doi: 10.2466/22.24.PMS.116.2.

# A PRELIMINARY ASSESSMENT OF THE RELIABILITY AND VALIDITY OF A COMPUTERIZED WORKING MEMORY TASK[1]

NICOLE LOVATO, LEON LACK,HELEN WRIGHT, EVA KEMPS,
MEGAN CANT, AND JOANNE HUMPHREYS

School of Psychology
Flinders University, Australia

Summary.—The aim of the current study was to evaluate the psychometric properties of a recently developed measure of working memory: the Double Span Task. The Double Span Task is the first experimental task designed to assess all three components of Baddeley's (1992) model of working memory. The reliability of the Double Span Task was assessed in a sample of 105 older adults (M age = 64.3 yr., SD = 6.4). The internal consistency and test-retest reliability of the Double Span Task were good. The validity of the Double Span Task was assessed using a different sample of 49 older adults (M age = 70.0 yr., SD = 9.3). Performance on the Double Span Task was positively correlated with performance on a well-established measure of working memory, the Letter Number Sequencing Task. The Double Span Task also showed good discriminant validity. The Double Span Task is a reliable and valid measure of all three components of the working memory system.

The concept of working memory has been shown to play an important role in many aspects of everyday cognition, such as problem solving (Gilhooly, Logie, Wetherick, & Wynn, 1993), reasoning (Klauer, Stegmaier, & Meiser, 1997), and comprehension (Just & Carpenter, 1992). Baddeley (1992, p.556) defines working memory as "a brain system that provides temporary storage and manipulation of information necessary for complex cognitive tasks such as language comprehension, learning, and reasoning." Baddeley's (1992) model of working memory comprises two slave systems: the phonological loop, which processes speech-based material, and the visuo-spatial sketchpad, which is involved in the storage and manipulation of visual-spatial information. The two slave systems are monitored and coordinated by the central executive, an attentional-control system involved in response selection and inhibition (Baddeley, 1992; Martein, Kemps, & Vandierendonck, 1999). Baddeley (2000) has recently proposed a third slave system, the episodic buffer. The episodic buffer is proposed to integrate visual, spatial, and verbal information.

Although there are a growing number of studies investigating working memory performance, many of the available objective measures for doing so are not capable of measuring each component of the working memory system simultaneously (Martein, et al., 1999). Given the complex

[1]Address correspondence to Dr. Nicole Lovato, School of Psychology, Flinders University, GPO Box 2100, Adelaide, S.A., Australia, 5001 or email (nicole.lovato@flinders.edu.au).

DOI 10.2466/22.24.PMS.116.2.____ ISSN 0031-5125

interplay between the separate slave systems necessary to constitute effective working memory performance, conclusions derived from these studies are incomplete and potentially misleading. The primary aim of the current study was, therefore, to validate a recently developed measure of general working memory, the Double Span Task (DST), which is hypothesized to assess each component of working memory simultaneously.

Based on an initial presentation by Loisy and Roulin (1992) and later developed (Loisy & Roulin, 2003), Martein and colleagues (1999) were the first to publish an experimental task that can assess all components of the working memory model at the same time. Well-established measures of working memory, such as the Letter Number Sequencing Task and the N-back (also referred to as the modified lag task in the literature), are capable of assessing discrete components of the working memory process (Martein, et al. 1999). Previously used measures of working memory assess the central executive and one of the slave systems, typically the phonological loop. For example, the N-back task presents participants with a sequence of stimuli to which they are asked to indicate when a current stimulus matches one from n steps earlier (Kirchner, 1958). The load factor n can be increased or decreased to make the task more or less difficult, respectively. This task draws upon the resources of the phonological loop and the central executive, but makes no demands on the visuo-spatial sketchpad component of working memory.

