MEMORY, 2015
http://dx.doi.org/10.1080/09658211.2015.1076849

Routledge
Taylor & Francis Group

An improved spatial span test of visuospatial memory

David L. Woods[a,b,c,d], John M. Wyma[a], Timothy J. Herron[a] and E. William Yund[a]

[a]Human Cognitive Neurophysiology Laboratory, VANCHCS, 150 Muir Rd., Martinez, CA 94553, USA; [b]Department of Neurology, UC Davis, 4860 Y St., Suite 3700, Sacramento, CA 95817, USA; [c]Center for Neurosciences, UC Davis, 1544 Newton Ct., Davis, CA 95616, USA; [d]Center for Mind and Brain, UC Davis, 202 Cousteau Place, Suite 201, Davis, CA 95616, USA

ABSTRACT
In the widely used Corsi Block Test and Wechsler Spatial Span Tests, participants must reproduce sequences of blocks in the order touched by the examiner until two trials are missed at the same sequence length. The examiner records either the maximum number of blocks correctly reported or the total number of correct lists. Here, we describe a computerized spatial span test (C-SST) that uses psychophysical procedures to quantify visuospatial mean span (MnS) with sub-digit precision. Results from 187 participants ranging in age from 18 to 82 years showed that accuracy declined gradually with list length around the MnS (by ~30% per item). Simulation studies revealed high variance and biases in CBT and Wechsler measures, and demonstrated that the C-SST provided the most accurate estimate of true span (i.e., the sequence length producing 50% correct). MnS declined more rapidly with age than mean digit span (MnDS) measured in the same participants. Response times correlated with both MnS and MnDS scores. Error analysis showed that omission and transposition errors predominated, with weaker primacy and recency effects in spatial span than digit span testing. The C-SST improves the precision of spatial span testing and reveals significant differences between visuospatial and verbal working memory.

ARTICLE HISTORY
Received 16 December 2014
Accepted 20 July 2015

KEYWORDS
Digit span; meta-analysis; sex; processing speed; aging; gender; reaction time; computer; review

Introduction

Spatial span tests (SSTs) of visuospatial memory are widely used both in clinical settings and in developmental studies of spatial working memory (Berch, Krikorian, & Huha, 1998). The most common SST is the Corsi Block Test (CBT) (Corsi, 1972). In the CBT, the participant is presented with a set of nine blocks fixed to a checkerboard-sized board. The blocks are tapped in sequence by the examiner, beginning with a sequence length of two blocks. The participant's task is to reproduce the sequence by touching each block in the same order as the examiner. Two sequences are presented at each length. If the participant can correctly reproduce either sequence, list length is increased by one additional block. Testing ceases when the participant misses both test sequences, and maximal span is quantified as the length of the longest block sequence correctly reproduced (Beblo, Macek, Brinkers, Hartje, & Klaver, 2004; Claessen, van der Ham, & van Zandvoort, 2014; Fournet et al., 2012; Kessels, van den Berg, Ruis, & Brands, 2008; Kessels, van Zandvoort, Postma, Kappelle, & de Haan, 2000; Ostrosky-Solis, Jaime, & Ardila, 1998; Tamayo et al., 2012).

The Wechsler Memory Scale (WMS) 3rd edition SST uses the same trial delivery and test-termination rules with a set of 10 blocks (Wechsler, 1997), but scores the test by counting the total number of trials reported correctly, regardless of length (Lo, Humphreys, Byrne, & Pachana, 2012; Wiechmann, Hall, & O'Bryant, 2011; Wilde, Strauss, & Tulsky, 2004). Finally, Kessels and colleagues used a 9-block display but report both maximal span and total weighted span, the product of the maximal span and the number of total correct trials (Claessen et al., 2014; Kessels et al., 2000; Kessels et al., 2008).

Other investigators have used different sequence delivery and termination rules. For example, Capitani, Laiacona, and Ciceri (1991) presented three trials at each length and terminated testing when the participant missed two of three sequences, and Orsini et al. (1986) presented five trials at each list length and terminated testing when the participant missed three of five trials. In contrast, Farrell Pagulayan, Busch, Medina, Bartok, and Krikorian (2006) presented five trials at each length, but quantified maximal span as the length where a participant could reproduce at least one sequence correctly.

Table 1 provides a summary of spatial span measures obtained on different SSTs. Comparisons of the results from different studies are complicated by the different methods used for test administration and scoring (Berch et al., 1998). Nevertheless, significant age effects are evident in virtually all studies of spatial span (Brunetti,

CONTACT David L. Woods dlwoods@ucdavis.edu

Table 1. Normative studies of spatial span.

Manual CBT	N	Mean age	Max span	True span	Age-predicted true span	Diff.
Orsini et al. (1986)	1354	57.2	4.56 (1.08)	5.24	5.54	−0.30
Capitani et al. (1991)	495	53.4	5.03 (1.07)	5.76	5.63	0.13
Ostrosky-Solis (1998)	105	46.7	5.20	5.36	5.78	−0.42
Kessels et al. (2000)	70	31.2	6.2 (1.3)	6.38	6.15	0.23
Beblo et al. (2004)	48	25	6.2 (1.0)	6.38	6.30	0.08
Farrell et al. (2006)	94	21.7	7.1 (1.0)	6.00	6.38	−0.38
Kessels et al. (2008)	230	66.5	5.12 (0.78)	5.28	5.32	−0.04
Tamayo et al. (2012)	179	36	5.50	5.67	6.04	−0.37
Monaco et al. (2013)	362	54.2	5.38 (1.09)	5.56	5.61	−0.05
Claessen et al. (2014)	40	22.9	7.0 (1.2)	7.19	6.35	0.84
Computerised CBT	*N*	Mean age				
Brunetti et al. (2014)	107	32.3	5.62 (1.1)	5.79	6.12	−0.33
Claessen et al. (2014)	40	22.9	6.40 (1.5)	6.58	6.35	0.33
WMS: forward SS	*N*	Mean age	Total correct			
Wilde et al. (2004)	1250	48	8.01 (2.32)	5.71	5.75	−0.04
Lo et al. (2012)	339	57.9	7.49 (1.74)	5.43	5.52	−0.09
Wiechmann et al. (2011)	44	77	8.3 (2.54)	5.88	5.07	0.81
Computerised CBT	*N*	Mean age	Total correct			
Fournet et al. (2012)	382	70	6.14 (1.95)	4.92	5.23	−0.31

Notes: Studies have been divided into manual and computerised studies using the CBT and manual and computerised studies quantifying forward span in the WMS-III SST. Means and standard deviations are shown. True span values (the length where 50% of trials are correct) were estimated using Table 7 conversions. Age-predicted true spans were calculated using linear regression using the correlation ($r = -0.74$) between true span means and mean population ages for the different studies (slope per year = −0.024). Diff, differences between observed and age-predicted true spans.

Del Gatto, & Delogu, 2014; Capitani et al., 1991; Fournet et al., 2012; Monaco, Costa, Caltagirone, & Carlesimo, 2013; Orsini et al., 1986; Ostrosky-Solis et al., 1998; Wilde et al., 2004). Age effects are relatively large in most studies, with declines exceeding a full digit between the ages of 20 and 80 (Orsini et al., 1986). Moreover, in studies comparing age-related declines in spatial span and digit span tests, the rate of decline is typically greater for spatial span (Hester, Kinsella, & Ong, 2004; Orsini et al., 1986; Park et al., 2002; Wilde et al., 2004), consistent with greater age-related declines in spatial than verbal working memory. Other factors, including sex and education, have more variable effects on spatial span. Some studies have found better performance in male participants (Capitani et al., 1991; Fournet et al., 2012; Orsini et al., 1986), while others have failed to find significant sex differences (Kessels et al., 2008; Monaco et al., 2013). Similarly, education has a small but significant effect in many studies (Fournet et al., 2012; Kessels et al., 2008; Orsini et al., 1986), but other studies have failed to find significant education-related differences (Capitani et al., 1991).

Traditional SST paradigms suffer from four limitations: (1) *Biased sampling of span.* Because testing proceeds exhaustively through short lists and halts once longer lists are missed, more lists are sampled below than above maximal span. As a result, the maximal span measure is biased relative to "true span", the list length where participants would correctly identify 50% of lists (see below). In addition, fewer trials are presented to participants with shorter spans. (2) *Single digit measurement precision.* Previous studies have suggested that accuracy of report declines gradually with list length, with participants correctly reporting 20–30% of lists presented one digit above their maximal spans (Brunetti et al., 2014). This continuous function is poorly captured by single digit precision metrics such as maximal span. Moreover, many participants will have identical maximal span scores despite differences in performance (see below). (3) *Fixed stimulus sequences.* The layout of blocks remains unchanged throughout traditional tests, and each participant is tested with the same block sequences at each sequence length. However, the procedures used for creating the block layout and selecting the standard block sequences are not specified. Sequence selection is an important variable because previous studies have found that factors such as overall path distance and the number of path crossings can significantly influence performance; for example, some shorter sequences with greater path distance and more crossings are often more difficult than simpler sequences that contain more blocks (Busch, Farrell, Lisdahl-Medina, & Krikorian, 2005; Claessen et al., 2014; Kessels et al., 2000; Orsini, Simonetta, & Marmorato, 2004; Smirni, Villardita, & Zappala, 1983). Since path distance and configuration will inevitably differ between any two sequences, the sequences used at each list length will inevitably vary somewhat in difficulty. Moreover, the increases in difficulty associated with increasing list length may also be somewhat unequal. (4) *Examiner influences.* Test performance depends on difficult-to-standardise cues provided by the examiner. In most tests, the examiner touches each block at intervals of 1.0 or 1.5 s. However, examiners differ as to how the blocks are touched (Orsini et al., 1986), and some investigators use a pencil rather than their finger (Farrell Pagulayan et al., 2006).

