



The composite complex span: French validation of a short working memory task.

Corentin Gonthier, Noémie Thomassin, Jean-Luc Roulin

► To cite this version:

Corentin Gonthier, Noémie Thomassin, Jean-Luc Roulin. The composite complex span: French validation of a short working memory task.. Behavior Research Methods, 2016, 48 (1), pp.233-42. 10.3758/s13428-015-0566-3 . hal-01569884

HAL Id: hal-01569884

<https://hal.science/hal-01569884>

Submitted on 1 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Composite Complex Span:
French Validation of a Short Working Memory Task

Corentin Gonthier^{a,b}, Noémie Thomassin^{a,c}, & Jean-Luc Roulin^{a,c}

^a Laboratoire de Psychologie et Neurocognition, UMR CNRS 5105, Université Pierre Mendès France, BP 47, 38040 Grenoble Cedex 09, France

^b Department of Psychology, University of Grenoble, France

^c Department of Psychology, University of Savoy, France

Correspondence concerning this article should be addressed to Jean-Luc Roulin, Laboratoire LPNC, Domaine Universitaire de Jacob Bellecombette, BP 1104, 73011 Chambéry Cedex

E-mail: jean-luc.roulin@univ-savoie.fr

Abstract

Most studies in individual differences in the field of working memory research use complex span tasks to measure working memory capacity. Various complex span tasks based on different materials have been developed, and these tasks have proven both reliable and valid; several complex span tasks are often combined to provide a domain-general estimate of working memory capacity with even better psychometric properties. The present work sought to address two issues. Firstly, having participants perform several full-length complex span tasks in succession makes for a long and tedious procedure. Secondly, few complex span tasks have been translated and validated in French. We constructed a French working memory task labeled the Composite Complex Span (CCS). The CCS includes shortened versions of three classic complex span tasks: the reading span, symmetry span and operation span. We assessed the psychometric properties of the CCS, including test-retest reliability and convergent validity with Raven's Advanced Progressive Matrices and with an alpha span task; the CCS demonstrated satisfying qualities in a sample of 1093 participants. This work provides evidence that shorter versions of classic complex span tasks can yield valid working memory estimates. The materials and normative data for the CCS are also included.

Keywords

Working memory capacity; Complex span task; French validation; Psychometric properties; Short version

Working memory capacity (WMC), defined as the ability to maintain and manipulate information at the same time, is a central construct in human cognition. In particular, WMC is thought to play a role in a range of complex behaviors (Engle & Kane, 2004). Interestingly, WMC is subject to individual differences that appear relatively stable in time (Klein & Fiss, 1999); these individual differences are strongly related to fluid intelligence (Ackerman, Beier & Boyle, 2005) and more generally to performance in high-level cognitive tasks (Engle & Kane, 2004). It is therefore of interest to accurately measure individual differences in WMC.

The ubiquitous complex span tasks are certainly the most frequently used paradigm to assess working memory (for a review, see Conway et al., 2005; Redick et al., 2012). Complex spans are based on the model of simple span tasks, which require participants to memorize a series of stimuli presented in quick succession. Contrary to simple spans, however, complex spans interleave the presentation of to-be-remembered stimuli with a processing task – for example reading a sentence or solving a mathematical operation. This association of processing and storage requirements constitutes a direct operationalization of the definition of working memory. Complex spans typically demonstrate excellent psychometric properties (Conway et al., 2005; Redick et al., 2012): they have good internal consistency (Redick et al., 2012), stability over time (Klein & Fiss, 1999), and convergent and criterion validity (Redick et al., 2012). By contrast, other tasks frequently used as working memory measures are not nearly as successful: for example, the backward span is more strongly associated with short-term memory than with working memory (e.g. Engle, Tuholski, Laughlin, & Conway, 1999) and the *n*-back task demonstrates limited reliability as well as limited correlations with other working memory measures (Jaeggi, Buschkuhl, Perrig, & Meier, 2010, Redick & Lindsey, 2013).

Many different complex spans have been developed over the years. The seminal complex span was the reading span (Daneman & Carpenter, 1980). In the original version of the task, participants were asked to read a series of sentences and decide whether they were correct; the last word of each sentence had to be memorized for serial recall at the end of a trial. Other classic complex span tasks are the operation span, in which participants have to decide whether mathematical operations are correct while memorizing unrelated stimuli presented after each operation (Turner & Engle, 1989; Unsworth, Heitz, Schrock, & Engle, 2005), and the symmetry span, in which participants have to decide whether spatial displays are vertically symmetrical while memorizing spatial locations (Kane et al., 2004). Yet other complex span tasks exist, such as the counting span, navigation span or rotation span (see Kane et al., 2004). Despite being based on a variety of materials, such as visual, spatial, verbal and numeric stimuli, all these complex span tasks seem to assess the same underlying construct: latent variable analyses generally indicate that complex span tasks load on a common, domain-general factor, and that this domain-general factor has better predictive validity than domain-specific factors (e.g. Kane et al., 2004). Although the assessment of domain-specific WMC may be of interest, these results often lead studies in individual differences to combine several complex span tasks so as to obtain a domain-general estimate of WMC (for recent examples, see Redick & Engle, 2011; McVay & Kane, 2012; Unsworth, Brewer & Spillers, 2011).

