Assessing Working Memory Capacity Through Time-Constrained Elementary Activities

LUCIDI, Annalisa, et al.

DOI: 10.1080/00221309.2013.870121
Assessing Working Memory Capacity Through Time-Constrained Elementary Activities

Annalisa Lucidi, Vanessa Loaiza, Valérie Camos & Pierre Barrouillet


To link to this article: http://dx.doi.org/10.1080/00221309.2013.870121
Assessing Working Memory Capacity Through Time-Constrained Elementary Activities

ANNALISA LUCIDI
Université de Genève

VANESSA LOAIZA
VALÉRIE CAMOS
Université de Fribourg

PIERRE BARROUILLET
Université de Genève

ABSTRACT. Working memory (WM) capacity measured through complex span tasks is among the best predictors of fluid intelligence (Gf). These tasks usually involve maintaining memoranda while performing complex cognitive activities that require a rather high level of education (e.g., reading comprehension, arithmetic), restricting their range of applicability. Because individual differences in such complex activities are nothing more than the concatenation of small differences in their elementary constituents, complex span tasks involving elementary processes should be as good of predictors of Gf as traditional tasks. The present study showed that two latent variables issued from either traditional or new span tasks involving time-constrained elementary activities were similarly correlated with Gf. Moreover, a model with a single unitary WM factor had a similar fit as a model with two distinct WM factors. Thus, time-constrained elementary activities can be integrated in WM tasks, permitting the assessment of WM in a wider range of populations.

Keywords: fluid intelligence, individual differences, working memory

FOLLOWING BADDELEY’S (2000, p. 418) DEFINITION, cognitive psychology generally considers working memory (WM) as “a limited-capacity system allowing the temporary storage and manipulation of information necessary for such complex tasks as comprehension, learning, and reasoning.” These defining features are well reflected in the tasks used to assess WM capacity, known as complex span tasks. Indeed, contrary to simple span tasks that only require the
maintenance of lists of items for immediate serial recall, complex span tasks require participants to perform some task while maintaining memory items for further recall, thus involving storage and manipulation of information. Moreover, the concurrent task to be performed often involves a complex activity such as problem solving or comprehension. These characteristics are common to the most popular complex span tasks, such as Daneman and Carpenter’s (1980) reading span that requires participants to read sentences while remembering words, Turner and Engle’s (1989) operation span that requires participants to solve arithmetic equations instead of reading sentences, or Cases’s counting span (1985), in which participants count arrays of dots while remembering the results of the successive counts. WM spans measured through these tasks predict performance in a wide range of complex cognitive activities and more generally are the best single predictor of fluid intelligence (Gf; e.g., Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Kyllonen & Christal, 1990).

One of the main accounts for the relationship between complex spans and high-level cognition is that complex cognitive tasks involve coordination of activities and executive control, and that WM spans are good measures of executive attention and control due to their dual task requirements (Kane, Conway, Hambrick, & Engle, 2007). Conversely, simple span tasks (i.e., tasks that only require storage) tend to be poorer predictors of complex cognition (Ackerman, Beier, & Boyle, 2005; Engle et al., 1999; Kane et al., 2004). However, it should be noted that other studies have shown that a variety of tasks with no processing components were as predictive of higher-order cognition as complex span tasks. These include simple span tasks (Colom, Rebollo, Abad, & Shih, 2006; Unsworth & Engle, 2007), running memory span tasks (Broadway & Engle, 2010; Cowan et al., 2005), change detection in visual arrays (Shipstead, Redick, Hicks, & Engle, 2012), or immediate free recall (Unsworth & Engle, 2007; Unsworth, Spillers, & Brewer, 2010). Nonetheless, as noted by Conway and colleagues (2005), complex span tasks have been the most strongly embraced by researchers among all the available measures of WM capacity and in almost all branches of psychology. Accordingly, traditional complex span tasks, such as the reading span, the operation span, and the counting span tasks, are still routinely used for assessing WM capacity (for some recent examples, see Broadway & Engle, 2010; Burgess, Gray, Conway, & Braver, 2011; Dang, Bracken, Ferrer, & Liu, 2012; Kyndt, Cascallar, & Dochy, 2012; Kyndt, Dochy, Struyven, & Cascallar, 2012; Martinez & Colom, 2008; Redick, Unsworth, Kelly, & Engle, 2012; Redick, Broadway, et al., 2012; Rose, Olsen, Craik, & Rosenbaum, 2012; Shelton, Elliott, Matthews, Hill, & Drew Gouvier, 2010; Unsworth, Brewer, & Spillers, 2009). This persistent reliance on complex span tasks involving complex activities is probably due to their compliance with Baddeley’s canonical definition of WM as a system mainly involved in complex cognition and combining processing and storage.