Martein, et al. (1999) developed a task theorised to assess all components of working memory simultaneously. The Double Span Task (DST) is the first experimental task designed to assess all components of Baddeley's (1992) model of working memory. The DST involves the presentation of increasingly longer sequences of common objects, such as an apple and a shoe, which appear one at a time in random locations on a 4 x 4 grid. Immediately following the presentation of the last object in the sequence, participants are asked for different types of recall, which are indicated by the words 'pictures', 'positions', or 'both'. Participants are required to either verbally recall the names of the objects presented (prompted by the word 'pictures'), to indicate the location where the objects were displayed on the corresponding location on an empty 4 x 4 grid (prompted by the word 'locations'), or to name the objects while indicating corresponding locations on an empty 4 x 4 grid (prompted by the word 'both'). Martein, et al.(1999) originally developed a manual version of the DST, which was later computerized by Kemps and Tiggemann (2005).

Within the context of Baddeley's (1992) model of working memory, it is theorised that during the presentation of objects, all systems must be activated in order to encode the objects (phonological loop), the locations

(visuo-spatial sketch pad), and their combinations (phonological loop, vi-

suo-spatial sketch pad, central executive, and the episodic buffer; Martein, et al., 1999). The single recall of objects or locations are proposed to activate either the phonological loop or the visuo-spatial sketchpad, respectively. Each of these processes however, also relies on the central executive to monitor the rehearsed information and retrieve the appropriate information when prompted. The combined recall of objects and their locations requires integrating the information from two different modalities (verbal information stored in the phonological loop, and spatial information stored in the visuo-spatial sketchpad) and matching them together, also requiring the central executive and episodic buffer. Therefore, unlike the single recall components of the task, 'double recall' or 'both' trials are theorized to rely heavily on the episodic buffer to integrate, store, and monitor matched pairs of stimuli from two different modalities, the phonological loop and the visuo-spatial sketchpad (Baddeley, 2000). Thus the 'double-recall' or 'both' trials are hypothesised to assess the capacity of all of the components of the working memory system simultaneously (Martein, et al.,1999). The central executive is hypothesised to play a particularly important role in the selection of the relevant object or location information, or both when responding to trials.

used to establish the discriminant validity of the DST as it is theorized

Despite its promising potential utility, the DST has not yet been validated against well-established, standardised measures of working memory, or been assessed for reliability. Therefore, the aim of the current study is to validate the DST against the Letter Number Sequencing (LNS) Task (Wechsler, 1997), a widely accepted measure of working memory (Economou, 2009; Shelton, Elliott, Hill, Calamia, & Gouvier, 2009). The LNS is theorized to draw upon the central executive and the phonological loop. Although the LNS is an incomplete measure of the entire working memory system, it is recognised as the best clinical measure of working memory to date (Shelton, et al., 2009).

The DST will also be validated against the Controlled Oral Word Association (COWA) Task (Benton, Hamsher, & Sivan, 1994), which is a commonly used measure of executive functioning (Bell-McGinty, Podell, Franzen, Baird, & Williams, 2002; Ross, Calhoun, Cox, Wenner, Kono, & Pleasant, 2007). The COWA relies on working memory only in the sense that people need to continually update information to avoid repetition of words. However, the COWA is unlikely to draw much upon the phonological loop, in that repetitions of words are not penalized, they simply do not increase the score. The COWA does not use the visual-spatial sketchpad at all. More importantly, the task requires participants to tap into their extensive vocabulary, which is stored in long-term memory, to generate words that begin with a particular letter. In this study, the COWA was

to assess distinctly different components of cognitive functioning. This study also aims to assess the test-retest reliability of the Double Span Task.

Hypothesis 1. Performance on the DST and the LNS will be strongly, positively correlated, as they are both measures of working memory.

Hypothesis 2. Performance on the Double Span Task and the COWA will be positively but only weakly correlated. A weak correlation is predicted because unlike the LNS, the COWA is a measure of executive functioning, requiring little if any working memory.

## Method

This study was approved by the Social and Behavioural Ethics Committee of Flinders University.