Computerised versions of the CBT using standard rules for sequence delivery and scoring have been utilised in previous studies to eliminate the influence of the examiner

(Beblo et al., 2004; Brunetti et al., 2014; Claessen et al., 2014; Fournet et al., 2012; Vandierendonck, Kemps, Fastame, & Szmalec, 2004). We used a computerised version of the SST (C-SST) that modified the sequence delivery and scoring rules following principles developed in a previous computerised test of digit span (Woods et al., 2011), resulting in three major changes: (1) *Balanced sampling of span*. Sequence lengths were adjusted adaptively using a staircase procedure: list lengths either increased following each correct report or were reduced following two successive errors. This assured adequate sampling of list lengths above and below the participant's true span. All participants received 14 trials regardless of their performance. (2) *Sub-digit measurement of span*. Performance was quantified using a psychophysical estimate of mean span (MnS) that took into consideration the participant's performance on all trials and had sub-digit precision. Traditional maximal span and total correct metrics were also recorded. (3). *Randomised displays and sequences*. Both the spatial layout of squares and the item sequences were selected randomly on each trial. This assured that performance variations due to differences in list length were not biased by the particular sequences used for testing, and also enabled the analysis of the effects of display layout, path distance, and path crossing on performance.

In addition, the C-SST measured the average response time (ReT) to the individual items in each sequence. Based on previous reports of strong relationships between processing speed and working memory capacity (Brown, Brockmole, Gow, & Deary, 2012; Holdnack, Xiaobin, Larrabee, Millis, & Salthouse, 2011; Mungas et al., 2014), we anticipated that a significant correlation would be observed between ReT and visuospatial span.

Previous studies of spatial memory have shown that errors are less frequent for the first and last items in list sequence, reflecting well-known primacy and recency effects. However, the primacy and recency effects seen in spatial span testing appear to be smaller than those observed in digit span testing (Smyth & Scholey, 1996b). Therefore, we also compared primacy and recency effects in digit and spatial span testing (Hurlstone, Hitch, & Baddeley, 2014).

We also analysed the types of errors that occurred. Previous studies using blocked sequences of fixed lengths have shown that transposition errors (swapping the position of two items) are the most common type of error (Hurlstone et al., 2014; Smyth & Scholey, 1996b). Here, we investigated the incidence of transposition errors, omission errors (failing to select an item), addition errors (adding an additional item), substitution errors (selecting an item that was not included in the test sequence), and permutation errors (complex reordering of the item sequence) as list length varied across trials.

Methods

Participants

We studied 189 participants who ranged in age from 18 to 82 years (mean 41.1 years), with 50% of the participants below the age of 30 and 33% above the age of 60. The participants were well educated (mean 14.6 years of education, range 10–20 years) and were predominantly (58%) male. Participants were recruited from advertisements in the San Francisco Bay Area on Craigslist (sfbay.craigslist.org), and from pre-existing control populations. Participants also indicated the number of hours/day that they used a computer: most of the participants used computers frequently (mean 2.4 hours/day), with 92.7% using computers for at least one hour/day. Participants were required to meet the following inclusion criteria: (a) fluency in the English language; (b) no current or prior history of bipolar disorder, mania, or schizophrenia; (c) no current substance abuse; (d) no concurrent history of neurologic disease known to affect cognitive functioning; (e) no history of hospitalisation for head trauma; (f) auditory functioning sufficient to understanding normal conversational speech; and (g) visual acuity normal or corrected to 20/40 or better. Participant ethnicities were 64% Caucasian, 12% African American, 14% Asian, 10% Hispanic/Latino, 2% Hawaiian/Pacific Islander, 2% American Indian/Alaskan Native, and 4% "other". All participants gave informed written consent following procedures approved by the Institutional Review Board of the VA Northern California Health Care System (VANCHCS) and were compensated for their participation.

Apparatus and stimuli

The C-SST was administered midway through a series of tests[1] and required 4–6 minutes to complete. Testing was performed in a quiet room using a standard PC controlled by Presentation software (Versions 13 and 14, NeuroBehavioral Systems, Albany CA). The test can be downloaded at http://www.ebire.org/hcnlab/cognitive-tests/SST. The C-SST display is shown in Figure 1(a). The locations of the 10 red squares presented on each test trial were determined by constrained random sampling and varied from trial-to-trial. Five squares were presented in each hemifield, with a single square presented in each horizontal row and in each vertical column (indicated by the dotted lines in Figure 1(a)). This resulted in 120 different possible layouts that were independently determined in each hemifield. In addition, small and random horizontal and vertical offsets prevented squares from being vertically or diagonally aligned. Thus, in contrast to the typical spatial span task, where the same block layout is used throughout

[1]The session included the following computerized tests and questionnaires: finger tapping, simple reaction time, Stroop, digit span forward and backward, phonemic and semantic verbal fluency, verbal list learning, spatial span, trail making, vocabulary, design fluency, the Wechsler Test of Adult Reading (WTAR), visual feature conjunction, risk and loss avoidance, delay discounting, the Paced Auditory Serial Addition Task (PASAT), the Cognitive Failures Questionnaire (CFQ), the Posttraumatic Stress Disorder Checklist (PCL), and a traumatic brain injury (TBI) questionnaire.

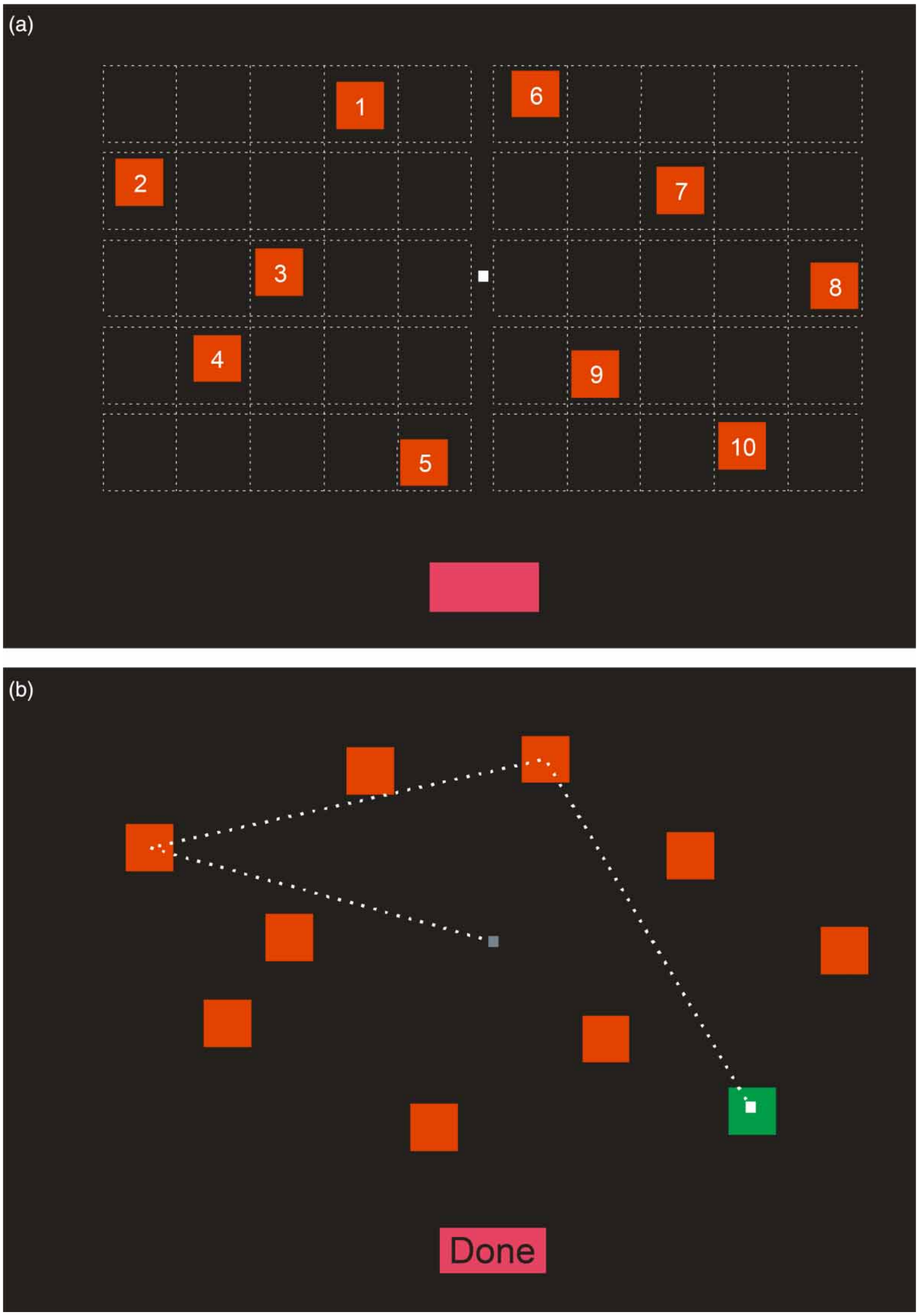

Figure 1. (a) The C-SST display. Ten squares were displayed at constrained random positions in the display matrix, with positions varying on each trial. Each square could appear in any of the five boxes per row in each hemifield, with the constraint that in each hemifield, each row and column contained one square. Squares 1–5 were in the left hemifield and squares 6–10 in the right hemifield. Numbers and dashed lines are for illustration only, and were not seen by participants. (b) A test sequence of length three. Display of the sequence "2-6-10". The sequence was shown by the moving cursor (small white square), with each square flashing green as it was selected (bottom right). When the sequence was complete, the "Done" button illuminated and the cursor returned to the center of the screen. The dashed line illustrates cursor displacement and was not visible to the participant.