The present work was motivated by two issues related to the practical use of complex span tasks. Firstly, although many researchers choose to combine multiple complex span tasks in the same protocol, this solution makes for a long procedure that can be tedious for the participant. Most studies employing more than one multiple complex span task have used the reading span, symmetry span and operation span; having a participant complete the most common versions of all three tasks (Unsworth et al., 2005) yields a total of 42 trials, or 192

stimuli to remember and 192 processing demands to carry out, without even taking into account the training phases for each task. This high number of trials makes it difficult to include other tasks in the same experimental session. It may also pose experimental problems by decreasing participant engagement in the task and increasing fatigue; this is not a trivial issue since complex spans are sensitive to task sequence, both because performing a complex span may decrease performance in subsequent tasks (Schmeichel, 2007) and because performance in complex span tasks can be lowered if demanding tasks have been previously completed in the same testing session (Healey, Hasher, & Danilova, 2011). These issues may be especially problematic in developmental or clinical settings. Moreover, a large number of trials encourages the buildup of proactive interference throughout the successive tasks, which can directly affect working memory performance (May, Hasher, & Kane, 1999; Lustig, May, & Hasher, 2001). This problem is especially critical for the assessment of working memory, since participants with a low working memory capacity are known to be more sensitive to proactive interference (Kane & Engle, 2000).

Importantly, the large number of trials included in common complex spans comes from the fact that they were designed as stand-alone tasks, sufficient to obtain a psychometrically sound measure of WMC by themselves. However, this constraint can be avoided: since the different complex spans are known to reflect a common underlying construct, we may consider the association of multiple complex spans as a single working memory test. If individual complex spans are viewed as subtests of a larger test, then they do not need to have individually sufficient psychometric properties and the number of trials per task can be reduced. In other words, it is possible to construct a working memory test including several complex span tasks serving as subtests, with only a low number of trials per subtest, as long as the total number of trials across all subtests is sufficient to obtain a reliable measure. This idea is supported by a recent work indicating that classic complex span tasks

retain significant validity even when reducing the number of trials by two thirds, and that combining shortened versions of multiple complex span tasks yields a better measure than using a full-length version of a single task (Foster et al., 2014). Shortened versions of complex spans tasks are also emerging in the literature, and these tasks demonstrate adequate psychometric properties (Oswald, McAbee, Redick, & Hambrick, in press).

Secondly, the range of available complex span tasks is limited for French-speaking samples. Certain types of span tasks are suitable for working memory assessment – such as time-constrained span tasks, which have been validated in French (Lucidi, Loaiza, Camos, & Barrouillet, 2014) – but these are not complex span tasks. Two versions of the reading span task and two versions of the operation span exist in French, but they all differ significantly from the widely used English-speaking versions of the tasks (Unsworth et al., 2005). The first version of the reading span task (Desmette, Hupet, Schelstraete, & van der Linden, 1995) is not computerized and only includes correct sentences, which means the only processing requirement is to read the sentences. The second version of the reading span (Delaloye, Ludwig, Borella, Chicherio, & de Ribaupierre, 2008) is computerized and includes incorrect sentences, but the sentences differ markedly in structure from English-speaking versions – their average length is 5.5 words (whereas the average length is 12.6 words in Unsworth et al., 2005), and half the sentences begin with the word *they*. Both versions of the task require participants to remember the last word of each sentence, rather than unrelated stimuli (as is the case in Unsworth et al., 2005); the words also have to be recalled orally, which precludes using the tasks in group sessions. The two versions of the operation span (Fournet et al., 2012) have participants memorize words or spatial locations instead of consonants; they present trials in ascending order of difficulty, rather than in pseudo-random order; and they have only been normed for older adults. Complex spans such as the symmetry span and operation span do not rely on verbal materials, which means they could be adapted by simply

translating the instructions; however, there may be differences in normative data between French and English-speaking samples. In particular, Unsworth and colleagues (Unsworth et al., 2005) recommend that all participants with accuracy lower than 85% on the processing task be excluded from the sample; we have observed that a very high number of participants consistently fail to reach this level of performance in work from our own laboratory, especially on the operation span.