### Participants

The reliability of the DST was assessed in a sample of 105 older adults (M age = 64.3 yr., SD = 6.4; 51 men, 54 women). The validity of the DST was assessed using a different sample of 49 older adults (M age = 70.0 yr., SD = 9.3; 27 men, 22 women). We chose to test the reliability and validity of the DST in older individuals because of its potential utility for assessing working memory in those vulnerable to developing impairments of working memory (Hedden & Gabrieli, 2004). It is well documented that tasks relying on working memory are vulnerable to the effects of ag (Craik, 1994). Research has consistently demonstrated that older individuals perform significantly worse than younger individuals on tasks that involve two simultaneous inputs, or necessitate re-organisation of material to be recalled (Daigneault & Braun, 1993; De Beni & Palladino, 2004). Participants suffering from moderate to severe depression, anxiety, or cognitive impairment were excluded from this study. Although no older adults were excluded from participation based on these criteria, six individuals withdrew their participation due to illness.

Participants were recruited from a variety of sources, including advertisements in the local newspapers and electronic media of metropolitan Adelaide, and announcements to social groups, such as Senior Citizens. The participants provided written, informed consent and they were assured their responses were anonymous and they could withdraw their participation at any time. Participants were compensated A$20 for their time.

### Screening Measures

Screening measures included the Vocabulary and Matrix Reasoning subtests of the Weschler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999) and the Depression Anxiety and Stress short form (DASS21; Lovibond & Lovibond, 1993). Based on participants' scores, these measures were used to ensure participants did not demonstrate moderate to severe symptoms of depression, anxiety, stress, or cognitive impairment.

Weschler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999).—The WASI is a general measure of adult intelligence (Psychological Corporation, 1999). Two subtests of the WASI were used to give an indication of global cognitive functioning. The Vocabulary task was used as an assessment of crystallized, verbal intelligence whereas the Matrix Reasoning task was used to assess nonverbal fluid reasoning. The total score for each subtest was converted to age-group-specific T scores (M = 50, SD = 10), which were then summed to form an estimated full-score WASI intelligence quotient (M = 100, SD = 15; Psychological Corporation, 1999). Participants with a full-scale, WASI IQ below 70 were excluded from participation as such scores are indicative of severe cognitive impairment (American Psychiatric Association, 1994). Research has established high internal consistency (r = .94) and inter-rater reliability (r = .98) for both subtests (Psychological Corporation, 1999). Convergent validity has also been established between the WASI and the Weschler Adult Intelligence Scale–Third Edition, with reasonable correlations for both the Vocabulary (r = .88) and Matrix Reasoning (r = .66) subtests.

Depression Anxiety and Stress short form (DASS21; Lovibond & Lovibond, 1993).—The DASS21 is a set of three self-report scales, which measure current states of depression, anxiety, and stress. Scores for each subscale range from 0 – 42 and the overall score ranges from 0 – 126. In each case, higher scores indicate greater depression, anxiety, or stress. Normal levels of depression, anxiety, and stress are indicated by scores on each subscale of < 9, <7, and <14, respectively, (Lovibond & Lovibond, 1993). The DASS21has high internal consistency (Cronbach's alpha) for each subscale ranging from .94 to .96 for the Depression subscale, .87 to .89 for the Anxiety subscale, and .91 to .93 for the Stress subscale Brown, Chorpita, Korotitsch, & Barlow, 1997) (Antony, Bieling, Cox, Enns, & Swinson, 1998. The concurrent validity of the DASS21 has been supported by high correlations between the Depression subscale and the Beck Depression Inventory (r = .79), and the Anxiety subscale and the Beck Anxiety Inventory (r = .85) (Antony, et al., 1998).

## Primary Outcome Measures

Double Span Task (DST).—The Double Span Task (DST) is a 36-item performance task that assesses working memory performance (Martein, et al., 1999). The Double Span Task used in the current study was a computerized version of the double span memory task developed by Kemps and Tiggemann (2005).

