

# Construct Validity and Reliability of Working Memory Tasks for People with Aphasia

## INTRODUCTION

Working memory (WM) is a cognitive system that maintains activation of select information in the service of goal-directed behavior (Baddeley, 2003). Previous research suggests that WM contributes to language impairments in aphasia (e.g., Caspari et al., 1998; Friedman & Gvion, 2012). However, the WM measures used in studies of neurologically healthy adults typically involve verbal production and often use digits or letters, both of which may be susceptible to errors in people with aphasia (PWA). Thus, it is difficult to separate WM deficits from general language deficits in PWA. Tasks that have been used to measure WM in PWA include pointing tasks, n-back tasks (e.g., Christensen & Wright, 2010) and non-linguistic tasks such as block span (Lang & Quitz, 2010). To our knowledge, no study has established the construct validity of measures used with PWA by comparing performance on them to measures used more commonly with neurologically healthy populations.

The goals of this study were to (1) develop a battery of measures to assess WM in people across a range of aphasia severities and (2) establish the construct validity of those measures by comparing neurologically healthy adults' performance on the new battery to a well-established ("gold standard") measure of WM.

## EXPERIMENT ONE

Experiment 1 compared a battery of tests designed to measure WM in PWA to a "gold standard."

## METHODS

### **Participants**

Neurologically healthy younger (ages 18-30) and older adults (ages 60-85) participated (n=20 per group).

### **Materials**

This study used a composite score obtained by summing performance on three WM measures (see below) as the "*Gold-Standard*." Waters and Caplan (2003) reported that a composite score based on these measures had test re-test reliability of .85 in a study of neurologically healthy adults.

Alphabet Span: Participants repeated word lists after alphabetizing them.

Subtract 2: Participants repeated number lists after subtracting 2 from each digit.

Reading Span: Participants read sentence lists and repeated the last word of each sentence.

Participants progressed from sets of two to eight items on *Alphabet* and *Subtract 2*, and sets of six on *Reading span*.

The *Aphasia Battery* consisted of the following seven tasks:

Listening Span: Participants listened to sentence lists and repeated the last word of each sentence (adapted from Thompson, unpublished). Each sentence was ten syllables long. The final

words were high frequency, monosyllabic nouns. Participants progressed from sets of one to six sentences.

1-Back and 2-Back: Participants pressed a button to indicate when a picture (of a piece of fruit) was the same as one that occurred one or two items previously (Christensen & Wright, 2010). Both reaction time and percent correct data were collected.

Picture Span (Forward & Backward): The experimenter read a list of high frequency, monosyllabic nouns before displaying a sheet with 9 pictures. Participants pointed to the images in either forward or backwards order (depending on the version).

Square Span (Forward & Backward): Participants were presented with a 3x3 grid of black squares. The experimenter pointed to a series of squares that the participant replicated in either forwards or backwards order.

For picture and square span, testing progressed from sets of two to six items.

## **Procedures**

All participants completed all tasks over two sessions. Ten older adults returned an average of 19 weeks after their last session (range: 12-27 weeks) to examine test-retest reliability of the battery.

## **RESULTS & DISCUSSION**

Item scores were calculated as the total number of correct trials across all span levels. Span scores were also calculated. However, in the interests of brevity only item scores (which are more sensitive) and selected statistical analyses are reported here.

### **Construct Validity**

Correlations were computed to determine the relation between the new measures and the gold standard composite scores (see Table 1). Correlations greater than .80 were only obtained by summing two or more of the new measures.

### **Test-Retest Reliability**

Correlations were calculated for the 10 participants who completed the battery twice. Most of the measures showed test re-test reliability ranging from .73 to .89 ( $p$ 's < .01). There were three exceptions: for *Listening span*,  $r = .97$ ,  $p < .001$ , and for 1-back reaction times and 2-back accuracy,  $r$ 's < .39,  $p$ 's > .26. Similar to Waters and Caplan (2003), the composite gold standard score was more reliable than any of the individual measures,  $r = .92$ ,  $p < .001$ . Similarly, the new composite scores (see Table 1) showed test re-test reliability of .94 ( $p < .001$ ).

## **EXPERIMENT 2**

Experiment 2 piloted the aphasia battery in a diverse group of PWA to determine which of the tests provided the most feasible method of measuring WM, regardless of aphasia severity or the presence of apraxia of speech.

## METHODS

### Participants

Nine PWA participated (ages 39 - 68). The diagnosis of aphasia was independently confirmed with a comprehensive testing battery. Three PWA had significant apraxia of speech.

### Materials & Procedures

The materials and procedures were identical to those described above except that PWA (1) did not complete the “gold standard” battery and (2) tested to span on *Listening span*, meaning that they stopped after they correctly responded to fewer than 2/5 trials at any set size.

## RESULTS & DISCUSSION

PWA and nine age-matched controls' performance on the WM tasks is summarized in Table 2. Wilcoxon Sign Rank tests showed that PWA performed better on the square span than picture span tasks, in both forward and backward versions. Mann-Whitney U-tests showed that controls outperformed PWA on all tasks.

Listening span was the only task that presented significant difficulty for the PWA, likely because it required verbal production.

## GENERAL DISCUSSION

The results suggest that a composite score based on the picture span forward and backward tasks provides a reliable measure of WM in PWA that is highly correlated with the gold standard composite score.

## REFERENCES

- Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*, 189-208.
- Caspari, I., Parkinson, S. R., LaPointe, L. L., & Katz, R. C. (1998). Working memory and aphasia. *Brain and Cognition, 37*, 205-223.
- Christensen, S. C., & Wright, H. H. (2010). Verbal and non-verbal working memory in aphasia: what three *n*-back tasks reveal. *Aphasiology, 24*, 752-762.
- Lang, C. J., & Quitz, A. (2012). Verbal and nonverbal memory impairment in aphasia. *Journal of Neurology, 259*, 1655-1661.
- Waters, G. S., & Caplan, D. (2003). The reliability and stability of several working memory measures. *Behavior Research Methods, Instruments, & Computers, 35*, 550-564.

## TABLES

**Table 1: Correlations between the Gold Standard Composite score and new measures**

Task		Gold Standard Composite
Listening Span		0.73*
Picture Span	Forward	0.74*
	Back	0.72*
Square Span	Forward	0.59*
	Back	0.39*
1-Back	Accuracy	-0.08
	Reaction Time	-0.01
2-Back	Accuracy	0.45*
	Reaction Time	-0.11
<i>New Composite Scores</i>		
Picture Back + Picture Forward + 2-Back Acc		0.82*
Picture Back + Pictures Forward		0.81*

\*p < .05

**Table 2: Mean (Range) for PWA and Age-Matched Controls (n=9 per group)**

Task	PWA	Controls
Listening Span	7.3 (0 – 27)	75.8 (59 – 101)
Picture Span	Forward	113.8 (92-148)
	Backward	110.2 (88-152)
Square Span	Forward	119.7 (82-161)
	Backward	108.9 (81-125)
1-Back Accuracy	95.3 (87 – 100)	98.8 (93-100)
2-Back Accuracy	81.6 (77 – 86)	89.3 (81-98)
<i>New Composite Scores</i>		
Picture Back + Picture Forward + 2-Back Acc	167.8 (132 – 184)	313.3 (266-388)
Picture Back + Pictures Forward	83.2 (48 – 102)	224.0 (181-300)