (Mdn) at Three Interval Conditions

Condition		Participants with Aphasia (n=37)			Control Participants (n=16)		
		Interval Condition			Interval Condition		
		1-sec UF	*5-sec UF*	*5-sec F*	*1-sec UF*	*5-sec UF*	*5-sec F*
Word Repetition	Mean	0.83	0.81	0.67	1.00	0.99	0.98
	SD	0.20	0.24	0.26	0.02	0.03	0.05
	Range	0.33–1.00	0.00–1.00	0.00–1.00	0.93–1.00	0.93–1.00	0.80–1.00
	Mdn	0.93	0.87	0.67	1.00	1.00	1.00
Nonword Repetition	Mean	0.51	0.43	0.18	0.88	0.83	0.63
	SD	0.30	0.30	0.21	0.12	0.18	0.27
	Range	0.00–1.00	0.00–0.93	0.00–0.67	0.67–1.00	0.33–1.00	0.20–0.97
	Mdn	0.53	0.40	0.13	0.90	0.87	0.73

Table 5

Picture Naming: Mean Proportion Correct, with Standard Deviation (SD), Range and Median (Mdn) at Three Interval Conditions

	Participants with Aphasia (n=37)			Control Participants (n=17)		
	Interval Condition			Interval Condition		
	1-sec UF	***5-sec UF***	***5-sec F***	***1-sec UF***	***5-sec UF***	***5-sec F***
Mean	0.68	0.65	0.68	0.98	0.98	0.97
SD	0.30	0.3	0.28	0.03	0.02	0.04
Range	0.00–1.00	0.00–1.00	0.00–0.97	0.93–1.00	0.93–1.00	0.87–1.00
Mdn	0.77	0.73	0.73	1.00	1.00	1.00

Table 6

Sentence Repetition: Mean Proportion Correct, with Standard Deviation (SD), Range and Median (Mdn) at Three Interval Conditions

		Participants with Aphasia (n=35)			Control Participants (n=16)		
		Interval Condition			Interval Condition		
Condition		***1-sec UF***	***5-sec UF***	***5-sec F***	***1-sec UF***	***5-sec UF***	***5-sec F***
Unpadded	Mean	0.72	0.67	0.55	1.00	0.99	0.97
	SD	0.29	0.31	0.32	0.01	0.02	0.04
	Range	0.02–1.00	0.04–1.00	0.02–1.00	0.96–1.00	0.94–1.00	0.88–1.00
	Mdn	0.82	0.78	0.58	1.00	1.00	0.98
Padded	Mean	0.55	0.50	0.41	0.94	0.89	0.80
	SD	0.30	0.29	0.25	0.05	0.10	0.14
	Range	0.00–0.97	0.04–0.96	0.00–0.83	0.83–1.00	0.69–1.00	0.52–1.00
	Mdn	0.57	0.44	0.36	0.97	0.91	0.85

Table 7

Sentence Comprehension: Mean Proportion Correct with Standard Deviation (SD), Range and Median (Mdn) at Three Interval Conditions

		Participants with Aphasia (n=38)			Control Participants (n=16)		
		Interval Condition			Interval Condition		
Condition		***1-sec UF***	***5-sec UF***	***5-sec F***	***1-sec UF***	***5-sec UF***	***5-sec F***
Lexical Distracter	Mean	0.91	0.9	0.86	0.99	0.99	0.99
	SD	0.12	0.15	0.16	0.03	0.03	0.03
	Range	0.60–1.00	0.50–1.00	0.40–1.00	0.90–1.00	0.90–1.00	0.90–1.00
	Mdn	1.00	1.00	.90	1.00	1.00	1.00
Reverse Semantic Role Distracter	Mean	0.69	0.68	0.68	0.93	0.96	0.99
	SD	0.21	0.19	0.18	0.10	0.05	0.03
	Range	0.33–1.00	0.33–1.00	0.30–1.00	0.66–1.00	0.89–1.00	0.90–1.00
	Mdn	0.69	0.67	0.69	1.00	1.00	1.00