In order to address both these issues, we constructed the Composite Complex Span (CCS), a French-speaking composite working memory task. The CCS included three subtests: the reading span, symmetry span and operation span. These tasks were chosen because they are the most widespread complex span tasks, because they have been validated in very large samples (Redick et al., 2012), and because they represent a variety of materials: with these three subtests, the CCS includes numeric, visuo-spatial and verbal content. All three subtests were designed to mimic the widespread English-speaking versions of the tasks (Unsworth et al., 2005). Because the three subtests were not intended to be used in isolation, they were shortened relative to the original versions by halving the number of trials. The CCS was entirely computerized and did not require oral responses from the participants, thus allowing for group administration.

Method

The Composite Complex Span

The CCS includes three subtests: the reading span, symmetry span and operation span, presented in this order. The whole procedure takes approximately 25 minutes. All three subtests have the same structure: in each trial, participants have to solve a series of simple processing problems while memorizing unrelated stimuli presented after each problem. At the end of a trial, a grid containing all possible to-be-remembered stimuli appears on the screen; participants have to click the cases of the grid corresponding to the stimuli they have seen, in

the correct order. An illustration of the operation span subtest is presented in Figure 1. The reading span subtest requires participants to tell whether sentences are correct while memorizing unrelated digits; the symmetry span requires participants to tell whether spatial displays are vertically symmetrical while memorizing spatial locations within a grid; and the operation span requires participants to tell whether mathematical operations are correct while memorizing consonants.

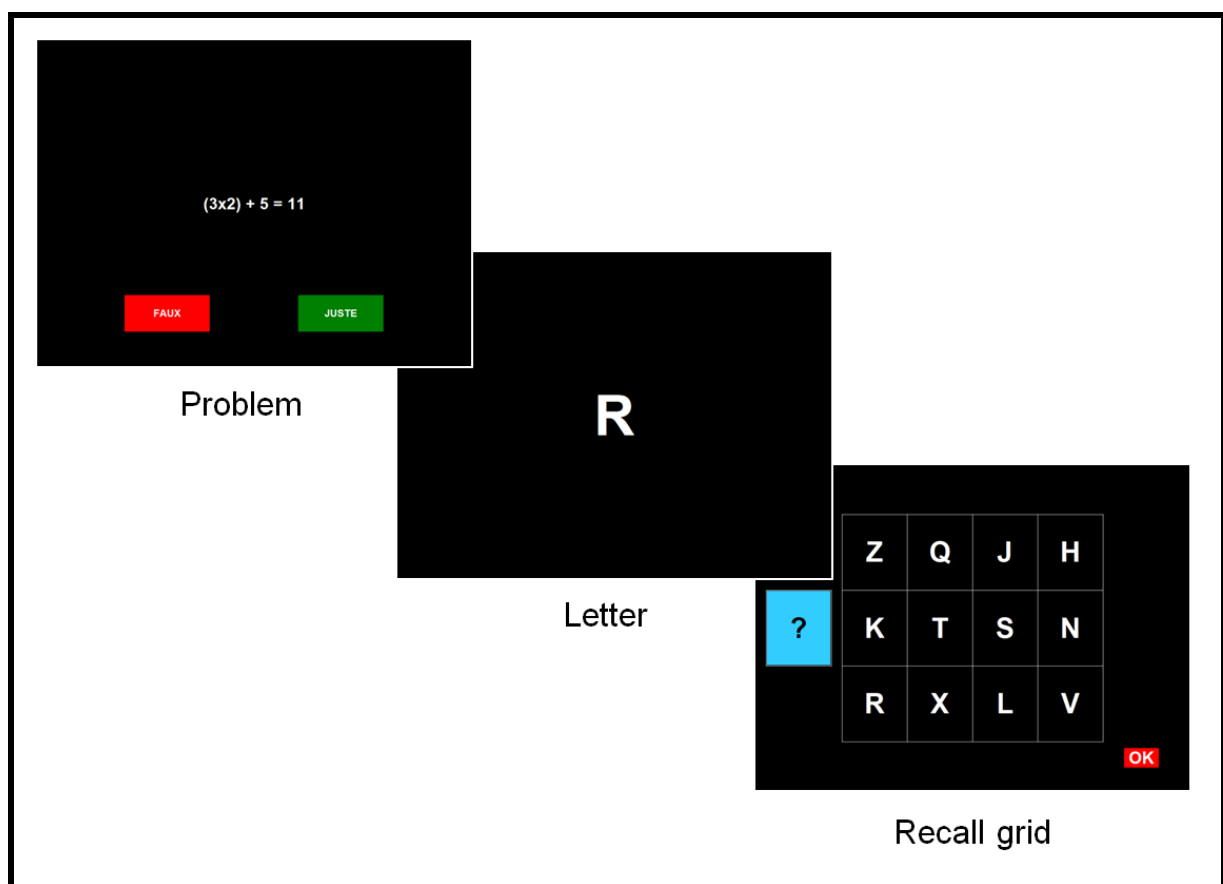


Figure 1. Illustration of the operation span subtest of the CCS. A series of problems and letters to memorize is followed by the recall grid.