However, one important prerequisite for the validity of these complex span tasks is that participants must have a relatively high level of performance on the
secondary task. In this respect, a possible caveat is that reading sentences fluently or verifying complex arithmetic operations requires a rather high level of education. This requirement constitutes an important constraint for developmental studies in which young children could experience difficulties in reading or mathematical skills. It could also hamper assessment of WM capacity in adult population. The UNESCO estimates that about 16% of the adult world population is illiterate, and even in western countries where primary schooling is universal, a large number of adults have difficulties in reading comprehension and arithmetic. To circumvent this difficulty, new complex span tasks could be created involving simpler activities. Indeed, Barrouillet, Lépine, and Camos (2008) have shown that even elementary processes such as subitizing, reading digits, or arithmetic problems solved through direct retrieval from long-term memory such as very simple additions with operands from 1 to 4 were sensitive to individual differences in WM capacity, with high-span individuals being systematically faster than low-span individuals. Moreover, complex activities such as counting large arrays of dots did not induce WM-related differences beyond what can be predicted from the concatenation of differences elicited by the elementary constituents of the task (subitizing of subgroups and recursive addition of the totals; Barrouillet et al., 2008). As a consequence, if WM differences are also reflected in the simplest cognitive processes, these processes could be used in turn to assess WM capacity. Complex span tasks involving these elementary processes instead of complex activities such as reading sentences, solving complex equations, or counting large arrays of objects should be as good as the traditional complex span tasks to measure WM capacity and to predict high level cognition and Gf.

We have already designed such WM tasks within the theoretical framework of the time-based resource-sharing (TBRS) model. For example, Lépine, Bernardin, and Barrouillet (2005) compared a traditional reading span task with a task in which participants had to read a series of a few letters instead of sentences for comprehension. In a second experiment, they compared Turner and Engle’s (1989) operation span with a task requiring adding or subtracting 1 to one-digit numbers instead of solving multiple-operation equations. Importantly, whereas the traditional tasks were experimenter-paced, the TBRS tasks were time-constrained with a limited time allowed to read each letter or add/subtract 1. In line with several other studies (e.g., Barrouillet, Bernardin, & Camos, 2004; Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007), results revealed that performing the processing component of the TBRS tasks at a faster pace led to poorer recall performance, but more importantly, that elementary intervening processes were as detrimental as complex activities for concurrent maintenance. For example, reading four letters at a pace of 675 ms per letter was as detrimental as reading sentences for veracity judgment. Although these findings indicate that performing elementary processes under time pressure disrupts the concurrent maintenance of information in WM, this does not mean that the TBRS tasks assess the same capacity as traditional
complex span tasks or that they would be as good of predictors of high-level cognition. The aim of the present study was to verify this point.

In the following study, we compared three traditional complex span tasks (the reading span, the operation span, and the counting span tasks) with three TBRS span tasks involving elementary processes. A first task, the reading letter span task, required maintaining a series of digits while reading letters. The second, the upgrading span task, involved maintaining consonants while performing elementary additions. A third task, the enumeration span task, consisted of maintaining consonants while recognizing dice-like patterns of dots. Tasks involving elementary processes (reading letters, solving simple additions, pattern recognition) should be as sensitive to assessing WM capacity as complex activities such as reading sentences, solving complex equations, or counting large arrays of dots. Thus, a high correlation between the latent variables from TBRS and traditional complex span tasks was expected. As well, the two constructs should have equivalent power in predicting fluid intelligence assessed through three Gf tasks (Cattell’s culture fair test, WAIS reasoning matrix, and Raven’s progressive matrices).

Method

Participants

A total of 116 undergraduate students (21 males) enrolled at the University of Geneva received either a course credit or 40 Swiss Francs (approximately 45 US Dollars) for participating. Participants ranged in age from 19 to 45 years (mean age = 22.06 years; SD = 5.73). All participants were native French speakers.