(Berch et al., 1998), different displays were presented on each trial of the C-SST.

Sequence selection was also random, with the constraint that a square could only be selected once in each sequence. For example, length-six sequences were randomly selected from 151,200 ($10 \times 9 \times 8 \times 7 \times 6 \times 5$) possible sequences. Thus, in contrast to most SSTs, each participant received a unique list of block sequences. The locations of all squares and the sequence of squares selected were recorded on each trial, permitting the quantification of various path parameters (e.g., distance, crossings, angles, etc.) that may influence sequence recall.

As seen in Figure 1(b), squares were selected by a computer-controlled cursor (a small white square) that was displaced during a 677 ms interval, followed by a 300 ms

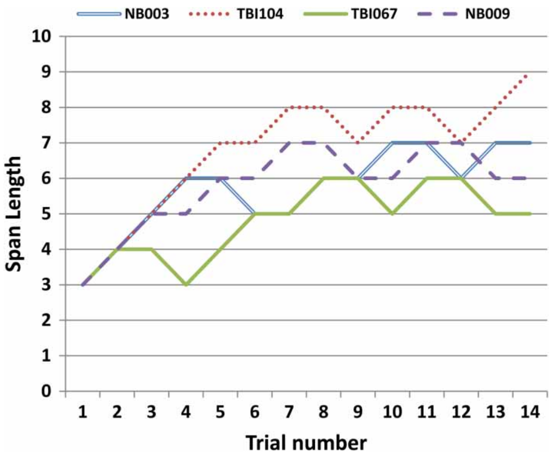

Figure 2. Span length on each trial in the 14-trial sequences for four different subjects. The span length either increased following each correct trial or was reduced following two successive errors.

interval when the selected square changed color in a green-red-green sequence (100 ms each) before finally returning to red to indicate a selection. The cursor continued to highlight each square in the sequence until the entire sequence was complete. Once the square sequence had been shown, the cursor returned to the center of the display and the "DONE" button appeared at the bottom of the screen.

During the response phase, the participant moved the cursor to each square and selected it by depressing the left mouse button (Razer Sidewinder, Carlsbad, CA). When the cursor fell within the boundaries of a square, the square changed colour from red to green. When selected with a mouse click, the square flashed red and green (each for 100 ms) and then changed back to red. The time to move the cursor and select each square was recorded (in ms). Once the participant selected a square, the selection could not be changed. When finished selecting squares, the participant clicked the "Done" button at the bottom of the screen. This was followed by a screen displaying a "Next" button that the participant clicked to begin the next trial.

There were four practice trials followed by 14 test trials. In the practice trials, the display contained six squares (three in each hemifield) and testing began at a sequence length of two. Practice trials included feedback.

Fourteen test trials were given in all, beginning with a sequence length of three squares. The sequence length increased following each trial that was reported correctly, and decreased following two successive misses. Figure 2 shows the sequence lengths tested in four participants. For example, participant TBI067 correctly reproduced the three-square sequence on trial 1, then missed two successive four-square sequences, reducing the span length back to three. On trials four and five, participant TBI067 correctly reproduced sequences of length three and four, then missed the 5-square sequence on trial six, but reproduced it correctly on trial seven. Thereafter, he missed all four 6-square sequences, but reproduced one of three 5-square sequences correctly.

Scoring metrics

The data from individual trials were analysed using different automated scoring metrics similar to those used in a previous study of digit span (Woods et al., 2011). The maximal span (MS) was the maximum trial length correctly reported before two successive trials were missed at the same length. Thus, the termination rule for MS was similar to that used in the CBT. We also measured the maximal length (ML) successfully repeated over the 14 trials, the mean length of the last five trials (Last5), and the total number of correct trials (TotC) over the entire 14-trial sequence. In addition, we estimated mean span (MnS), the extrapolated list length where 50% of lists would be correctly reported based on psychophysical estimation (Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004). The MnS baseline was set at 2.5 and was incremented by the fraction of digit strings accurately reported at each succeeding list length. Finally, we obtained the MnS *z*-score after regressing out the contributions of the two most important correlates, age and computer-use (see below). In addition, we measured response time (ReT), the average time to respond to each item, and obtained an age- and computer-use regressed *z*-score of the log-transformed mean ReT (Log-ReT-Z). We also obtained a mean digit span (MnDS) metric in forward span testing, as described in a previous experiment (Woods et al., 2011) that used a psychophysical procedure similar to that used for measuring the MnS. Age- and computer-use corrected MnDS *z*-scores were also obtained following similar procedures.

Statistical analysis

Correlation analysis was used to evaluate the effects of age, education, computer-use, and sex, and to develop normative regression functions. Pairwise effects were also analysed with Student's *t*-tests, using a model that assumes unequal variance in the different participant groups when appropriate. Group comparisons were further analysed using a multifactor mixed ANOVA. Separate ANOVAs were performed for age- and computer-use regressed *z*-scores (see below) for total completion time and movement velocity. Greenhouse-Geisser corrections of degrees of freedom were uniformly used in computing *p* values in order to correct for covariation within factors or interactions. Effect sizes are reported as partial ω^2 values.

Results

Span metrics

Figure 3 shows the participants' MnS scores (blue diamonds) as a function of age, and for comparison shows the same participants' MnDS forward digit span results (red squares), which have been presented elsewhere (Woods et al., 2011). The average results obtained with the different spatial span metrics are shown in Table 2.

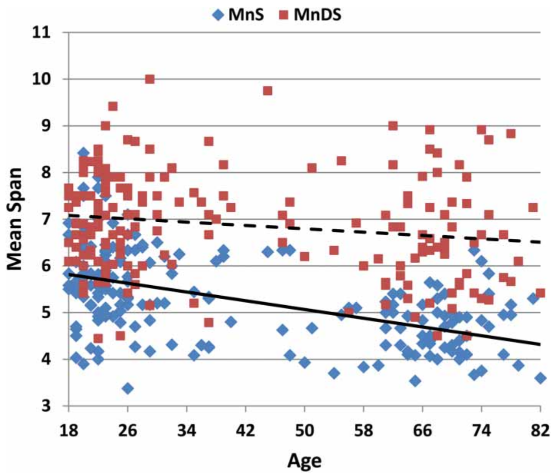

Figure 3. Mean spatial span (MnS,blue) and mean digit span (MnDS,red) for participants as a function of age. Age regression slopes are shown (dashed = digit span, solid = spatial span).

On spatial span trials, participants reported 5.95 (standard deviation = 0.92) of 14 trials correctly, including 3.14 (1.19) trials before making two errors at the same length to define their MS. MnS scores averaged 5.27 (1.01), and were slightly greater than the average MS score of 5.15 (1.20) [$F(1,186) = 3.98$, $p < .05$, partial $\omega^2 = 0.02$]. The standard deviation and coefficient of variation (CV) of the MS metric were also considerably larger than those of the MnS metric, as seen in Table 2.

Many participants produced identical scores on metrics that measured spatial span with single digit precision (i.e., MS, ML, and TotC). For example, 34.2% of participants had MS scores of 5. In contrast, the sub-digit precision of the MnS metric resulted in a more continuous distribution, with the most common MnS score (5.17) shared by only 4.3% of participants. As a result, MnS scores distinguished participants with identical MS scores. For example, among participants with MS scores of 5, MnS scores ranged from 3.53 to 6.67. As a consequence, the difference between the MS and MnS scores showed a relatively large variance (0.84), with 13.4% of participants showing MnS scores that were more than 1 digit larger than their MS scores, and 6.4% of participants showing MnS scores that were more than 1 digit smaller than their MS scores.

Participants were divided into quintiles based on MnS scores. Figure 4 shows the percentage of correct trials for each quintile at different sequence lengths, while Table 2 provides performance summaries for each quintile. The MnS scores increased from 3.95 to 6.75 for successive quintiles, with interquintile differences ranging from 0.49 to 0.95 digits and greater intersubject variance seen in the first and fifth quintiles. The slopes relating accuracy and sequence length were similar for different quintiles, with accuracy decreasing by approximately 30% per additional item from 1.0 digit below the MnS to 1.0 digit above.

The estimated 30%/item slope accords well with the results of Brunetti et al. (2014), who found that participants accurately reported 20–30% of sequences one digit above their MS. It is also in agreement with the results of Farrell Pagulayan et al. (2006), who measured a maximal span of 7.1 when participants recalled at least one trial in five presentations at a given list length, vs. a baseline span of 4.9 when participants correctly reported four trials in a row. Probability analysis shows that in order to accurately report at least one trial in five, participants would need to be correct on 13% of individual trials [i.e., $50\% = 1 - (0.87)^5$], while in order to identify four trials in a row, participants would need to be correct on 84% of individual trials [i.e., $50\% = 0.84^4$]. In Figure 4, the difference in accuracy on individual trials (71%) is equal to a 2.3-item difference in list length, similar to the 2.2 digit difference observed between maximal and baseline spans measured by Farrell Pagulayan et al. (2006). Figure 4 also shows that participants were less than 100% accurate at span lengths considerably shorter than their MnS. For example, participants in the 5th quintile (MnS = 6.75) produced occasional errors at lengths 3 and 4. This suggests that distractability effects should be included in any model of SST performance (see below).