The difficulty varies for the different subtests: set sizes range from 4 (four processing problems to solve interleaved with four stimuli to memorize) to 8 for the reading span, from 3 to 6 for the symmetry span and from 3 to 7 for the operation span. These set sizes were based

on the versions used by Unsworth and colleagues (from 3 to 7 for the reading span and operation span and from 2 to 5 for the symmetry span; Unsworth et al., 2005); set sizes were increased for the reading span and symmetry span because preliminary data acquired in a small sample ($N = 45$) suggested that these tasks were slightly too easy in our population. In order to shorten the duration of the testing session, the number of trials per set size was reduced when compared to the versions used by Unsworth and colleagues (which include three trials per difficulty level). Each subtest includes only one trial for the lowest and highest set sizes (for which less sensitivity is needed since there are fewer participants to discriminate at these levels of ability), and two trials for all other set sizes. The trials are presented in pseudo-random order (identical for all participants) to ensure that the set size of the current trial cannot be anticipated (Unsworth et al., 2005; see also St Clair-Thompson, 2012).

Each subtest is preceded by a training phase including three practice sessions, based on the procedure used by Unsworth and colleagues (Unsworth et al., 2005). Participants receive feedback on their performance after each trial in the practice sessions. The first practice session trains participants to memorize stimuli without a concurrent processing demand; for example in the reading span training, participants simply have to memorize and recall a series of digits. Participants complete three practice trials in this first session (one trial each of set sizes 2, 3 and 4). The second practice session trains participants to perform the processing task, without a memory requirement: for the reading span training, participants only have to tell whether sentences are correct. Participants initially complete fifteen practice trials in this session; however, if they fail to correctly answer at least 65% of trials, the practice session is repeated until they meet this criterion. There is no time constraint on this second practice session, but the participant's response times are registered and serve to calculate a time limit to complete the processing problems in the subsequent phases of the task. The time limit is calculated as the participant's mean response time plus 2.5 standard

deviations (Unsworth et al., 2005). If the participant fails to answer the processing problem within this delay during the third practice session or the real block of trials, the program registers an error and moves on to the next stimulus. This time limit ensures that participants cannot freely rehearse the series of to-be-remembered stimuli while they are supposed to answer a processing problem. The third and final practice session trains participants to perform the memory and processing tasks simultaneously and is similar to the real block of trials. Prior to beginning the third session, participants are instructed that the memory and the processing tasks are equally important, and that they should strive to remain above 85% of accuracy on the processing task at all times. Participants complete two practice trials in this session (one trial of set size 2 and one trial of set size 3).

Stimuli For The Complex Span Tasks

The reading span task.

All stimuli for the reading span subtest are presented in Supplemental material 1. To-be-remembered stimuli are digits from 1 to 9, counterbalanced across trials. The same digit never appears twice in the same trial, and no trial includes a meaningful sequence of numbers. The sentences for the processing task are based on the stimuli used by Desmette and colleagues (Desmette et al., 1995). Half the sentences were made nonsensical by replacing one selected word with another word incongruent to the meaning. All nonsensical sentences remained syntactically correct – e.g., *Un étranger apparut sur le seuil et tendit à la fille un petit sac de fenêtres* [A stranger appeared on the doorstep and handed the girl a small bag of windows]. The position of the incongruent word was situated between the middle point and the end of the sentence, counterbalanced across all trials. Each trial included between 25% and 75% of incorrect sentences.

The symmetry span task.

The stimuli for the symmetry span subtest are presented in Supplemental material 2. To-be-remembered stimuli are sequentially presented spatial locations in a 4x4 matrix; the stimuli are displayed to the participant as one square of the matrix colored in red. Spatial locations are counterbalanced across trials; the same location never appears twice within the same trial; and the locations never form a meaningful spatial pattern. The spatial displays for the symmetry judgment task were re-used from the classic computerized version of the symmetry span (Unsworth et al., 2005) with permission from the authors. These spatial displays are made up of black and white squares in a 8x8 matrix; half the displays are vertically symmetrical, and each trial includes between 25% and 75% of vertically symmetrical displays.

The operation span task.