Table 8

Effects of 5-second Unfilled Interval: Wilcoxon Signed-Rank Test Results

TALSA Subtest	*z*	*p* (2-tailed)
Phoneme Discrimination - Words (n = 38)	−1.10	0.272
Phoneme Discrimination - Nonwords (n = 38)	−1.10	0.056
Rhyme Judgments - Words (n = 38)	−0.65	0.519
Rhyme Judgments - Nonwords (n = 38)	−0.34	0.732
Lexical Comprehension (n = 39)	−0.78	0.434
Category Judgments - Pictures (n = 38)	−1.64	0.101
Category Judgments - Words (n = 38)	−0.44	0.659
Word Repetition (n = 37)	−0.14	0.888
Nonword Repetition (n = 37)	−3.06	0.002*
Picture Naming (n = 37)	−1.66	0.098
Sentence Repetition - Unpadded (n = 38)	−2.667	0.008
Sentence Repetition - Padded (n = 38)	−2.902	0.004*
Sentence Comprehension - Lexical Distracter (n = 38)	−0.468	0.640
Sentence Comprehension - Reverse Role (n = 38)	−0.332	0.740

* $p < .006$ = alpha level

Table 9

Effects of 5-second Unfilled Delay on Performance on TALSA Subtests with Time Interval Manipulations

TALSA Subtest and Total Number of Participants	Statistic	Worse performance after delay	Better performance after delay
	n	14	8
Phoneme Discrimination - Words (n = 38)	*Mean change (SD)*	−0.09 (0.09)	0.07 (0.03)
	Range of change	−0.05 to −0.35	0.05 to 0.10
	n	20	9
Phoneme Discrimination - Nonwords (n = 38)	*Mean change (SD)*	−0.07 (0.03)	0.07 (0.05)
	Range of change	−0.05 to −0.15	0.05 to 0.20
	n	9	17
Rhyme Judgments - Words (n = 38)	*Mean change (SD)*	−0.12 (0.06)	0.08 (0.04)
	Range of change	−0.05 to −0.25	0.05 to 0.20
	n	18	18
Rhyme Judgments - Nonwords (n = 38)	*Mean change (SD)*	−0.21 (0.12)	0.18 (0.13)
	Range of change	−0.05 to −0.45	0.05 to 0.55
	n	6	7
Lexical Comprehension (n = 39)	*Mean change (SD)*	−0.07 (0.02)	0.11 (0.06)
	Range of change	−0.06 to −0.12	0.06 to 0.19
	n	16	11
Category Judgments - Pictures (n = 38)	*Mean change (SD)*	−0.08 (0.04)	0.06 (0.02)
	Range of change	−0.05 to −0.15	0.05 to 0.10
	n	14	11
Category Judgments - Words (n = 38)	*Mean change (SD)*	−0.08 (0.05)	0.10 (0.08)
	Range of change	−0.05 to −0.20	0.05 to 0.30
	n	13	15
Word Repetition (n = 37)	*Mean change (SD)*	−0.16 (0.15)	0.11 (0.08)
	Range of change	−0.07 to −0.53	0.07 to 0.27
	n	26	7
Nonword Repetition (n = 37)	*Mean change (SD)*	−0.17 (0.10)	0.17 (0.15)
	Range of change	−0.07 to −0.40	0.07 to 0.47
Picture Naming (n = 37)	*n*	21	12

TALSA Subtest and Total Number of Participants	Statistic	Worse performance after delay	Better performance after delay
	Mean change (SD)	−0.12 (0.07)	0.11 (0.08)
	Range of change	−0.03 to −0.30	0.03 to 0.30
	n	20	11
Sentence Repetition - Unpadded (n = 38)	*Mean change (SD)*	−0.06 (0.11)	0.04 (0.03)
	Range of change	−0.02 to −0.38	0.02 to 0.10
	n	22	10
Sentence Repetition - Padded (n = 38)	*Mean change (SD)*	−0.11 (0.09)	0.05 (0.04)
	Range of change	−0.02 to −0.36	0.02 to 0.16
	n	7	6
Sentence Comprehension - Lexical Distracter (n = 38)	*Mean change (SD)*	−0.16 (0.08)	0.13 (0.05)
	Range of change	−0.10 to −0.30	0.10 to 0.20
	n	15	13
Sentence Comprehension Distracter - Reverse Role (n = 38)	*Mean change (SD)*	−0.15 (0.05)	0.15 (0.05)
	Range of change	−0.10 to −0.23	0.10 to 0.23