Response times (ReTs)

Mean ReTs were obtained by averaging response latencies for each selection. Table 2 includes mean ReTs from the different quintiles, and Figure 5 shows mean ReTs for individual participants as a function of age. Consistent with previous reports (Brunetti et al., 2014), participants took longer to select the first square in a sequence (mean 2398 ms, sd = 748 ms) than to select later squares (mean 1578 ms, sd = 378 ms) [$F(1,186 = 4443.65$, $p < .0001$, partial $\omega^2 = 0.70$], presumably reflecting the increased time needed to plan the response sequence (Hurlstone et al., 2014). Mean ReTs averaged 1742 ms (standard

Table 2. Spatial span performance metrics for all subjects and for subjects in different quintiles of performance.

	TotC	ML	MS	MnS	MnS-z	ReT	L-ReT-z	MnDS	MnDS-z
All	5.95 (0.92)	5.93 (1.01)	5.15 (1.20)	5.27 (1.01)	0.00 (1.00)	1742 (426)	0.00 (1.00)	6.87 (1.11)	0.00 (1.00)
Q1	5.05 (0.69)	4.72 (0.56)	3.97 (0.93)	3.95 (0.40)	−1.15 (0.72)	2040 (641)	0.42 (1.25)	6.21 (1.03)	−0.46 (0.86)
Q2	5.58 (0.69)	5.55 (0.56)	4.82 (0.69)	4.73 (0.20)	−0.28 (0.62)	1846 (381)	−0.04 (0.93)	6.80 (1.02)	0.09 (0.96)
Q3	5.80 (0.63)	5.80 (0.53)	5.11 (0.87)	5.22 (0.14)	−0.01 (0.59)	1590 (311)	0.01 (1.12)	6.78 (1.06)	−0.08 (0.98)
Q4	6.49 (0.56)	6.41 (0.50)	5.38 (0.88)	5.80 (0.23)	0.26 (0.49)	1419 (154)	−0.19 (0.70)	7.17 (1.10)	0.15 (1.09)
Q5	6.89 (0.71)	7.25 (0.60)	6.56 (0.91)	6.75 (0.55)	1.28 (0.65)	1373 (159)	−0.22 (0.80)	7.43 (1.01)	0.32 (0.96)

Notes: TotC, total correct; ML, maximum length reported; MS, Corsi Block span, maximal length before missing two successive trials; MnS, mean spatial span; MnS-*z*, age- and computer-use regressed MnS *z*-score; ReT, mean response time per item (ms); L-ReT-*z*, *z*-score of age- and computer-use regressed, log transformed ReT; MnDS, mean digit span estimated using procedures from Woods et al. (2011); MnDS-*z*, age- and computer-used regressed MnDS. Standard deviations are shown in parentheses.

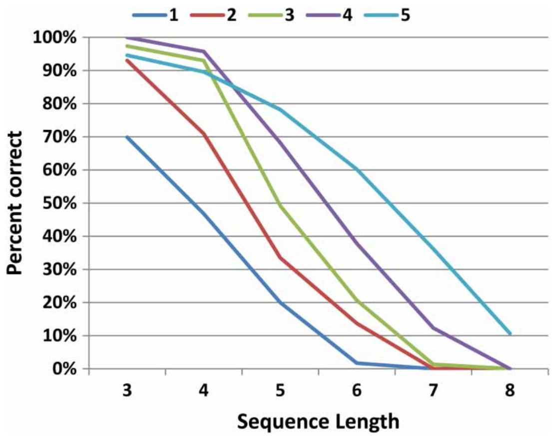

Figure 4. Percent of correct reports at different sequence lengths by quintile. Quintiles were defined by mean spatial span (MnS) scores.

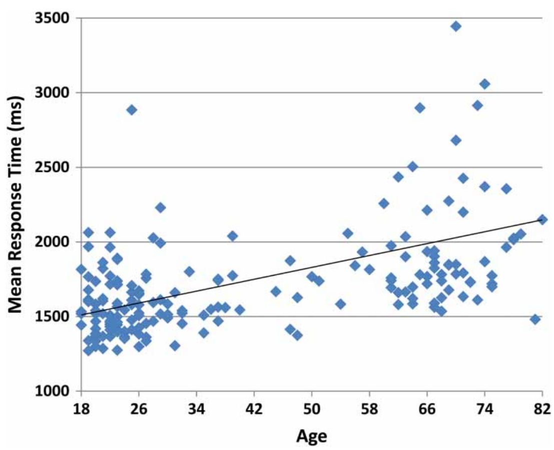

Figure 5. Mean response times (ReTs) as a function of age. Response times were averaged over responses to all items in the display.

deviation = 426 ms) and increased with age [$r = 0.50$, $t(185) = 7.85$, $p < .0001$] at a rate of 9.4 ms/year. ReTs also showed a strong negative correlation with MnS [$r = -0.46$, $t(185) = -7.05$, $p < .0001$].

ReTs were positively correlated with the results of other processing speed tests administered on the same day of testing, including simple reaction time [$r = 0.37$, $t(181) = 5.36$, $p < .0001$] (Woods, Wyma, Yund, Herron, & Reed, 2015b), choice reaction time [$r = 0.43$, $t(171) = 6.23$, $p < .0001$] (Woods, Wyma, Yund, Herron, & Reed, 2015a), completion times for Trail Making Test part A [$r = 0.60$, $t(162) = 9.55$, $p < .0001$] and part B [$r = 0.47$, $t(162) = 6.78$, $p < .0001$] (Woods, Wyma, Yund, & Herron, 2015c), and the time per question to complete a self-paced questionnaire [$r = 0.63$, $t(150) = 10.06$, $p < .0001$] (Woods, Yund, Wyma, Ruff, & Herron, 2015d). The correlations between measures of mouse movement speed (e.g., Trails A and question-completion time) exceeded that observed for simple and choice reaction time measures [minimum $z = 2.24$, $p < .03$]. MnS scores also correlated negatively with all of these additional measures of processing speed [range $r = -0.28$ to $r = -0.40$, $p < .0002$ for all comparisons].

The ReT distribution was positively skewed (skew = 2.98), so ReTs were log-transformed before further analysis. Both age and computer-use independently influenced log-transformed ReT (log-ReT) measures [total $r = 0.60$, age: $t(185) = = 8.06$, $p < .0001$; computer-use: $t(185) = -4.58$, $p < .0001$]. log-ReT *z*-scores were calculated from log-ReT values corrected for age and computer-use using multiple linear regression.

Factors influencing spatial span

Table 3 shows the correlation matrix for demographic variables and metrics. Different spatial span metrics were strongly correlated with each other, with the MnS metric showing stronger average correlations with other C-SST metrics (mean $r = 0.79$) than with the MS metric (mean $r = 0.59$). Age showed significant negative correlations with all of the span metrics, with the strongest correlation seen for the MnS [$r = -0.49$, $t(185) = 7.65$, $p < .0001$]. MnS scores decreased with a slope of −0.023/year; for example, the predicted MnS of an average 70 year-old would be reduced by 1.17 digits with respect to that of an average 20 year-old. Age also showed a small positive correlation with education [$r = 0.17$, $t(185) = 2.35$, $p < .02$] and a negative correlation with computer-use [$r = -0.22$, $t(185) = -3.08$, $p < .003$].

No significant sex differences were seen for any metric. Education showed only a weak correlation with MnS [$r = 0.13$, $t(185) = 1.78$, $p < .05$, one-tailed]. However, computer-use was more strongly associated with performance, with the strongest correlation seen with MnS [$r = 0.28$, $t(185) = 3.97$, $p < .0002$]. Multiple regression analysis showed that both age and computer-use independently influenced MnS [total $r = 0.52$, age: $t(185) = = 7.07$, $p < .0001$; computer-use: $t(185) = 2.78$, $p < .006$]. MnS *z*-scores, corrected for age and computer-use, were negatively correlated with log-ReT *z*-scores [$r = -0.27$, $t(185) = -3.814$, $p < .0002$], and also showed negative correlations with age- and computer-use corrected *z*-scores on the other processing speed tests [range $r = -0.11$ to $r = -0.34$].

Digit span and spatial span

MnS and MnDS were positively correlated [$r = 0.38$, $t(185) = 5.59$, $p < .0001$]. MnDS scores averaged 6.87 (sd = 1.11), significantly above MnS spans [$F(1,186) = 340.01$, $p < .0001$, partial $\omega^2 = 0.65$]. Like MnS, MnDS was negatively correlated with age [$r = -0.17$, $t(185) = -2.35$, $p < .02$, one-tailed]. However, the correlation was weaker [$z = 2.19$, $p < .03$] than the correlation between MnS and age, and showed a shallower slope (−0.009 digits/year), as seen in Figure 3. As a result, the difference between MnDS and MnS (mean = 1.60, standard deviation = 1.19) increased significantly with age [$r = 0.26$, $t(185) = 3.66$, $p < .0005$].

MnDS correlated significantly with computer-use [$r = 0.28$, $t(185) = 3.97$, $p < .0002$], but was only minimally affected by education [$r = 0.14$, $t(185) = 1.92$, $p < .03$, one-

Table 3. Correlation matrix for spatial span.