The stimuli for the operation span subtest are presented in Supplemental material 3. To-be-remembered stimuli are consonant letters chosen for their visual and phonological distinctiveness (for example, the task includes the letter N but not the letter M; a total of 11 different letters are used), counterbalanced across trials. The same letter never appears twice within the same trial, and the letters never form a meaningful sequence. The mathematical operations for the processing task follow the same structure as the original operation span (Turner & Engle, 1989; Unsworth et al., 2005). Each operation string includes two simple operations and a stated result – e.g., $(2 \times 2) + 7 = 11$. The operands include all digits from 1 to 9; the first operation in the string can be a multiplication or a division and the second operation can be an addition or a subtraction, counterbalanced across trials. The correct result of the operation string is always an integer comprised between 1 and 20. The stated result is incorrect in half the operation strings, and each trial includes between 25% and 75% of correct operations.

Scoring Method

Performance in the CCS was scored with the partial credit load method (Conway et al., 2005); in other words, participants are awarded one point per correctly recalled stimulus in each trial. With this scoring method, a participant correctly recalling four out of five stimuli in a trial of set size 5 would get four points. The partial credit method is the preferred scoring method for complex span tasks (Conway et al., 2005; Redick et al., 2012); we adopted the load version because it produced slightly more normal distributions in our sample. This scoring method yields one working memory score for each subtest. Working memory scores on each subtest are then transformed into z-scores and the three z-scores are averaged, yielding a single composite working memory score. Processing accuracy scores, calculated as the percentage of processing problems correctly answered by the participant, are also retrieved for each subtest. Participants with less than 85% accuracy on a processing task are typically excluded from the sample (Conway et al., 2005; Unsworth et al., 2005); however, various studies performed in our laboratory suggested that this criterion is too strict in French student samples. For this reason, we instead elected to exclude participants who score in the bottom 5th percentile of the distribution of processing accuracy scores. When a participant scores below the exclusion criterion in a single subtest, their working memory score is calculated as the average of their scores on the two other subtests; when a participant scores below the criterion in two or all three subtests, their data are discarded entirely.

Validation Procedure

Convergent validity tasks.

Two tasks were used to assess the predictive validity of the CCS. The first task was set II of Raven's Advanced Progressive Matrices (APM; Raven, Raven & Court, 1998), a test of fluid intelligence. Set II of the APM is made up of 36 items of ascending difficulty; each item comprises a matrix of nine geometric patterns that follow various logical rules. On each

item, the bottom-right piece of the matrix is missing, and the participant has to select the correct piece to complete the matrix among eight alternatives. Working memory demonstrates consistent correlations with fluid intelligence, and the APM are frequently used to test convergent validity when validating complex span tasks (e.g. Redick et al., 2012; Unsworth et al., 2005).

Because we wanted to ensure that the CCS correlates with working memory tasks other than complex spans, we chose the alpha span as a second convergent validity measure (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000). This working memory task requires participants to read a series of words and to recall the first letter of each word in alphabetical order. The alpha span is not a complex span with interleaved presentation of processing problems and to-be-remembered stimuli; instead, the processing requirement in the task is to rearrange the first letters of each word in alphabetical order. We constructed a French version of the alpha span for this validation study (stimuli are presented in Supplemental material 4). The alpha span included five practice trials with set sizes ranging from 2 to 8, and eight target trials with set sizes ranging from 4 to 8, similar to the reading span and operation span subtests. A pre-test experiment conducted in a sample of 104 participants revealed that the alpha span correlated with Raven's APM ($r = .49, p < .001$), indicating convergent validity. Internal consistency was also satisfying for both the alpha span ($\alpha = .68$) and the APM ($\alpha = .75$).

Validation sample.

A total of 1093 participants completed the CCS (mean age = 20.79 years, $SD = 4.61$; 142 male). These data were collected over the course of three years, in the context of several different experiments not reported here. All participants were university students participating for course credit; they were recruited at the University of Savoy or at the University of Grenoble, France. The following inclusion criteria were observed: having

French as a first language, having no history of neurologic disorders, and taking no psychoactive drugs. All participants provided written informed consent prior to the experimental session. A subset of these 1093 participants ($N = 303$) performed the task on two separate occasions, allowing for the examination of test-retest reliability. The test-retest data was collected incidentally over multiple experiments, which means the delay between the two testing sessions varied (median = 57 days, range = 13 – 398). Two other subsets of participants additionally completed either the APM ($N = 184$) or the alpha span ($N = 249$) in the same session as the CCS, allowing for the examination of convergent validity.

Results

Descriptive statistics

Among the total sample of 1093 participants, 20 participants (1.8%) were excluded because they failed to reach the accuracy criterion on the processing tasks in two or all three subtests. Another 99 participants (9.1%) failed to reach the accuracy criterion in a single subtest, and their working memory scores were calculated on the basis of the two other subtests. The remaining 974 participants (89.1%) performed adequately in all three subtests. Most participants needed a single practice session on the processing task to reach the accuracy criterion in each subtest; more than one practice session was required for 12 participants in the reading span (1.1%), 6 participants in the symmetry span (0.5%), and 33 participants in the operation span (3.0%).