Material and Procedure

The tasks were presented over the course of two sessions, each lasting approximately 1 hour and administered 2–3 weeks apart. All participants completed six complex span tasks in the first session administered individually, and three Gf tasks in the second session run in groups of 2 to 7 participants. Task order for the WM tasks was partially counterbalanced following a Latin-square design. For the reasoning tasks, two orders were defined by alternating the two timed tasks (Cattell’s Culture Fair and Raven’s Progressive Matrices Advanced tests), the Advanced Matrix Reasoning subtest being always last because it was time-unlimited. The WM tasks were administered with E-prime software (Schneider, Eschman, & Zuccolotto, 2012). The reasoning tasks were presented in a paper and pencil format. All of the tasks were instructed and completed in French.

The Working Memory Span Tasks

There were three traditional complex span tasks (the reading span, the operation span, and the counting span tasks) and three TBRS tasks (the reading letter
span, the upgrading span, and the enumeration span tasks). In all tasks, series of two to five or six memory items with three trials of each length were presented successively on screen in ascending length. From pre-tests and to avoid ceiling effects, the maximum length for the counting, the reading letter and the enumeration span tasks was increased to 6.

All the tasks had the same structure: Memory items were presented successively on screen for 1000 ms, each item being followed by a blank screen (i.e., an interstimulus interval, ISI) of 500 ms and then by a distracting task that differed between tasks. Each trial began by a screen indicating the number of items to be remembered (e.g., “2 digits to be memorized”). Participants were asked to read and recall memory items aloud in correct order when presented with the word “rappel” (recall). Two practice trials of length 2 preceded each WM task. Following Kane and colleagues’ (2004) scoring method, a memory item was scored as correct only if it was recalled in correct serial position. The number of correct memory items within each trial was converted into a proportion-correct score, and then the mean proportion-correct score was computed over all trials in each span task for each participant.

**Reading Span (RS)**

Participants maintained series of digits randomly drawn from 1–9. Each digit was followed by a sentence that participants read aloud at their own pace and verified whether it made sense or not by saying aloud “yes” or “no.” As soon as the participant had given her response, the experimenter pressed a key triggering the display of another digit on screen, and so on. Half of the 48 sentences made sense. The sentences were selected from Desmette, Hupet, Schelstraete, and Van der Linden (1995). To create nonsense sentences, one word was changed in the sentence (e.g., “les poubelles du restaurant étaient souvent le point de ralliement des astéroïdes (pauvres)” [bins in the restaurant were often the focal point of asteroids (poor people)]).

**Operation Span (OS)**

Adapted from Kane and colleagues (2004), participants memorized series of consonants. All consonants were used except W, which is trisyllabic in French. Each letter was followed by an equation such as \((4 \times 3) - 2 = 9\). Equations began with a multiplication or a division followed by a one-digit number to add or subtract from the product or dividend. Participants were asked to read the equation aloud and to verify it by responding “yes” or “no.” The result was correct for half of the equations. As soon as the participant gave her response, the experimenter pressed a key to move on to the next memory item.

**Counting Span (CS)**

Participants memorized series of consonants. As in Engle et al.’s (1999) CS task, an array containing 4–9 dark blue circles, 1,3,5,7, or 9 dark blue squares, and
1–5 light green circles on a gray background was presented in the inter-memory item interval. Participants were asked to count the number of blue circles in each display aloud, repeating the total at the end of the count. For example, for four dark blue circles, the participant would say “one, two, three, four . . . four”. As soon as the participants had repeated the total, the experimenter pressed a key triggering the appearance of the next memory item.

**Reading Letter Span (RLS)**

Participants had to remember series of digits. Each digit was followed by four consonants to be read aloud. Consonants appeared successively on screen for 600 ms after an ISI of 200 ms.

**Upgrading Span (US)**

Memoranda were consonants followed by series of four digits appearing successively on screen at a rate of one digit every 1600 ms, with 1200 ms of display and a 400 ms ISI. Participants were asked to add 2 to each digit and to give their response aloud.

**Enumeration Span (ES)**

Memory items were consonants. Each consonant was followed by four dice-like patterns of dots successively presented on screen. Each pattern appeared for 600 ms after an ISI of 200 ms. Participants were asked to say aloud the total of dots in each pattern.