Participants were seated approximately 45 cm in front of a 17-inch laptop computer. For each trial, participants were presented with a sequence of pictures, which appeared one at a time in a different, random location on a 4x4 grid. The pictures were presented at a rate of one per 1.5 sec., with an inter-stimulus interval of 0.5 sec. Immediately following the presentation of the last picture in the sequence, the word 'pictures', 'positions', or 'both' appeared on the screen. These words were used to prompt the participants to complete one of three tasks. Participants were required to verbally recall the names of the objects presented (prompted by the word 'pictures'), to indicate the location where the objects were displayed by using the mouse to click on the corresponding location on an empty 4x4 gird (prompted by the word 'locations'), or to name the objects while clicking on their corresponding locations on an empty 4x4 grid (prompted by the word 'both'). Participants were instructed that all three types of recall of objects, locations, or both had to be in the same sequential order as the presentation. Participants were not informed prior to each trial whether they would be required to recall the names of the objects, their positions, or both.

This task used 14 different objects that have high name, concept, and image agreement, and are highly familiar to participants and visually simple (Martein, et al., 1999). Participants were familiarized with these objects prior to commencing the task. Participants were also provided with six practice trials to familiarize themselves with the task. On any given trial, no two objects were the same or appeared in the same location. The objects and locations also differed between trials so that identical objects and locations were not used on consecutive trials.

The sequences used in this task progressively increased beginning with 2 objects and increasing to 6 objects. Participants were required to complete 6 consecutive trials at each sequence length. Of these, there were 2 object recall trials, 2 location recall trials, and 2 both object and location recall trials in random order.

Four scores were derived from the DST: a single Object recall score, a single Locations recall score, a Double recall score, and an Overall score. The single Objects and single Locations recall scores were the sum of correct responses for the trials when participants were asked to verbally recall the names of the objects or to recall the locations of the objects, respectively. The Double recall score was the sum of correct responses for the trials when participants were asked to recall the names of the objects while they indicated their corresponding location on an empty grid. For the Double recall score, participants received one point for each correctly named object and position pair in correct sequential order. Thus, the Double recall score for a trial was the total number of correct pairs in correct order. The Overall score was the sum of the total scores from the three different types of recall (single recall of objects, locations, and the double recall). Possible scores ranged from 0–40 for single recall of Objects, 0–40 for single recall of Locations, 0–40 for Double recall, and 0–120 for Overall recall. In each case, higher scores indicated better working memory.

Letter Number Sequencing Task (LNS).—The LNS assessed working memory (Wechsler, 1997). The LNS was comprised of 30 trials. Each trial involved the experimenter reading aloud a series of letters and numbers in an alternating pattern (e.g. V-1-J-5). On each trial, participants had to repeat the letters in alphabetical order aloud to the experimenter (e.g., J–V), and secondly, repeat the numbers in numerical order (e.g., 1-5). The alpha-numerical sequences ranged from 1 letter and 1 number to 4 letters and 4 numbers starting from 1 letter/number and incrementing to 4 of each. Each trial was scored as either 1: Correctly recalled sequence of letters and numbers or 0: Incorrectly recalled sequence. Possible scores range between 0–30, with higher scores indicative of greater working memory performance. The LNS has high internal consistency (r = .88) and test retest reliability (r = .76; Sattler & Ryan, 2009).

Controlled Oral Word Association Test (COWA).—The COWA assesses executive functioning, purportedly including cognitive flexibility, strategic planning, and working memory (Benton, et al., 1994). The COWA required participants to generate words that begin with a specific letter of the alphabet (e.g., F, A, S). Participants were given 60 seconds to generate as many different words as they can beginning with the letter F. This procedure was then repeated for the letters 'A' and 'S'. Participants were instructed that proper nouns (names of people and places), numbers (e.g., four), repeated words with different endings (i.e., flat, flatter), or repeating a previously repeated word did not add to their score. An overall score was calculated from the sum of words produced for all three letters. Higher scores indicated greater executive functioning. The COWA has high internal consistency (r = .83) and test re-test reliability (r = .74; Ruff, Light, & Parke, 1996).