	C-use	Educ	Sex	TotC	ML	MS	MnS	MnS-Z	ReT	ReT-Z	MnDS
Age	−0.22	0.17	0.10	−0.44	−0.48	−0.33	−0.49	0.00	0.50	0.00	−0.17
C-use		0.34	0.03	0.19	0.26	0.27	0.28	0.00	−0.39	0.00	0.28
Educ			0.05	0.08	0.12	0.15	0.13	0.18	−0.06	−0.02	0.14
Sex				−0.12	−0.07	−0.05	−0.09	−0.05	0.02	−0.03	−0.07
TotC					0.68	0.39	0.77	0.63	−0.44	−0.25	0.32
ML						0.66	0.89	0.73	−0.46	−0.24	0.34
MS							0.72	0.61	−0.34	−0.15	0.31
MnS								0.85	−0.47	−0.23	0.38
MnS-Z									−0.21	−0.27	0.29
ReT										0.79	−0.32
ReT-Z											−0.19

Notes: C-use, computer use; Educ, education. See Table 2 for other abbreviations. Given the sample size ($N = 189$), correlations at $|r| > 0.19$ are significant at $p < .01$, uncorrected.

tailed] and not significantly influenced by sex [$r = -0.07$]. ReTs in the C-SST were negatively correlated with MnDS scores [$r = -0.32$, $t(185) = 4.61$, $p < .0001$]; that is, participants with shorter ReTs had longer digit spans, as well as longer spatial spans. Multiple regression was used to analyse the combined effects of age and computer-use on MnDS scores: computer-use exerted a significant influence [$t(184) = 3.53$, $p < .0006$], but the age factor failed to reach significance [$t(184) = -1.58$, $p < .12$]. Regression coefficients were used to calculate MnDS *z*-scores, corrected for age and computer-use, which correlated with MnS *z*-scores [$r = 0.31$, $t(185) = 4.44$, $p < .0001$] and showed significant negative correlations with log-ReT *z*-scores [$r = -0.20$, $t(185) = -2.78$, $p < .007$].

Serial position analysis

Figure 6 shows the probability of errors at each serial position for spatial spans at list lengths of 5 and 6 and for digit spans at lengths 6 and 7. Spatial span testing showed small primacy and recency effects; errors occurring mid-list were somewhat more common than errors in the initial and final locations. However, serial position effects were much larger in digit span testing, with fewer errors in the initial and final

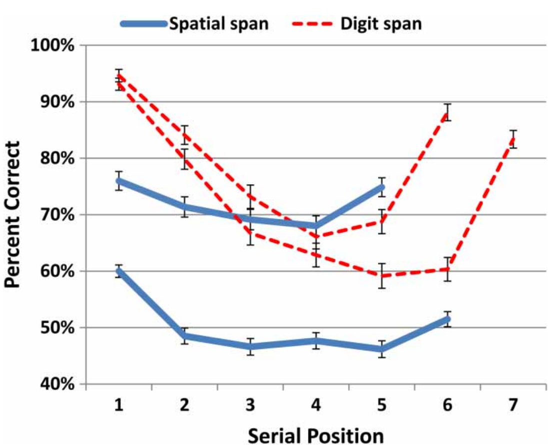

Figure 6. Serial position functions in spatial span and digit span testing. Percent correct at each position in list lengths of five and six for spatial span testing and list lengths of six and seven for digit span testing. Error bars show standard errors of the mean.

positions than in mid-list positions. Statistical comparisons confirmed significant differences in the serial positions of errors in digit span and spatial span testing for lists of length five through eight. Specifically, we counted the errors at each position in both digit span and spatial span trials of the same length for participants who completed both tests and found that Chi-squared independence tests failed for trials of length five [$\chi^2_{(4)} = 24.75,\ p < .0001$], six [$\chi^2_{(5)} = 89.57,\ p < .0001$], seven [$\chi^2_{(6)} = 72.01,\ p < .0001$], and eight [$\chi^2_{(7)} = 23.65,\ p < .003$].

Types of errors

Table 4 shows the percentage of the different types of errors made during spatial and digit span testing. Transposition and omission errors predominated during spatial span testing. More than 80% of transposition errors occurred when participants transposed successive items (e.g., the "1-6-5-2-8" became "1-6-2-5-8"). In addition, transposition errors were somewhat more common when items were spatially close to each other in the display. For example, on trials where only a single transposition error occurred, the screen distance between transposed locations was smaller than the mean distance to other transposable locations [paired *t*-test (actual vs. mean possible), $t(373) = -3.69$, $p < .001$].

Errors of all types increased with increasing list length, with transposition and omission errors increasing in parallel during spatial span testing. At most list lengths, addition errors, substitution errors, and permutation errors occurred at rates that were more than five standard errors of the mean below the average rates for omission and transposition errors.

Table 5 shows the percentage of omission and transposition errors at serial positions during digit span and spatial span testing. In both tests, omission errors were rare in the first position and increased throughout the list, declining in the final position. However, transposition errors showed a different pattern in the two tests. In the C-SST, transposition errors were relatively common in the first position and slightly increased over the sequence before declining to minimal values in the final position. On the other hand, transposition errors in digit span testing remained rare in

Table 4. Types of errors (in mean percent per trial) in spatial span (top) and digit span (bottom) for trials of different lengths.

Length	N	PC	Omit	Add	Sub	Trans	Perm
Spatial span							
3	247	87.0% (2.2%)	2.8% (1.4%)	4.9% (1.0%)	2.4% (1.5%)	6.1% (1.5%)	0.0% (0.0%)
4	439	69.0% (2.2%)	11.4% (1.4%)	9.3% (2.0%)	12.5% (1.5%)	11.4% (1.6%)	2.5% (0.8%)
5	716	43.2% (1.9%)	24.7% (1.4%)	15.6% (1.7%)	19.7% (1.7%)	25.3% (1.7%)	6.9% (1.0%)
6	664	29.0% (1.8%)	39.9% (1.7%)	20.5% (2.1%)	25.6% (2.1%)	37.5% (2.1%)	13.6% (1.4%)
7	385	18.7% (2.0%)	55.6% (2.3%)	22.1% (2.6%)	26.5% (3.2%)	50.7% (3.5%)	21.8% (2.2%)
Digit Span (Woods et al., 2011)							
5	322	81.1% (2.2%)	4.7% (1.3%)	3.1% (1.0%)	3.7% (1.2%)	10.9% (1.7%)	1.2% (0.6%)
6	522	53.6% (2.2%)	23.2% (2.9%)	13.2% (1.5%)	8.2% (1.5%)	23.4% (2.0%)	7.5% (1.2%)
7	595	39.2% (2.0%)	45.2% (4.30%)	16.9% (1.7%)	9.8% (1.4%)	33.1% (2.2%)	13.1% (1.4%)
8	475	24.4% (2.0%)	88.8% (7.4%)	18.3% (2.0%)	11.4% (1.6%)	38.3% (2.7%)	21.7% (1.9%)
9	225	18.2% (2.6%)	80.4% (7.8%)	24.0% (3.3%)	20.0% (3.2%)	48.0% (4.2%)	32.9% (3.3%)

Notes: *N*, number of trials; PC, percentage correct; Omit, omission; Add, addition; Sub, substitution; Trans, transposition; Perm, permutation. Numbers in parentheses show standard errors of the mean. Multiple errors could occur in a single trial.

Table 5. Percentages of transposition (Trans) and omission (Omit) errors.

Length	Position	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	7 (%)	8 (%)
Spatial span									
5	Trans	10.3	13.8	11.5	9.5	5.5			
	Omit	3.1	4.1	5.5	7.4	4.6			
6	Trans	10.2	16.3	15.5	12.1	12.4	8.1		
	Omit	2.0	4.7	5.9	8.3	10.5	8.4		
7	Trans	13.4	17.7	13.5	16.9	16.9	14.6	8.6	
	Omit	3.4	4.4	6.8	10.7	9.9	10.1	10.4	
Digit span (Woods et al., 2011)									
6	Trans	0.6	5.9	11.7	16.3	10.2	2.1		
	Omit	0.8	3.8	3.6	5.8	7.1	2.1		
7	Trans	1.0	5.9	12.3	15.0	15.8	13.3	3.0	
	Omit	2.2	5.2	8.1	7.4	8.1	8.6	5.7	
8	Trans	2.1	6.1	9.5	12.6	16.2	15.0	13.1	2.1
	Omit	4.2	8.2	11.2	12.2	11.6	12.8	19.0	9.7

Notes: Shown for different serial positions for trials of different length during spatial span (top) and digit span (bottom) testing.

the first and last positions relative to mid-list positions, even in long lists. Chi-squared independence tests revealed a significant difference in the serial position functions of transposition errors in spatial and digit span testing in 6-item lists [$\chi^2_{(5)} = 12.67$, $p < .03$], with a trend towards significant differences in 7-item lists [$\chi^2_{(6)} = 7.52$, $p < .06$].

Factors affecting trial difficulty: distance, crossings, and clustering

Square positions varied pseudo-randomly from trial-to-trial, and item sequence selection was random during spatial span testing. As a result, trials of similar sequence lengths were associated with differences in path distance and number of path crossings. Table 6 shows the mean and range of path distance and number of crossings at different sequence lengths, and shows the correlations of both measures with trial outcome and trial completion time. Path distance did not correlate significantly with outcome, except for a weak trend at sequence length 5 [$r = -0.07$, $t(607) = -1.73$, $p < .05$, one-tailed], and path crossings were weakly related to outcome only at sequence lengths of 4 [$r = -0.07$, $t(952) = -2.17$, $p < .05$] and 7 [$r = -0.10$, $t(629) = -2.52$, $p < .02$]. In contrast, both distance and the number of crossings were associated with increased ReTs at all sequence lengths (range $r = 0.08$ to $r = 0.18$).

Two additional spatial proximity effects were also noted. First, when participants made substitution errors, they tended to select squares in the display that were spatially close to squares that had actually appeared in the sequence. In trials where only a single substitution error occurred, the distance between the target square and substituted square was smaller than the mean distance to all other potentially substitutable locations [$t(205) = -7.40$, $p < .0001$]. Second, clustering exerted a small but significant effect on the difficulty of different trials. In a regression that included sequence lengths adjusted for each

Table 6. Correlations of trial properties with trial outcomes for different sequence lengths.