**Reasoning Tasks**

Gf was measured using three standardized intelligence tests. For each test, one point was assigned for each correct item and then the total number of points was divided by the number of items for a proportion-correct score.

**Cattell’s Culture Fair Test (Scale 3 Form A)**

This test was composed of four separate and timed (2.5 to 4 min) subtests involving reasoning about shapes and figures (Cattell, 1986). In the subtest *Series*, participants selected the best option among six alternatives to complete a series of four figures. In subtest *Classification*, they chose the two figures that were different from the others in a set of five. In the *Matrices* subtest, they chose the one, which completed a matrix of 4 to 9 boxes among six alternatives. The final *Conditions* subtest presented sets of abstract figures containing lines and a single dot. Participants had to assess relationship between the figure, the lines and the dot, and to choose the one in which the dot could be placed according to the same relationship among five alternatives.
Matrix Reasoning Subtest (WAIS-R Third Edition; Psychological Corporation, 1997)

This untimed nonverbal reasoning test contained 26 items: three sample items and 23 regular items (Weschler, 2000). Each item presented a pattern of colored figures arranged in a matrix with one figure missing. Participant selected the figure that completed the matrix among five response options.

Raven’s Progressive Matrices Advanced (Abbreviated Version, ECPA, 2009)

Each item presented a pattern of eight black and white figures arranged in a 3 $\times$ 3 matrix. The lower right corner of the matrix was missing, and participants selected the best one of eight multiple-choice alternatives to fill the empty corner. Following four practice items, participants had 20 minutes to complete 23 items that increased in difficulty.

Results

One participant was excluded from the analysis due to his performance on the WAIS task being four standard deviations below the group mean. To be sure that task order did not affect performance, two analyses of variance (ANOVA) were performed, one for the WM tasks and one for the Gf tasks, with task order as between-subject factor and tasks as within-subject factor. In both analyses, task order was not significant, $F_s < 1$, as well as its interaction with the tasks, $p_s > .20$. The correlation matrix and descriptive statistics of each measure for the remaining 115 participants are shown in Table 1. The measures generally reached traditionally accepted norms of reliability, skewness and kurtosis. Correlations demonstrated discriminant validity with high intercorrelations among tasks theorized to measure the same construct (average correlations of .49, .48, and .45 for TBRS tasks, traditional tasks, and Gf tasks respectively). The TBRS and traditional tasks correlated with each other (average correlation of .49), with correlations often higher between than within categories of tasks, suggesting that TBRS tasks are valid indicators of WM capacity. However, some correlations between WM and Gf tasks proved lower than usually observed and sometimes non-significant. These low correlations could be due to a truncation of the range for Gf scores because we may suspect undergraduate students to have higher than average intelligence. However, the comparison of our sample with the norms did not confirm this hypothesis. The mean score of our sample was very close to the norm for Cattell’s test (mean score of the sample: .56, norm: .56, SD of .09 and .10, respectively) and the matrix subtest of the WAIS (.84 and .85, respectively$^1$) whereas our sample was lower than the norm for Raven’s test (mean of .46 and .56 and SD of .17 and .15, respectively). Thus, it is difficult to assume that non-significant correlations could be due to a range restriction of Gf scores on our sample. Moreover, for the significant correlations, their range from .20 to .28 did not diverge from what is usually reported. For example, Conway, Cowan, Bunting,
TABLE 1. Correlation Matrix for Measures of Task Performance