## Procedure

Following completion of the screening measures (WASI and DASS21), eligible participants arranged to come to Flinders University Laboratory for performance testing. For assessment of reliability, 105 participants attended the laboratory on two occasions, 3 months apart. For validity testing, 49 participants attended the University on one occasion only to complete the DST, the LNS, and the COWA.

On arrival at the University, participants were asked to complete the DST on the laptop computer. Participants were presented with standard instructions on the computer screen and given six trials to practice the task. The practice trials consisted of two presentations only, with two of each recall type. The experimenter clarified the procedure of the task if participants were unsure during the practice trials. Following the practice trials, participants then completed all 36 trails of the task, including 6 trials of each level of difficulty incrementing from 2–6 presentations per trial. The experimenter recorded the names of the objects recalled, while the computer program recorded the positions indicated by the participant. No feedback was given once the trials began. Following completion of the DST, participants were then required to complete the LNS and the COWA.

Following final participation, all individuals were verbally debriefed about the objectives of the study and reimbursed for their time and travel costs.

### Analysis

The current study utilized a correlational, non-experimental design to validate the DST using the raw LNS and COWA scores. Age was entered as a covariate in analyses, conducted to assess the relationship between each of the DST components with the LNS and COWA. Test-retest reliability, Cronbach's alpha, and McDonald's omega were also calculated as an assessment of the reliability of the Double Span Task. All analyses were conducted using Predictive Analytic Software (PASW) release Version 18.0.0 (SPSS, Inc., 2009, Chicago, IL). McDonald's omega was calculated using R release Version 2.15.2 (The R Foundation for Statistical Computing, 2012, Vienna, Austria).

## Results

### Data Screening

Table 1 below shows the means and standard deviations for all screening and primary outcome measures including the DST, the LNS, and the COWA. Histograms of distributions for all variables including the DST, the LNS, and the COWA were inspected for normality. Scores on all components of the DST (single recall of Objects, single recall of Locations, Double recall, and Overall memory score), LNS, and COWA met the assumption of normality according to the Kolmogorov-Smirnov statistic. No outliers were detected.

### Reliability

To investigate the reliability of the DST, Cronbach's alpha was calculated for single recall of Objects, single recall of Locations, and Double recall components of the DST, as well as the overall score. Cronbach's alpha coefficient for the Overall score was .81, which demonstrates good internal consistency. The Cronbach's alpha coefficient and McDonald's omega for the separate components of the DST are shown in Table 2 below.

TABLE 1

Descriptive Statistics for Screening Measures Including Weschler Abbreviated Scale of Intelligence (WASI; Vocabulary Subtest [normed T score], Matrix Reasoning Subtest [normed T score] and Estimated Full-scale Intelligence Quotient [normed standard score]), Depression Anxiety and Stress Scale (DASS; Depression, Anxiety and Stress Sub- scales and Overall Score [raw scores]), Double Span Task (DST; Object, Location, Double, and Overall Recall [raw scores]), Letter Number Sequencing Task (LNS; raw score), and Controlled Oral Word Association (COWA; raw score) Task

| Measure | M | SD |
|---|---|---|
| Weschler abbreviated scale of intelligence | | |
| Vocabulary subtest | 57.33 | 7.91 |
| Matrix reasoning subtest | 57.88 | 9.27 |
| Estimated full-scale IQ | 113.31 | 12.04 |
| Depression anxiety and stress scale | | |
| Depression | 1.71 | 3.46 |
| Anxiety | 1.84 | 2.33 |
| Stress | 2.84 | 3.14 |
| Overall | 6.45 | 7.30 |
| Double Span Task | | |
| Object | 24.71 | 6.54 |
| Location | 19.78 | 5.70 |
| Double | 16.14 | 4.94 |
| Overall | 87.35 | 18.19 |
| Letter number sequencing task | 17.60 | 2.43 |
| Controlled oral word association | 36.13 | 10.06 |

To further examine the reliability of the DST, Pearson product-moment correlations were conducted for performance on all recall components of the task across the two testing occasions. These correlations reveal adequate correlations for performance on all components across the two testing occasions (Table 2).