Descriptive statistics for working memory scores and processing accuracy scores are presented in Table 1. Overall, the working memory scores for each subtest were normally distributed. For the reading span and symmetry span subtests, processing accuracy scores showed high kurtosis coefficients, indicating a floor effect (similar to Redick et al., 2012); this floor effect on processing scores is a desirable feature of complex spans since the processing task is only intended as a distraction rather than a sensitive psychometric measure

(Redick et al., 2012). For the operation span, processing accuracy scores were approximately normally distributed, indicating the absence of a floor effect.

Table 1. *Descriptive statistics for working memory and processing accuracy scores.*

| Measure | Mean | SD | Skewness | Kurtosis |
|--------------------------------|-------|-------|----------|----------|
| Composite working memory score | -0.01 | 0.80 | -0.60 | 0.57 |
| Reading span | | | | |
| Working memory score | 32.12 | 8.71 | -0.22 | -0.39 |
| Processing accuracy score | 88.09 | 9.81 | -2.10 | 6.06 |
| Symmetry span | | | | |
| Working memory score | 18.37 | 5.38 | -0.55 | -0.23 |
| Processing accuracy score | 89.55 | 8.89 | -1.79 | 5.37 |
| Operation span | | | | |
| Working memory score | 28.88 | 7.48 | -0.87 | 0.68 |
| Processing accuracy score | 80.93 | 12.32 | -1.02 | 1.00 |

Note. The range of possible working memory scores is 0-48 for the reading span, 0-27 for the symmetry span, and 0-40 for the operation span. Processing accuracy refers to the percentage of correct answers.

Working memory and processing accuracy scores as a function of percentile in the sample are presented in Table 2. These data confirm the presence of a floor effect for processing accuracy on the reading span and symmetry span and the absence of this floor effect for processing accuracy on the operation span. In the latter case, most participants demonstrated adequate performance on the processing task except for participants in the bottom 5th percentile who scored barely above chance level. No floor or ceiling effect appeared for working memory scores on any subtest.

Table 2. *Percentiles for working memory and processing accuracy scores.*

| Measure | 5 th | 25 th | 33.3 th | 50 th | 66.6 th | 75 th | 95 th |
|--------------------------------|-----------------|------------------|--------------------|------------------|--------------------|------------------|------------------|
| Composite working memory score | -1.41 | -0.52 | -0.33 | 0.04 | 0.37 | 0.61 | 1.14 |
| Reading span | | | | | | | |
| Working memory score | 18 | 26 | 28 | 32 | 37 | 39 | 46 |
| Processing accuracy score | 71 | 85.5 | 87.5 | 89.5 | 93.75 | 95 | 98 |
| Symmetry span | | | | | | | |
| Working memory score | 9 | 14 | 16 | 19 | 22 | 23 | 26 |
| Processing accuracy score | 74 | 85.25 | 89 | 92.5 | 92.5 | 96.5 | 100 |
| Operation span | | | | | | | |
| Working memory score | 15 | 23 | 27 | 30 | 33 | 34 | 39 |
| Processing accuracy score | 57.5 | 75 | 77.5 | 82.5 | 87.5 | 90 | 97.5 |

Note. The range of possible working memory scores is 0-48 for the reading span, 0-27 for the symmetry span, and 0-40 for the operation span. Processing accuracy refers to the percentage of correct answers.

Reliability

Internal consistency of the working memory scores was computed for each subtest with the Kane et al. (2004) method: the proportion of correctly recalled stimuli was calculated for each trial and a Cronbach's α was calculated across all trials. The values of Cronbach's α were satisfying, with values above .70 for the reading span ($\alpha = .72$), the symmetry span ($\alpha = .72$) and the operation span ($\alpha = .76$). These values are comparable to the coefficients reported by Redick et al. (2012), indicating that the decrease in the number of trials did not critically affect the reliability of the subtests. An omega total coefficient was also computed to estimate the internal consistency of the full scale; this coefficient is similar

to Cronbach's alpha but offers a better estimate of reliability for multidimensional scales, as is the case here (see Revelle & Zinbarg, 2009). Internal consistency was even higher for the full scale than for the subtests ($\omega_t = .86$).