Table 10

Effects of 5-second Filled Interval: Wilcoxon Signed-Rank Test Results

TALSA Subtest	*z*	*p* (2-tailed)
Phoneme Discrimination - Words (n = 38)	−2.87	0.004 *
Phoneme Discrimination - Nonwords (n = 38)	−4.78	<0 .001 *
Rhyme Judgments - Words (n = 38)	−3.99	<0 .001 *
Rhyme Judgments - Nonwords (n = 38)	−2.60	0.009
Lexical Comprehension (n = 39)	−3.36	0.001 *
Category Judgments - Pictures (n = 38)	−3.57	< 0.001 *
Category Judgments - Words (n = 38)	−4.43	<0 .001 *
Word Repetition (n = 37)	−3.99	< 0.001 *
Nonword Repetition (n = 37)	−5.04	<0.001 *
Picture Naming (n = 37)	−1.57	0.117
Sentence Repetition - Unpadded (n = 38)	−3.97	<0.001 *
Sentence Repetition - Padded (n = 38)	−3.88	<0.001 *
Sentence Comprehension - Lexical Distracter (n = 38)	−2.37	0.018
Sentence Comprehension - Reverse Role (n = 38)	−0.03	0.976

* $p < .006$ = alpha level

Table 11

Synonymy and Rhyming Triplet Judgments with Memory Load Variations: Mean Proportion Correct, with Standard Deviation (SD), Range and Median (Mdn)

Subtest		Participants with Aphasia (ST n=37, RT n=38)		Control Participants (n=17)	
		Interval Condition		Interval Condition	
		2-Choice Format	*3-Choice Format*	*2-Choice Format*	*3-Choice Format*
Synonymy Triplet Judgments	Mean	0.87	0.74	0.98	0.94
	SD	0.11	0.17	0.04	0.06
	Range	0.53–1.00	0.38–1.00	0.88–1.00	0.85–1.00
	Mdn	0.88	0.75	1.00	0.95
Rhyming Triplet Judgments	Mean	0.87	0.73	1.00	0.96
	SD	0.15	0.18	0.01	0.07
	Range	0.40–1.00	0.23–1.00	0.97–1.00	0.80–1.00
	Mdn	0.93	0.75	1.00	1.00

Table 12

Span Measures with Linguistic Variations: Mean Span with Standard Deviation (SD), Range and Median (Mdn)

a. Digit and Word Span

	Participants with Aphasia (n = 36)				Control Participants (n = 16)			
	Digits		Words		Digits		Words	
	Pointing	*Repetition*	*Pointing*	*Repetition*	*Pointing*	*Repetition*	*Pointing*	*Repetition*
Mean Span	3.40	3.78	3.06	3.17	6.69	6.69	5.60	5.60
SD	1.68	1.59	1.21	1.30	0.66	0.62	0.81	1.18
Range	1.00–6.80	1.20–7.00	1.00–5.50	1.00–6.00	4.60–7.00	4.60–7.00	4.00–7.00	3.00–7.00
Mdn	3.20	3.50	2.80	2.80	7.00	7.00	5.50	5.60

b. Repetition Span for Words Varied for Frequency and Imageability

	Participants with Aphasia (n = 36)				Control Participants (n = 9)			
	Frequency and Imageability Type				Frequency and Imageability Type			
	HF/HI	*HF/LI*	*LF/HI*	*LF/LI*	*HF/HI*	*HF/LI*	*LF/HI*	*LF/LI*
Mean Span	2.63	2.35	2.34	2.13	4.93	4.80	4.78	4.51
SD	1.3	1.37	1.17	1.31	0.75	0.90	0.76	0.76
Range	0.00–5.00	0.00–4.80	0.00–4.40	0.00–4.80	4.00–6.00	3.80–6.00	4.00–6.00	3.40–5.40
Mdn	2.80	2.40	2.40	2.00	5.00	5.00	4.80	4.80