Length	N	%hit	Distance	Crossings	CT (s)	hit/Dist	hit/Cross	ReT/Dist	ReT/Cross
4	609	69.2%	35.7 (16.3–66.5)	0.23 (0–1)	7.06	−0.07	0.01	0.18	0.10
5	954	44.9%	48.0 (23.8–85.3)	0.71 (0–3)	8.74	−0.05	−0.07	0.13	0.11
6	954	32.7%	59.3 (33.1–101.4)	1.41 (0–6)	9.68	−0.04	0.01	0.14	0.11
7	631	20.3%	71.4 (43.8–111.6)	2.44 (0–10)	10.66	−0.05	−0.10	0.08	0.10

Notes: Mean values for distance and number of crossings are shown with the range in parentheses. The final four columns show the correlations of outcome (hit) with distance (Dist) and crossings (Cross), and responses times (ReT) with distance and crossings.

TABLE 7. Simulations showing the relationship between different metrics and procedures in spatial span testing.

True span	CBT-MS	WMS tot	Orsini	Capitani	TotC	ML	MS	MnS	Last5
3.0	2.87 (*0.87*)	2.86 (*1.37*)	2.34 (*0.67*)	2.28 (*0.76*)	4.49 (*0.61*)	3.74 (*0.58*)	3.00 (*0.76*)	3.15 (*0.44*)	3.72 (*0.64*)
3.5	3.36 (*0.81*)	3.85 (*1.33*)	2.84 (*0.66*)	2.79 (*0.75*)	4.80 (*0.57*)	4.10 (*0.48*)	3.37 (*0.78*)	3.57 (*0.40*)	4.16 (*0.57*)
4.0	3.86 (*0.91*)	4.71 (*1.48*)	3.32 (*0.70*)	3.24 (*0.82*)	5.16 (*0.59*)	4.68 (*0.59*)	3.88 (*0.87*)	4.08 (*0.47*)	4.71 (*0.65*)
4.5	4.33 (*0.84*)	5.72 (*1.41*)	3.85 (*0.68*)	3.76 (*0.80*)	5.48 (*0.58*)	5.05 (*0.49*)	4.36 (*0.82*)	4.53 (*0.43*)	5.14 (*0.57*)
5.0	4.85 (*0.93*)	6.61 (*1.56*)	4.32 (*0.70*)	4.21 (*0.88*)	5.84 (*0.59*)	5.65 (*0.61*)	4.85 (*0.91*)	5.05 (*0.50*)	5.69 (*0.65*)
5.5	5.33 (*0.86*)	7.64 (*1.50*)	4.83 (*0.69*)	4.75 (*0.86*)	6.16 (*0.56*)	6.03 (*0.50*)	5.34 (*0.83*)	5.52 (*0.47*)	6.13 (*0.59*)
6.0	5.83 (*0.97*)	8.49 (*1.69*)	5.31 (*0.73*)	5.18 (*0.93*)	6.52 (*0.60*)	6.60 (*0.62*)	5.85 (*0.92*)	6.02 (*0.53*)	6.69 (*0.66*)
6.5	6.32 (*0.91*)	9.56 (*1.68*)	5.82 (*0.71*)	5.71 (*0.95*)	6.84 (*0.56*)	6.98 (*0.51*)	6.32 (*0.87*)	6.49 (*0.49*)	7.13 (*0.59*)
7.0	6.81 (*1.01*)	10.44 (*1.79*)	6.29 (*0.75*)	6.14 (*1.05*)	7.21 (*0.59*)	7.54 (*0.62*)	6.82 (*0.97*)	7.01 (*0.57*)	7.69 (*0.65*)
7.5	7.30 (*0.96*)	11.40 (*1.87*)	6.84 (*0.73*)	6.67 (*1.06*)	7.52 (*0.57*)	7.93 (*0.53*)	7.31 (*0.90*)	7.47 (*0.52*)	8.11 (*0.59*)
8.0	7.79 (*1.04*)	12.07 (*1.99*)	7.29 (*0.77*)	7.08 (*1.17*)	7.91 (*0.60*)	8.49 (*0.63*)	7.79 (*0.97*)	7.97 (*0.58*)	8.64 (*0.62*)
8.5	8.23 (*0.95*)	12.89 (*2.00*)	7.81 (*0.75*)	7.62 (*1.16*)	8.24 (*0.60*)	8.87 (*0.55*)	8.27 (*0.88*)	8.44 (*0.53*)	9.05 (*0.52*)

Notes: Each simulation was based on a "true span", the length where subjects were 50% correct. Each cell represents the result of 10,000 simulations using the termination rule of each test. Numbers in parentheses show standard deviations. The model incorporated a slope of 30%/item, centered on the true spans shown. Subjects were assumed to miss 5% of trials that would otherwise be correct due to attentional lapses. CBT-MS, Corsi Block Test maximal span; WMS tot, Total forward span on the WMS III SST; Orsini, maximal span for three of five trials correct; Capitani, maximal span for two of three trials correct. The last five columns show the metrics of the current test (see Table 2 for abbreviations).

participant's MnS, path distance, path crossings, path angles, and the minimum distance between display elements, sequence length produced the expected highly significant effect [$t(3421) = 31.18$, $p < .0001$], while among the other factors, only minimal distance reached weak statistical significance [$t(3421) = 2.85$, $p < .01$].

Discussion

An improved test of visuospatial span

We describe a brief computerised test of visuospatial span, the C-SST, that uses adaptive procedures to adjust list lengths in order to repeatedly sample spans above and below report thresholds. The C-SST also provides response time measures of processing speed that were shown to correlate with both spatial and digit span. The psychophysically-based MnS metric was shown to have several superior properties in comparison to traditional maximal span and total correct metrics, including reduced standard deviations and higher correlations with demographic variables that influence spatial span (e.g., age). The MnS metric also has higher test-retest reliability than maximal span or total correct metrics (Woods, Wyma, Herron, & Yund, submitted). In addition, simulation studies show that the MnS metric provides a more accurate estimate of participants' true span than other metrics (see below).

Simulating performance in different spatial-span paradigms

We performed simulation studies in order to further elucidate the differences between the span metrics. We simulated performance in different paradigms by assuming that the probability of correct report changed as a logistic function with a slope of 30% per item (e.g., Figure 4) relative to the participant's true span (i.e., 50% correct). These simulations enabled the comparison of maximal spans reported for variants of the CBT that use different sequence delivery rules, and also enabled the comparison of studies using similar sequence delivery rules but different span metrics (e.g., total correct vs. maximal span) (Berch et al., 1998).

Table 7 shows the results of such simulations for true spans ranging from 3.0 to 8.5. Ten thousand simulations were performed at each true span following the termination rules for each test, calculating performance independently for each trial, and assuming that accuracy never exceeded 95%/trial, even on short lists, due to inattention (see Figure 4). Several important results are evident. The MnS metric provides accurate estimates of true span (range −0.06 to 0.15) at all simulated true spans, with small standard deviations (range 0.40 to 0.58). In contrast, the CBT maximal span underestimates the true span by increasing amounts as true span increases (range −0.13 to −0.27), with standard deviations that are approximately twice as large as those of the MnS metric. More serious underestimations occur when the maximal span is estimated using the "3 of 5" termination rule of Orsini (range −0.65 to −0.71) and the "2 of 3" termination rule of Capitani (range −0.71 to −0.92).

The reason for the systematic underestimation of true span with the CBT is that testing often ceases prematurely. For example, in the traditional CBT, a participant with a true span of 5 will miss two length-5 trials on 25% of presentations, resulting in a maximal span of 4. Thus, fewer than 75% of participants will be tested at length 6, where two failures would produce a span of 5. In the CBT variants of Orsini et al. (1986) and Capitani et al. (1991), the undersampling bias is more severe. In these tests, the maximal span is the longest sequence correctly reported on two of three or three of five presentations, respectively. Since a participant with a true span of N has only a 50% chance of producing two of three (or three of five) correct reports in a length-N sequence, 50% of participants with true spans of 5 will be scored as having spans of 4; that is, testing at length 6 (where failure would produce a maximal span of 5) would occur in less than 50% of participants.

In contrast, the method of Farrell Pagulayan et al. (2006) (not included in Table 7) would produce a maximal span well above the true span. They quantified maximal span as the longest sequence correctly reported on any of five sequences, and terminated testing at list lengths where participants missed all five trials. Participants only need to be 13% correct on individual trials in order to have a 50% chance of detecting one trial in five. In other words, assuming a slope of 30% per item around the true span, the maximal span reported by Farrell Pagulayan et al. (2006) would be slightly more than one digit above true span.

The standard deviation of MS measures was also much larger than the standard deviation of the MnS metric. This difference in standard deviations is due to three factors. First, the MnS metric in the C-SST is based on trials that are both above and below the true span, while MS primarily samples sub-span list lengths. Second, the C-SST always includes 14 trials and thus avoids early test termination. In contrast, the number of trials presented in other tests varies with span. For example, a participant with a true span of 5.0 will receive an average of 8.61 trials in the standard CBT, and 22.5 trials in the Orsini variant of the CBT, where testing ceases after missing three of five trials. The larger number of trials explains the slightly reduced variance of the MS metric in the Orsini paradigm when compared to the standard CBT paradigm. Third, MnS variance is reduced relative to that of MS because MnS is calculated with sub-digit precision, which more closely corresponds to the continuous underlying distribution of true span. This reduces the MnS standard deviation compared to that of MS, even when a larger number of trials is used to estimate maximal span.