Test-retest reliability of the working memory scores was calculated as the correlation between scores on the first session and scores on the second session. The length of time between the two administrations of the task was added as a covariable in the analysis. Correlation coefficients were moderate for the reading span, $r(285) = .61$, the symmetry span, $r(286) = .69$, and the operation span, $r(283) = .66$. These values are lower than the test-retest reliability coefficients reported by Redick et al. (2012). However, test-retest reliability was higher and above .70 for the composite working memory score, $r(298) = .77$; this value is similar to the results reported in Redick et al. (2012) and indicates satisfying test-retest reliability.

On average, working memory scores were higher on the second session for the symmetry span, the operation span, and the composite working memory score (all $ps < .001$), indicating a practice effect. However, the effect was relatively small; on average, participants recalled 1.6 more stimuli on the second session of the symmetry span (out of a total of 27) and 1.7 more stimuli on the second session of the operation span (out of a total of 48). The practice effect did not reach significance for the reading span, $F(1, 285) = 2.44$, $p = .12$, $\eta^2_p = .01$; on average, participants recalled 0.8 more stimuli on the second session of this subtest (out of a total of 48).

Validity

Convergent validity was assessed by examining the correlations between the three subtests. For reference, Redick et al. (2012) reported the following average correlation coefficients between the reading span, symmetry span and operation span in four different samples: $r = .46$ for the reading span and symmetry span, $r = .63$ for the reading span and

operation span, and $r = .47$ for the symmetry span and operation span. In the CCS, working memory scores were moderately correlated across the three subtests (see Table 3). As can be seen, these correlation coefficients are lower than those reported by Redick et al. (2012), but not disproportionately so, suggesting that the short versions of the subtests retained satisfying validity.

Table 3. *Cross-task correlations for the working memory and convergent validity measures*

| Measure | Reading span | Symmetry span | Operation span | Alpha span | Raven's APM |
|-----------------|--------------|---------------|----------------|------------|-------------|
| Reading span | - | .33 | .53 | .45 | .18 |
| Symmetry span | .33 | - | .38 | .41 | .37 |
| Operation span | .53 | .38 | - | .48 | .20 |
| Composite score | .80 | .75 | .82 | .54 | .39 |

Concurrent validity was assessed as the correlation between the working memory scores and performance on the APM and the alpha span task (see Table 3). As expected, the working memory composite score correlated with Raven's APM, $r(182) = .39, p < .001$. This correlation is close to usually observed values: Redick et al. (2012) reported an average coefficient of $r = .36$ for the correlation between complex span tasks and Raven's matrices in 11 different samples. Performance on the three individual subtests of the CCS also correlated with the APM, although the correlation coefficients were much lower than for the global score. The working memory composite score also correlated with the alpha span, $r(247) = .54, p < .001$. Again, this correlation is close to the expected value: for example, Oberauer et al. (2000) reported a .49 correlation between a similar alpha span task and a version of the reading span. Performance on the three subtests also correlated with the alpha span; the correlations for the subtests and the composite score were close in magnitude.

Confirmatory factor analysis

To provide a more powerful test of the internal consistency and convergent validity of the CCS, the data were submitted to a confirmatory factor analysis (CFA; for a similar procedure, see Lewandowsky, Oberauer, Yang, & Ecker, 2010). All items in a subtest were assumed to load on a latent variable representing the score on this subtest, and the three latent variables representing the three subtests were assumed to load on a general factor representing working memory capacity. Measurement errors for each item were assumed to be uncorrelated. The resulting model is represented in Figure 2. The fit of this model was excellent [$\chi^2(206) = 286.72$, $p < .001$; $\chi^2/df = 1.39$; comparative fit index (CFI) = 0.979; root-mean square error of approximation (RMSEA) = 0.02; standardized root mean square residual (SRMR) = 0.029; see Hu & Bentler, 1999, for details on the fit indices]. All items in each subtest loaded on their respective latent variables, and the latent variables for each subtest loaded on the general factor representing working memory capacity. In other words, the three subtests of the CCS demonstrated both internal consistency and convergent validity.