<table>
<thead>
<tr>
<th>Measures</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cattell</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. WAIS-R</td>
<td>.49</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Raven’s</td>
<td>.43</td>
<td>.43</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Reading</td>
<td>.24</td>
<td>.20</td>
<td>.21</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Upgrading Span</td>
<td>.11</td>
<td>.06</td>
<td>.03</td>
<td>.36</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Enumeration</td>
<td>.20</td>
<td>.24</td>
<td>.25</td>
<td>.48</td>
<td>.63</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Reading Span</td>
<td>.27</td>
<td>.28</td>
<td>.26</td>
<td>.50</td>
<td>.28</td>
<td>.55</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Operation</td>
<td>.13</td>
<td>.00</td>
<td>.02</td>
<td>.50</td>
<td>.45</td>
<td>.55</td>
<td>.43</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>9. Counting</td>
<td>.15</td>
<td>.17</td>
<td>.11</td>
<td>.51</td>
<td>.54</td>
<td>.56</td>
<td>.43</td>
<td>.56</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>115</th>
<th>115</th>
<th>115</th>
<th>115</th>
<th>115</th>
<th>115</th>
<th>115</th>
<th>115</th>
<th>115</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0.56</td>
<td>0.84</td>
<td>0.46</td>
<td>0.81</td>
<td>0.76</td>
<td>0.76</td>
<td>0.80</td>
<td>0.62</td>
<td>0.84</td>
</tr>
<tr>
<td>SD</td>
<td>0.09</td>
<td>0.10</td>
<td>0.17</td>
<td>0.10</td>
<td>0.16</td>
<td>0.12</td>
<td>0.14</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>Cronbach’s α</td>
<td>0.55</td>
<td>0.67</td>
<td>0.75</td>
<td>0.68</td>
<td>0.76</td>
<td>0.77</td>
<td>0.70</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>Skewness</td>
<td>−0.46</td>
<td>−1.22</td>
<td>−0.17</td>
<td>−0.53</td>
<td>−0.72</td>
<td>−0.88</td>
<td>−1.19</td>
<td>−0.24</td>
<td>−1.05</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>−0.10</td>
<td>1.45</td>
<td>0.26</td>
<td>−0.40</td>
<td>0.40</td>
<td>1.45</td>
<td>1.35</td>
<td>−0.24</td>
<td>1.16</td>
</tr>
</tbody>
</table>

Note. Bolded correlations are significant at the $p < .05$ level. Note that Cronbach’s $\alpha$, Skewness, and Kurtosis were obtained from the current sample.

Therriault, and Minkoff (2002) reported a correlation of .29 between reading span and Cattell’s test and a non-significant correlation ($r = .15$) between reading span and Raven’s test. These correlations were of .24 and .28 respectively in Engle and cousins (1999). Jurden (1995) reported a correlation of .20 and Ackerman, Beier, and Boyle (2002) of .23 between the reading span and Raven’s test. Moreover, the intercorrelations between the three traditional tasks that ranged between .43 and .56 were similar to those reported in prominent studies like Engle et al. (1999; range .32–.51) or Conway and colleagues (2002; range .49–.53). Finally, it should be noted that all the span tasks that failed to significantly correlate with reasoning tests (i.e., operation, updating and counting span tasks) involved mathematical skills. In previous studies, we already observed that TBRS span tasks involving mathematical skills were not as good predictors of school achievements as those involving simpler activities like reading letters (Barrouillet, Camos, Morlaix, & Suchaut, 2008; Lépine, Barrouillet, & Camos, 2005).

As a first exploration of our data set, we performed an Exploratory Factor Analysis using an oblique (Promax) rotation method that revealed a two-factor structure (Table 2). All the complex span tasks loaded on the same factor that
TABLE 2. Exploratory Factor Analysis Models of Fluid Intelligence and Working Memory Task Performance

<table>
<thead>
<tr>
<th>Measure</th>
<th>Rotated Solution Factor</th>
<th>Unrotated Solution Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Cattell</td>
<td>.240</td>
<td>.590</td>
</tr>
<tr>
<td>WAIS R</td>
<td>.187</td>
<td>.637</td>
</tr>
<tr>
<td>Raven’s</td>
<td>.146</td>
<td>.935</td>
</tr>
<tr>
<td>Reading Letters Span</td>
<td>.661</td>
<td>.343</td>
</tr>
<tr>
<td>Upgrading Span</td>
<td>.778</td>
<td>.062</td>
</tr>
<tr>
<td>Enumeration Span</td>
<td>.834</td>
<td>.375</td>
</tr>
<tr>
<td>Reading Span</td>
<td>.665</td>
<td>.482</td>
</tr>
<tr>
<td>Operation Span</td>
<td>.805</td>
<td>.040</td>
</tr>
<tr>
<td>Counting Span</td>
<td>.748</td>
<td>.208</td>
</tr>
</tbody>
</table>

Correlations

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.279</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. Boldface used to help distinguish between the factors.