The Wechsler total correct metric of performance used in the WMS-III SST was problematic for three reasons. First, the total correct metric conflates inconsistent performance with memory capacity. For example, a participant who misses one trial at lengths of 2, 3, 4, 5, 6, and 7, and then fails twice at list length 8, will have the same total correct score as another participant who accurately reports all trials at lengths 2, 3, and 4, but then fails twice at list length 5. Second, the variance of the total correct metric is increased, like that of the MS metric, because of the 2-miss termination rule. As a result, different participants receive different numbers of trials. Third, the CV is higher for the total correct metric than the MS metric because participants perform imperfectly even at lengths well below their true spans. For example, for participants with true spans of 5, the CV is 24% for the total correct metric, 19% for the MS metric, and 10% for the MnS metric. Thus, it is unsurprising that the CVs of the published studies using the total correct metric are high (Table 1). This reduces sensitivity to clinical abnormalities. For example, the reported total correct score for a group of 10 patients with Korsakoff's syndrome was 19% lower than that of the control group, but the differences failed to reach statistical significance due in large part to the high CV (32%) of the control means (Wilde et al., 2004).

The simulation results shown in Table 7 also permit the results from different SST paradigms to be translated into true span scores, as shown in column 5 of Table 1. Such comparisons reveal that the true spans estimated from the WMS total correct metric are similar to the true spans estimated from the CBT MS metric for participant groups of similar age (e.g., Lo et al., vs. Monaco et. al). Converting different MS scores into true spans also reduces apparent discrepancies in results. For example, the average MS reported by Orsini et al. (1986) differed from that of Farrell Pagulayan et al. (2006) by more than 2.5 digits (4.56 vs. 7.1). However, when translated into true spans, the differences were reduced to less than 1.0 digit.

Translating the results into true spans makes it possible to perform meta-analyses of previous studies without the confusions introduced by different span metrics. Meta-analysis revealed a strong correlation between the mean population ages of the participant groups shown in Table 1 and their estimated true spans [$r = 0.74$, $t(16) = 4.01$, $p < .002$]. Column five shows the age-predicted true spans based on the age-regression slope (−0.024/year) observed in Table 1 data. Correcting for age resulted in a further reduction of discrepancies between studies. For example, the studies of Orsini et al. (1986) and Farrell Pagulayan et al. (2006) had MS scores that differed by 2.5 digits and true spans that differed by 0.76 digits. However, these differences largely disappeared following age-correction: the observed true spans were only slightly less than the age-predicted values in both studies (−0.30 and −0.38 digits, respectively). Indeed, all the true span estimates from CBT and WMS data were within 0.5 digits of age-predicted means, with the exception of the manual test administered by Claessen et al. (2014), which used simplified, non-standard paths, and the WMS test of Wiechmann et al. (2011), who studied highly educated, older female participants.

However, the estimated true spans in the current experiment (5.23) showed a relatively large discrepancy with the age-predicted true span (5.95) that would be expected in the CBT for participants with a mean age of 41.1 years. This discrepancy suggests that the C-SST was more difficult than the standard CBT. Increased C-SST difficulty is also consistent with the unusually large difference that we observed between forward digit span and spatial span (1.60 digits) compared to the digit-span vs. spatial-span differences reported in previous studies (0.50 to 1.1 digits) (Kessels et al., 2008; Monaco et al., 2013; Orsini et al., 1986; Wilde et al., 2004).

Several factors may account for the increased difficulty of the C-SST. First, computerised SSTs may be generally more difficult than manual tests (Claessen et al., 2014) due to differences in visuospatial memory for 2D displays in comparison with physical objects, or because less familiar responses are required (e.g., moving the mouse vs. touching an object). Second, the C-SST block layout varied randomly from trial to trial. As a result, participants were first exposed to the block layout when the trial began. In contrast, an identical layout is used in the CBT, with the blocks remaining visible between trials. Third, the non-random paths used in the CBT may be easier than the random paths used in the C-SST.

The influence of age on performance

As in previous reports (Capitani et al., 1991; Fournet et al., 2012; Kessels et al., 2000; Lo et al., 2012; Monaco et al., 2013; Orsini et al., 1986; Park et al., 2002; Wilde et al., 2004), we found that increasing age was associated with reductions in spatial span. The slope that related age with MnS scores on the C-SST (−0.023 digits/year) was virtually identical to the average slope that related the mean population age with true span in the studies summarised in Table 1 (−0.024 digits/year). It was also similar to the slopes relating age and estimated true span in the studies of Wilde et al. (2004) and Monaco et al. (2013) (−0.021 and −0.025, respectively). However, it was steeper than the slope (−0.015) estimated from the studies of Orsini et al. (1986) and Lo et al. (2012), and shallower than the slope (−0.043) from the study of Fournet et al. (2012).

The age slope in the current study was well-fit by linear regression. As in previous reports (Orsini et al., 1986; Park et al., 2002), spatial span and digit span scores showed significant positive correlations. However, we found significantly greater age-related declines in spatial span than in digit span, consistent with the results of some previous studies (Orsini et al., 1986; Park et al., 2002; Wilde et al., 2004), but not others (Kessels et al., 2008; Monaco et al., 2013).

The influence of education, sex, and computer-use on visuospatial span

We found only minimal influences of education on C-SST performance. This likely reflects the fact that most of our participants were well educated (minimum 10 years of formal education, mean 14.6 years). Previous studies of participants with a broader range of education (e.g., grade school to college) have generally found stronger correlations between education and performance (Capitani et al., 1991; Fournet et al., 2012; Kessels et al., 2008; Monaco et al., 2013; Orsini et al., 1986).

We found that computer-use was more strongly correlated with both MnS and ReT measures than was education. While increased familiarity with computers would be expected to facilitate performance on computerised tasks, increased computer-use may also index a more general level of current intellectual engagement among participants. Thus, we also found a higher correlation between computer-use and digit span than between education and digit span, even though digit span was measured with standard verbal report.

We found no significant effects of sex on any performance metric. While studies of less well-educated participants have often reported superior spatial spans in male participants (Capitani et al., 1991; Fournet et al., 2012; Grossi, Matarese, & Orsini, 1980; Orsini et al., 1986), more highly educated participant populations generally fail to show sex differences (Kessels et al., 2008; Tamayo et al., 2012).

Processing speed and working memory capacity

Both MnS and MnDS metrics correlated with ReTs in the current task and with other measures of processing speed, including simple reaction time, choice reaction time, and completion times on trail making tests and questionnaires. The results are consistent with previous observations that increased processing speed is associated with superior recall in working memory tasks (Hurlstone et al., 2014; Park et al., 2002; Smyth & Scholey, 1996a).

Factors that influence trial difficulty

Consistent with some previous findings (Smyth & Scholey, 1994) but in contrast to others (Busch et al., 2005; Guerard & Tremblay, 2012; Orsini et al., 2004), we found only a small influence of path distance and crossings on trial difficulty. However, path distance and crossings were not evaluated parametrically. Thus, the relatively weak effects that we observed may have reflected the trial-to-trial variation in the display layout and sequence lengths, as well as the use of random paths, which can result in complex variations in Gestalt properties (e.g., continuation, symmetry, etc.) that may have influenced report accuracy independently of path distance and crossings.

Serial position effects in spatial and digit span testing

The primacy and recency effects seen in the C-SST were similar in magnitude to serial gradients in spatial memory seen in previous studies (Guerard & Tremblay, 2012; Smyth & Scholey, 1996b), and were less striking than the primacy and recency gradients seen in digit span testing, where errors on the first and last item remain rare, even in supra-span lists. These differences primarily reflected an increased incidence of errors involving the first element in spatial span lists. The results suggest that positional marking (Hurlstone et al., 2014) is less robust in the visuospatial than verbal domain. Such differences may also help to explain another modality difference. Because the positional gradient is relatively weak in spatial span, backward and forward spatial spans are comparable in length, while the strong positional marking in digit-span memory results in clear forward-span superiority (Wilde & Strauss, 2002).

Types of errors in spatial and digit span testing

Error analysis revealed similar types of errors on C-SST and digit span testing. In both cases, transposition errors predominated at shorter lengths, while omission errors became more common as lengths increased. In our paradigms, the occurrence of omission errors was increased relative to their incidence in most previous studies (Hurlstone et al., 2014) because participants were more extensively tested at list lengths above their true spans.

Conclusions

We describe an adaptive C-SST that quantifies visuospatial memory with a psychophysically-based mean spatial span (MnS) metric with sub-digit precision. The MnS metric had lower variance than other widely used, current metrics, including maximum span and total correct trials. Accuracy was strongly affected by list length, declining at a rate of approximately 30%/item around the MnS, with minimal additional influence of path distance and crossings. An analysis of incorrect trials showed that omission and transposition errors predominated. Serial position functions showed primacy and recency effects that were weaker in spatial span than digit span testing, consistent with a difference in the strength of positional marking in visuospatial and verbal domains. Visuospatial span declined more rapidly with age than digit span in the same participants, suggesting a more rapid age-related decline in spatial than verbal working memory. Faster response times were associated with increased spatial and digit spans, consistent with previously reported correlations between processing speed and working memory capacity. Simulation studies showed that the MnS metric provided more accurate estimates of true span (i.e., the list length where participants are 50% correct) than other metrics. The simulations also enabled a meta-analysis of previous studies of visuospatial span which revealed that the C-SST provides a more accurate measure of spatial working memory than the CBT or the WMS III SST.