TABLE 2

Cronbach's Alpha Coefficient and McDonald's Omega for Components of Double Span Task: Object Recall, Location Recall, Double Recall and Overall Recall and Correlation Coefficients for Performance on Two Separate Occasions

| Double Span Task: Recall Type | Cronbach's Alpha | McDonald's Omega | r | p |
|---|---|---|---|---|
| Object | .66 | .72 | .59 | <.001 |
| Location | .69 | .67 | .70 | <.001 |
| Double | .69 | .83 | .71 | <.001 |
| Overall | .81 | .87 | .74 | <.001 |

### Validation

Pearson product-moment correlations were conducted separately between LNS, COWA, and Matrix Reasoning performance, and performance on all recall components of the DST to establish the validity of the DST as a measure of working memory. After controlling for age, Pearson product-moment correlations revealed moderately strong positive correlations between LNS performance and all recall components of the DST (see Table 3 below). No statistically significant correlations were found between any type of recall on the DST and performance on the COWA. Moderate positive correlations were observed between performance on the Matrix Reasoning task and the Object, Location, and Overall recall components of the DST.

Direct comparisons of the DST/LNS correlation and DST/COWA correlations were also conducted to establish an indication of the discriminant validity of the DST. These comparisons are displayed in Table 3 above and indicate statistically significant differences ($p < .05$) on all four components.

TABLE 3

Correlations of Double Span Task Performance (Object Recall, Location Recall, Double Recall and Overall Recall) with Letter Number Sequencing Task, Controlled Oral Word Association Task and Matrix Reasoning Performance Controlling for Age. Right Column Displays Significance of Difference Between Double Span/Letter Number Sequencing Task and Double Span/Controlled Oral Word Association Task Correlations

| | Letter Number Sequencing | | Controlled Oral Word Association | | Matrix Reasoning | | Difference in Correlations |
|---|---|---|---|---|---|---|---|
| Recall type | r | p | r | p | r | p | p |
| Object | .51 | .013* | .22 | .24 | .27 | .05 | <.05 |
| Location | .35 | .150 | –.16 | .37 | .28 | .03 | <.01 |
| Double | .58 | .003* | –.02 | .88 | .12 | .11 | <.001 |
| Overall | .55 | .006* | .04 | .88 | .30 | .02 | <.005 |

*$p < .01$ (2-tailed).

## Discussion

The first pre-requisite in establishing validity is to ensure high reliability. Reliability coefficients demonstrate the DST is a reliable measure of performance. The internal consistency of the DST, as measured by Cronbach's alpha and McDonald's omega, was acceptable to good (.66–.87) for this population of older adults. Likewise, the test-retest reliability of the DST was good. Scores on all components of the DST (Object recall, Location recall, Double recall, and Overall recall) were strongly correlated when retested 3 months later.

The main aim of the current study was to validate the DST as a measure of working memory using a population of older adults. It was expected that performance on the DST and the LNS would be strongly, positively correlated as both measures assess working memory. A positive but weak correlation was predicted between performance on the DST and the COWA. Unlike the LNS, the COWA does rely on executive functioning but to a lesser extent. Results have supported the use of the DST as a measure of working memory performance. Recall performance on all components of the DST, except Location, was positively correlated with performance on the LNS, a well-established measure of working memory. In contrast, no statistically significant correlations were found between performance on any component of the DST and performance on the COWA, a broader assessment of executive functioning. Weak validity for the visuo-spatial component for the DST was demonstrated by positive correlations between recall performance on the Object, Location, and Overall components of the DST with performance on the Matrix Reasoning task. The apparent lower correlation between performance on the Matrix Reasoning task and the Double recall component of the DST is likely to reflect the lack of reliance on the central executive when performing the Matrix Reasoning task.