clearly differed from the second factor on which the reasoning tasks loaded. To test our prediction that the TBRS complex span tasks are as good predictors of Gf as the traditional complex span tasks, we performed confirmatory factor analyses (CFA). In a first CFA, we tested how a WM construct derived from TBRS tasks is related to the WM factor derived from traditional tasks, and to what extent the TBRS latent variable was as predictive of Gf as the factor derived from traditional tasks. For this purpose, we tested a three-factor model distinguishing between the traditional and TBRS WM tasks in relation to a Gf factor. The fit of this model was adequate (Model 1 in Table 3). Despite a significant $\chi^2$, the ratio between $\chi^2$ and degrees of freedom was 2.03, indicating acceptable fit. In addition, the model met traditional fit criteria: a comparative fit index (CFI) of .90 (Hu & Bentler, 1995) and a root-mean-square-error of approximation (RMSEA) of .10 (Browne & Cudeck, 1993). This analysis indicated high loadings of each of the tasks onto its theoretical construct, with comparable loading for TBRS and traditional tasks (Figure 1A). As we predicted, the two WM constructs were very highly correlated and both constructs significantly correlated with Gf. Although the correlations between each of the WM factors and Gf were significantly different (.37 versus .30,
### TABLE 3. Confirmatory Factor Analysis Models of Fluid Intelligence and Working Memory Task Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>AIC</th>
<th>NFI</th>
<th>NNFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1. Three factors (Figure 1, A)</td>
<td>48.78*</td>
<td>24</td>
<td>0.10</td>
<td>0.93</td>
<td>90.78</td>
<td>0.87</td>
<td>0.89</td>
</tr>
<tr>
<td>Model 2. Two factors (Figure 1, B)</td>
<td>49.50*</td>
<td>26</td>
<td>0.09</td>
<td>0.93</td>
<td>87.50</td>
<td>0.87</td>
<td>0.90</td>
</tr>
<tr>
<td>Model 3. Two factors (Figure 1, C)</td>
<td>1.45</td>
<td>8</td>
<td>0.00</td>
<td>1.00</td>
<td>27.45</td>
<td>0.99</td>
<td>1.08</td>
</tr>
<tr>
<td>Model 4. Two factors (Figure 1, D)</td>
<td>1.20</td>
<td>4</td>
<td>0.00</td>
<td>1.00</td>
<td>23.20</td>
<td>0.99</td>
<td>1.07</td>
</tr>
</tbody>
</table>

*Significant chi-square at the $p < .05$ level. RMSEA: root mean square error; CFI: comparative fit index; AIC: Akaike’s information criterion; NFI: normed fit index; NNFI: non-normed fit index.

**FIGURE 1.** Models of task performance on the fluid intelligence (Gf) and working memory (WM) tests, variability of which is modeled as supported by three related constructs (Model A), two related constructs including all the complex span tasks (Model B), or excluding the complex span tasks that require arithmetic skills (Model C), or with only two TBRS span tasks (Model D). Solid lines signify a significant loading or correlation at the $p < .05$ level.
$Z = 3.97, p < .01$), fixing the pair of correlations to equivalence did not significantly impair the model fit compared to the original three-factor model shown in Figure 1A, $\chi^2(1)$ difference $= 0.65, p = .42$. This suggests that the traditional WM tasks and the TBRS WM tasks share a substantial degree of variance, and similarly account for variability in Gf.

Thus, we tested whether the three-factor model could be improved upon if the variability of the traditional and TBRS WM tasks was instead accounted for by a common, unitary WM construct. A summary of this second tested CFA can be found in Table 3 (Model 2), and is visually depicted in Figure 1B. The fit of this second model was slightly better with lower ratio between $\chi^2$ and degrees of freedom (1.90), lower RMSEA and AIC, and NNFI now reaching .90. However, the models were not significantly different from one another, $\chi^2(2)$ difference $= 0.72, p = .70$. In a last CFA, we discarded from the previous model those complex span tasks that poorly correlated with reasoning tasks. This model in which the reading letters, the enumeration and the reading span tasks loaded on the same WM factor (Model 3 in Table 2; depicted in Figure 1C) proved to have an excellent fit, with a non-significant $\chi^2$ and excellent NFI. It also showed a high correlation between the two latent variables ($r = .50$). Interestingly, the fit of this model proved even slightly better when the traditional reading span task was removed, with a correlation between the two latent variables remaining almost unchanged (Model 4 in Table 2; Figure 1D). These latter analyses suggest that the TBRS span tasks are at least as good predictors of fluid intelligence as the traditional complex span tasks.