Acknowledgements

The content is solely the responsibility of the authors and does not necessarily represent the official views of the Department of Veterans Affairs or the US Government. We would like to thank Ben Edwards, Oren Poliva, Masood Younus, Nabeel Rahman, and Kerry Hubel who gathered data used in this report, and Robert Hink who developed the MySQL database.

Disclosure statement

DLW is affiliated with NeuroBehavioral Systems, Inc., the developers of Presentation software that was used to create these experiments.

Funding

This research was supported by a VA Rehabilitation Research and Development Grants [CX000583] and [CX001000] to DLW.

References

Beblo, T., Macek, C., Brinkers, I., Hartje, W., & Klaver, P. (2004). A new approach in clinical neuropsychology to the assessment of spatial working memory: The block suppression test. *Journal of Clinical and Experimental Neuropsychology* (*Neuropsychology, Development and Cognition: Section A*), *26*(1), 105–114.

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi Block-tapping task: Methodological and theoretical considerations. *Brain and Cognition*, *38*(3), 317–338.

Brown, L. A., Brockmole, J. R., Gow, A. J., & Deary, I. J. (2012). Processing speed and visuospatial executive function predict visual working memory ability in older adults. *Experimental Aging Research*, *38*(1), 1–19. doi:10.1080/0361073X.2012.636722

Brunetti, R., Del Gatto, C., & Delogu, F. (2014). eCorsi: Implementation and testing of the Corsi Block-tapping task for digital tablets. *Frontiers in Psychology*, *5*, 939. doi:10.3389/fpsyg.2014.00939

Busch, R. M., Farrell, K., Lisdahl-Medina, K., & Krikorian, R. (2005). Corsi Block-tapping task performance as a function of path configuration. *Journal of Clinical & Experimental Neuropsychology*, *27*(1), 127–134. doi:10.1080/138033990513681

Capitani, E., Laiacona, M., & Ciceri, E. (1991). Sex differences in spatial memory: A reanalysis of block tapping long-term memory according to the short-term memory level. *Italian Journal of Neurological Science*, *12*(5), 461–466.

Claessen, M. H., van der Ham, I. J., & van Zandvoort, M. J. (2014). Computerization of the standard Corsi Block-tapping task affects its underlying cognitive concepts: A pilot study. *Applied Neuropsychology Adult*, 1–9. doi:10.1080/23279095.2014.892488

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* (Dissertation Abstracts International, 34(02), 891B). University Microfilms No. AAI05–77717.

Farrell Pagulayan, K., Busch, R. M., Medina, K. L., Bartok, J. A., & Krikorian, R. (2006). Developmental normative data for the Corsi Block-tapping task. *Journal of Clinical & Experimental Neuropsychology*, *28*(6), 1043–1052.

Fournet, N., Roulin, J. L., Vallet, F., Beaudoin, M., Agrigoroaei, S., Paignon, A., … Desrichard, O. (2012). Evaluating short-term and working memory in older adults: French normative data. *Aging and Mental Health*, *16*(7), 922–930. doi:10.1080/13607863.2012.674487

Grossi, D., Matarese, V., & Orsini, A. (1980). Sex differences in adults' spatial and verbal memory span. *Cortex*, *16*(2), 339–340.

Guerard, K., & Tremblay, S. (2012). The effect of path length and display size on memory for spatial information. *Experimental Psychology*, *59*(3), 147–152. doi:10.1027/1618-3169/a000137

Hester, R. L., Kinsella, G. J., & Ong, B. (2004). Effect of age on forward and backward span tasks. [Comparative Study]. *J Int Neuropsychol Soc*, *10*(4), 475–481. doi: 10.1017/S1355617704104037

Holdnack, J. A., Xiaobin, Z., Larrabee, G. J., Millis, S. R., & Salthouse, T. A. (2011). Confirmatory factor analysis of the WAIS-IV/WMS-IV. *Assessment*, *18*(2), 178–191. doi:10.1177/1073191110393106

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. doi:10.1037/a0034221

Kessels, R. P., van den Berg, E., Ruis, C., & Brands, A. M. (2008). The backward span of the Corsi Block-Tapping task and its association with the WAIS-III digit span. *Assessment*, *15*(4), 426–434.

Kessels, R. P., van Zandvoort, M. J., Postma, A., Kappelle, L. J., & de Haan, E. H. (2000). The Corsi Block-tapping task: Standardization and normative data. *Applied Neuropsychology*, *7*(4), 252–258.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, *116*(4 Pt 1), 2395–2405.

Lo, A. H., Humphreys, M., Byrne, G. J., & Pachana, N. A. (2012). Test-retest reliability and practice effects of the Wechsler Memory Scale-III. *Journal of Neuropsychology*, *6*(2), 212–231. doi:10.1111/j.1748-6653.2011.02023.x

Monaco, M., Costa, A., Caltagirone, C., & Carlesimo, G. A. (2013). Forward and backward span for verbal and visuo-spatial data: Standardization and normative data from an Italian adult population. *Neurological Science*, *34*(5), 749–754. doi:10.1007/s10072-012-1130-x

Mungas, D., Heaton, R., Tulsky, D., Zelazo, P. D., Slotkin, J., Blitz, D., … Gershon, R. (2014). Factor structure, convergent validity, and discriminant validity of the NIH Toolbox Cognitive Health Battery (NIHTB-CHB) in adults. *Journal of the International Neuropsychology Society*, *20*(6), 579–587. doi:10.1017/S1355617714000307

Orsini, A., Chiacchio, L., Cinque, M., Cocchiaro, C., Schiappa, O., & Grossi, D. (1986). Effects of age, education and sex on two tests of immediate memory: A study of normal subjects from 20 to 99 years of age. *Perceptual and Motor Skills*, *63*(2 Pt 2), 727–732. doi:10.2466/pms.1986.63.2.727

Orsini, A., Simonetta, S., & Marmorato, M. S. (2004). Corsi's Block-tapping test: Some characteristics of the spatial path which influence memory. *Perceptual and Motor Skills*, *98*(2), 382–388. doi:10.2466/pms.98.2.382-388

Ostrosky-Solis, F., Jaime, R. M., & Ardila, A. (1998). Memory abilities during normal aging. *International Journal of Neuroscience*, *93*(1–2), 151–162.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320.

Smirni, P., Villardita, C., & Zappala, G. (1983). Influence of different paths on spatial memory performance in the Block-Tapping test. *Journal of Clinical Neuropsychology*, *5*(4), 355–359.

Smyth, M. M., & Scholey, K. A. (1994). Characteristics of spatial memory span: Is there an analogy to the word length effect, based on movement time? *Quarterly Journal of Experimental Psychology A*, *47*(1), 91–117.

Smyth, M. M., & Scholey, K. A. (1996a). The relationship between articulation time and memory performance in verbal and visuospatial tasks. *British Journal of Psychology*, *87*(Pt 2), 179–191.

Smyth, M. M., & Scholey, K. A. (1996b). Serial order in spatial immediate memory. *Quarterly Journal of Experimental Psychology A*, *49*(1), 159–177. doi:10.1080/713755615

Tamayo, F., Casals-Coll, M., Sanchez-Benavides, G., Quintana, M., Manero, R. M., Rognoni, T., … Pena-Casanova, J. (2012). [Spanish normative studies in a young adult population (NEURONORMA young adults Project): Norms for the verbal span, visuospatial span, letter-number sequencing, trail making test and symbol digit modalities test]. *Neurologia*, *27*(6), 319–329. doi:10.1016/j.nrl.2011.12.020

Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi Blocks task. *British Journal of Psychology*, *95*(Pt 1), 57–79.

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

Wiechmann, A., Hall, J. R., & O'Bryant, S. E. (2011). The utility of the spatial span in a clinical geriatric population. *Aging Neuropsychology and Cognition*, *18*(1), 56–63. doi:10.1080/13825585.2010.510556

Wilde, N. J., & Strauss, E. (2002). Functional equivalence of WAIS-III/WMS-III digit and spatial span under forward and backward recall conditions. *Clinical Neuropsychology*, *16*(3), 322–330. doi:10.1076/clin.16.3.322.13858

Wilde, N. J., Strauss, E., & Tulsky, D. S. (2004). Memory span on the Wechsler Scales. *Journal of Clinical & Experimental Neuropsychology*, *26*(4), 539–549.

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical & Experimental Neuropsychology*, *33*(1), 101–111.

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Hink, R. F., & Reed, B. (2011). Computerized analysis of error patterns in digit span recall. *Journal of Clinical & Experimental Neuropsychology*, *33*(7), 721–734. doi:10.1080/13803395.2010.493149

Woods, D. L., Wyma, J., Herron, T. J., & Yund, E. W. (submitted). The effects of repeated testing, malingering, and brain injury on computerized measures of visuospatial memory span. *Frontiers in Human Neuroscience*, Submitted.

Woods, D. L., Wyma, J. M., Yund, E. W., & Herron, T. J. (2015c). The effects of aging, malingering, and traumatic brain injury on computerized trail-making test performance. *PLoS ONE*, *10*(6), e0124345. doi:10.1371/journal.pone.0124345

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015a). Age-related slowing of response selection and production in a visual choice reaction time task. *Frontiers in Human Neuroscience*, *9*, 193. doi:10.3389/fnhum.2015.00193

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015b). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience i*, *9*, 131. doi:10.3389/fnhum.2015.00131

Woods, D. L., Yund, E. W., Wyma, J. M., Ruff, R., & Herron, T. J. (2015d). Measuring executive function in control subjects and TBI patients with question completion time (QCT). *Frontiers in Human Neuroscience*, *9*, 288. doi:10.3389/fnhum.2015.00288