Performance on the LNS was strongly, positively correlated with Overall recall, single recall of Objects, and Double recall components of the DST. Performance on the LNS was only moderately, but not significantly positively correlated with single recall of Locations on the DST. This is not surprising given that there is no visuo-spatial component to the LNS. Given the LNS has been established as a measure of working memory (Economou, 2009; Shelton, et al., 2009), this finding supports the use of the DST as a measure of working memory. Unlike the DST, the LNS does not assess all components of the working memory system simultaneously, but rather the ability of one slave system and the central executive (Martein, et al., 1999). The phonological loop is required in the LNS to remember the letters and numbers, while the central executive is required for arranging the letters into alphabetical order and numbers in ascending order for accurate recall. It could be argued that the LNS is similar to the single recall of objects in the DST, which also utilises the phonological loop and the central executive. The current study did assess the difference in correlations between performance on the LNS and single recall of Objects and Locations. Although the correlation between the performance on the LNS and the single recall of Objects ($r = .50$) was greater than the correlation for the single recall of Locations ($r = .35$), this difference was not statistically significant ($t = 0.95$, $p > .05$).

Performance on the COWA was not significantly related to any component of the DST. The COWA has been suggested as a measure of executive function; however, it remains questionable whether working memory is utilised during the COWA. Although the COWA requires individuals to produce as many words as possible that begin with a certain letter, presumably tapping the phonological store rather than visuo-spatial abilities, researchers have argued it is more a measure of long term memory and cognitive flexibility (creative thinking) than working memory. Ross, et al.(2007) argued that the COWA involves accessing and retrieving information previously stored in long-term memory. These authors argued working memory is only required to remember the words that an individual has already verbalised and inhibit incorrect responses. However, given participants were not penalised for repetition errors (despite repeated words taking up time, which could be utilized for other scorable words), their motivation to keep already verbalised words in working memory may not be a priority in this task. The significant differences between the DST/LNS correlations and the DST/COWA correlation provides support for the discriminant validity of the DST as a measure of working memory.

The current findings support the use of the Double Span Task as a measure of working memory and represent an initial step in evaluating the reliability and validity of the DST. The correlations reported are moderate in strength, suggesting additional sources of potential variance in working memory performance. This study did not examine the validity of the DST and separate components of working memory. Future research should use separate measures of the phonological loop, visuo-spatial sketchpad, executive function, and episodic buffer to establ ity of the DST as a measure of these separate components. Future studies should also consider developing normative data to allow comparisons of performance on the DST across age. The DST, reliant on working memory in two different realms, may be particularly sensitive to ageing effects and thus require comparisons across age. Furthermore, the normative data would be useful in comparisons with groups suffering pathologies (e.g., depression, anxiety, chronic fatigue, PTSD, insomnia, neurodegenerative disorders, etc.) that may interfere with working memory overall and its specific components.

## REFERENCES

AMERICAN PSYCHIATRIC ASSOCIATION. (1994) Diagnostic and Statistical Manual of Mental Disorders. (4th ed.). Washington, D.C.: Author.

ANTONY, M., BIELING, P., COX, B., ENNS, M., & SWINSON, R. (1998) Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychological Assessment, 10(2), 176-181.

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

BADDELEY, A. (1992) Working memory. Science, 255(5044), 556-559.

BELL-MCGINTY, S., PODELL, K., FRANZEN, M., BAIRD, A. D., & WILLIAMS, M. J. (2002) Standard measures of executive function in predicting instrumental activities of daily living in older adults. International Journal of Geriatric Psychiatry, 17(9), 828-834.

BENTON, A. L., HAMSHER, K. D. S., & SIVAN, A. B. (1994) Multilingual Aphasia Examination. (3rd ed.) Iowa City: A J A Associates, Inc.