**Discussion**

In summary, the results suggested that spans from the TBRS tasks are valid measure of WM capacity and as good of predictors of Gf as the traditional span tasks, both at the task and latent variable level. Specifically, the best-fitting model comprised two factors: one representing a unitary WM construct, which was significantly correlated to the second Gf factor. Furthermore, dropping the tasks that exhibited lower correlations with Gf and keeping two TBRS tasks along with the sole traditional reading span task substantially improved the fit of this two-factor model. This suggests that time-constrained complex span tasks can be used to assess WM capacity and reliably predict Gf, thereby extending the viable repertoire of WM tasks available to researchers.

These results have practical implications for WM capacity measurement. Complex span tasks involving simpler activities do not rely as much as traditional span tasks on prior school acquisitions and would consequently extend the population for which WM capacity could be assessed. In the same way, these tasks can be administered in children as soon as they are able to read digits or letters, favoring the use of the same tasks across a wide age range in life-span studies and developmental comparisons. We have already verified that TBRS tasks are
as good of predictor of school achievement in 12-year-old children as traditional span tasks (Lépine et al., 2005). Another strength of complex span tasks involving simpler activities is to provide us with more reliable WM capacity measures on large samples of individuals who greatly vary in school achievement, because it can be expected that individual variability is lower in elementary processes than in complex cognitive activities such as reading comprehension or mathematics. Thus, the WM capacity measure should be less contaminated by proficiency in performing the secondary task. Moreover, the elementary and computer-paced activities used in TBRS tasks are easily adaptable for neuroscientific investigations such as ERP and fMRI studies, as testified by pilot studies in our lab (Vergauwe, Hartstra, Barrouillet, & Brass, submitted).

At a more theoretical level, the present results suggest that the validity of the traditional span tasks in measuring WM capacity does not rely on the complexity of their processing component that can be replaced by the time-constrained succession of elementary cognitive steps. As we noted above, other tasks even without processing components are also good predictors of Gf and high-level cognition. Why is it that tasks varying so greatly in structure and content can similarly predict Gf? According to Cowan and colleagues (2005), a way to account for this phenomenon is to assume that the attention-demanding aspects of the tasks underpin their correlation with intelligence. This correlation would be diminished when memory performance relies on non-attentional mechanisms of maintenance, such as verbal rehearsal. Thus, a reliable WM task is a task that prevents or discourages the use of such mechanisms. The introduction of computer-paced elementary activities in complex span tasks would have the same effect as complex cognitive activities in preventing verbal rehearsal. Another way to understand the same phenomenon would be to suppose, in line with the TBRS model, that processing has an effect on storage by occupying attention and thereby preventing maintenance activities. In this respect, even a very simple process performed at a sufficiently fast pace could have the intended effect of distracting attention from maintenance in WM (Lépine, Bernardin, & Barrouillet, 2005).

In sum, the present results establish that WM capacity can be assessed through the proficiency in maintaining memory items while performing elementary activities under temporal constraints. Though a received view conceives WM as a system mainly involved in complex cognition such as comprehension, learning, and reasoning, the capacity of this system seems to depend on the efficiency of fairly simple and elementary processes. The present findings suggest that a large variety of valid complex span tasks could be easily constructed by using time-constrained elementary activities as processing components. These tasks would permit better adaptability to constraints in experimental settings and designs, or specificities of the populations under study.
NOTE

1. Standard deviation non available for the norm in the French version of the WAIS manual. SD for our sample of .11.

AUTHOR NOTES

Annalisa Lucidi is currently a post-doctoral fellow at University of Geneva. Vanessa Loaiza is currently a post-doctoral fellow at University of Fribourg. Valérie Camos is professor at University of Fribourg, where she heads the Fribourg Center for Cognition, a multidisciplinary center, and the Laboratory for Cognitive Development. Pierre Barrouillet is professor and head of the Developmental Cognitive Psychology (DeCoPsy) group at University of Geneva, where he is also director of the Archives Jean Piaget.

FUNDING

This study was supported by grants 100014_126910/1 and 100014_137860 from the Swiss National Science Foundation.

REFERENCES


*Original manuscript received July 1, 2013*

*Final version accepted November 22, 2013*