c. Word and Nonword Repetition Span

	Participants with Aphasia (n = 36)		Control Participants (n = 16)	
	Word	**Nonword**	**Word**	**Nonword**
Mean Span	2.79	1.55	4.84	2.92
SD	1.14	1.00	0.24	0.68
Range	*0.50–5.00*	*0.20–4.40*	4.20–5.00	1.60–3.80
Mdn	*2.70*	*1.40*	5.00	2.90

d. Probe Memory Span

	Participants with Aphasia (n = 39)		Control Participants	
	Semantic	**Phonological**	**Semantic (n = 15)**	**Phonological (n = 16)**
Mean Span	2.97	3.22	6.10	6.70

d. Probe Memory Span

	Participants with Aphasia (n = 39)		Control Participants	
	Semantic	**Phonological**	**Semantic (n = 15)**	**Phonological (n = 16)**
SD	1.79	2.31	0.82	0.84
Range	0.00–7.00	0.00–7.00	4.64–7.00	3.73–7.00
Mdn	2.71	2.97	5.88	7.00

Note. a – c = spans based on recall in serial order and d = probe span (recognition recall).

Table 13

Internal Consistency: Chronbach's Alpha Values and Interpretation

Subtest	Condition	Cronbach's alpha	Interpretation
	1s Unfilled	0.52	Poor
Phoneme Discrimination	5s Unfilled	0.61	Questionable
	5s Filled	0.37	Unacceptable
	1s Unfilled	0.78	Acceptable
Rhyming Judgments	5s Unfilled	0.93	Excellent
	5s Filled	0.81	Good
Synonymy Triplet	2-Choice	0.86	Good
Judgments	3-Choice	0.8	Good
Rhyming Triplet	2-Choice	0.86	Good
Judgments	3-Choice	0.83	Good

Table 14

Sensitivity of the TALSA: TALSA Results for Participants with Western Aphasia Battery AQ Scores above 93.8

		TUEL5 (WAB = 94.3)	TUFD26 (WAB = 95)
Phoneme Descrimination (5s UF)	*Words*	0.95 *	1.00
	Nonwords	0.90 *	1.00
Rhyming Judgments (5s UF)	*Words*	1.00	1.00
	Nonwords	1.00	1.00
Lexical Comprehension (5s UF)		1.00	1.00
Category Judgments (5s UF)	*Pictures*	1.00	0.85 *
	Words	0.75 *	1.00
Word-Nonword Repetition (5s UF)	*Words*	0.87 *	1.00
	Nonwords	0.07 *	0.93
Picture Naming (5s UF)		0.87 *	0.97
Sentence Repetition (5s UF)	*Unpadded*	0.86 *	1.00
	Padded	0.44 *	0.96
Sentence Comprehension (5s UF)	*Lexical Distracter*	1.00	1.00
	Reversible Semantic Role Distracter	0.56 *	0.89
Synonymy Triplet Judgments	*2-Choice Format*	1.00	0.95
	3-Choice Format	1.00	0.95
Rhyming Triplets	*2-Choice Format*	0.97 *	0.97 *
	3-Choice Format	0.90	0.87
Digit and Word Span	*Digits Pointing*	3.40 *	5.20 *
	Digits Repetition	3.80 *	5.40 *
	Words Pointing	2.40 *	3.80 *
	Words Repetition	2.20 *	3.80
Repetition span for words varied for frequency and imageability	*HF/HI*	1.80 *	3.40 *
	HF/LI	2.00 *	3.00 *
	LF/HI	2.40 *	3.00 *
	LF/LI	2.00 *	3.40
Word and Nonword Repetition Span	*Words*	2.60 *	3.60 *
	Nonwords	1.40 *	2.40

		TUEL5 (WAB = 94.3)	TUFD26 (WAB = 95)
Probe Memory Span	*Semantic*	2.89*	4.64
	Phonological	6.99	7.00

*≥ 2 SDs below control performance