BROWN, T. A., CHORPITA, B. F., KOROTITSCH, W., & BARLOW, D. H. (1997) Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. Behaviour Research and Therapy, 35(1), 79-89.

CRAIK, F. I. (1994) Memory changes in normal aging. Current Directions in Psychological Science, 3(5), 155-158.

DAIGNEAULT, S., & BRAUN, C. M. (1993) Working memory and the Self-Ordered Pointing Task: further evidence of early prefrontal decline in normal aging. Journal of Clinical & Experimental Neuropsychology: Official Journal of the International Neuropsychological Society, 15(6), 881-895.

DE BENI, R., & PALLADINO, P. (2004) Decline in working memory updating through ageing: intrusion error analyses. Memory, 12(1), 75-89.

ECONOMOU, A. (2009) Memory score discrepancies by healthy middle-aged and older individuals: the contributions of age and education. Journal of the International Neuropsychological Society, 15(6), 963-972.

GILHOOLY, K. J., LOGIE, R. H., WETHERICK, N. E., & WYNN, V. (1993) Working memory and strategies in syllogistic-reasoning tasks. Memory & Cognition, 21(1), 115-124.

HEDDEN, T., & GABRIELI, J. D. E. (2004) Insights into the aging mind: a view from cognitive neuroscience. Nature Reviews, Neuroscience, 5, 87-96.

JUST, M. A., & CARPENTER, P. A. (1992) A capacity theory of comprehension: individual differences in working memory. Psychological Review, 99(1), 122-149.

KEMPS, E., & TIGGEMANN, M. (2005) Working memory performance and preoccupying thoughts in female dieters: evidence for a selective central executive impairment. British Journal of Clinical Psychology, 44(Pt 3), 357-366.

KIRCHNER, W. K. (1958) Age differences in short-term retention of rapidly changing information. Journal of Experimental Psychology, 55(4), 352-358.

KLAUER, K. C., STEGMAIER, R., & MEISER, T. (1997) Working memory involvement in propositional and spatial reasoning. Thinking and Reasoning, 3(1), 9-47.

LOISY, C., & ROULIN, J.-L. (2003). Multiple short-term storage: A task to evaluate the coordination function of the central executive. *International Journal of Psychology, 38*(1), 1-10.

LOISY, C., & ROULIN, J-L. (1992). *Multiple short-term storage in working memory: A new experimental approach.* Paper presented at the Fifth conference of the European Society for Cognitive Psychology. Paris, France.

LOVIBOND, S., & LOVIBOND, P. (1993) Manual for the Depression Anxiety Stress Sacle (DASS). NSW, Australia: Psychology Foundation Monograph.

MARTEIN, R., KEMPS, E., & VANDIERENDONCK, A. (1999) The role of working memory in a double span task. Psychologica Belgica, 39(1), 15-27.

PSYCHOLOGICAL CORPORATION. (1999) Wechsler Abbreviated Scale of Intelligence (WASI) manual. San Antonio: TX: Author.

ROSS, T. P., CALHOUN, E., COX, T., WENNER, C., KONO, W., & PLEASANT, M. (2007) The reliability and validity of qualitative scores for the Controlled Oral Word Association

Test. *Archives of Clinical Neuropsychology, 22*(4), 475-488.

Ruff, R. M., Light, R. H., & Parker, S. B. (1996) Benton Controlled Oral Word Association Test: reliability and updated norms. *Archives of Clinical Neuropsychology, 11*(4), 329-338.

Sattler, J., & Ryan, J. (2009) *Assessment with the WAIS–IV*. California: Sattler Publisher, Inc.

laboratory and clinical Working Memory Tests and their predic- tion of fluid intelligence. Intelligence, 37(3), 283.

WECHSLER, D. (1997) Wechsler Adult Intelligence Scale. (5th ed.) San Antonio, TX: Har- court Assessment.