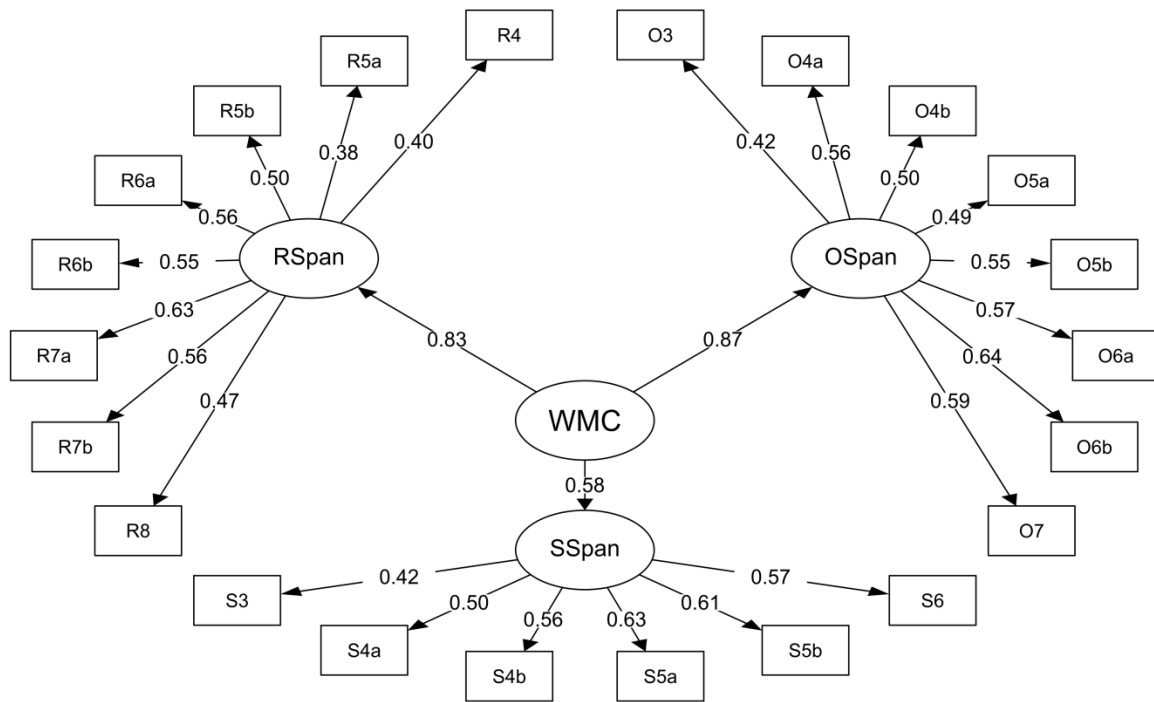


Figure 2. Confirmatory factor analysis model for the CCS. All correlations and loadings are standardized estimates. WMC = working memory capacity; RSpan = reading span; SSpan = symmetry span; OSpan = operation span. For each item, the uppercase letter indicates the subtest and the digit indicates set size. Measurement errors are not depicted.

Psychometric properties as a function of exclusion criteria

The CCS does not use the same exclusion criteria as the classic version of the tasks (Unsworth et al., 2005): the data of a participant on a subtest are only excluded if this participant scores in the bottom 5th percentile for processing accuracy, rather than if the participant scores below 85% processing accuracy. As presented in Table 2, this resulted in much more lenient exclusion criteria in our sample: for example, only participants scoring below 57.5% processing accuracy were excluded on the operation span. This raises the possibility that the CCS might have demonstrated different psychometric properties when using the more stringent criterion of 85% accuracy. In order to test this possibility, the data

were re-analyzed for each subtest separately after excluding participants scoring below 85% accuracy. Although this procedure resulted in the exclusion of a large number of participants (more than half the sample for the operation span), the three subtests demonstrated comparable psychometric properties (see the results in Table 4). In other words, excluding only participants who scored in the bottom 5th percentile did not seem to alter the psychometric qualities of the task¹.

Table 4. *Psychometric qualities of the CCS as a function of exclusion criteria*

| Measure | Reading span | N | Symmetry span | N | Operation span | N |
|---------------------|--------------|----------|---------------|----------|----------------|----------|
| Cronbach's α | .72/.68 | 1026/847 | .72/.72 | 1023/894 | .76/.69 | 1010/537 |
| Test-retest | .61/.56 | 286/221 | .69/.60 | 287/229 | .66/.66 | 284/91 |
| APM | .18/.13 | 173/153 | .37/.36 | 178/156 | .20/.18 | 175/91 |
| Alpha span | .45/.45 | 234/212 | .41/.39 | 233/218 | .48/.53 | 230/160 |

Note. The values are presented for the two different exclusion criteria: participants scoring in the bottom 5th percentile for processing accuracy / participants scoring below 85% processing accuracy, respectively. Test-retest = correlation between test and re-test performance; APM = correlation with Raven's APM; Alpha span = correlation with the alpha span.

Discussion

This article presented the CCS, a composite working memory task including short versions of three complex spans, the reading span, symmetry span and operation span. The CCS demonstrated satisfying reliability and validity. Observed values for internal consistency, test-retest stability, and concurrent validity were quite close to the values reported for English-speaking versions of the subtests (Redick et al., 2012). Performance on the CCS appeared relatively stable in time, and the task showed the expected correlations

¹ Two alternative exclusion criteria were also tested: processing accuracy less than 65% and processing accuracy less than 75%. The psychometric properties of the CCS were virtually unaffected in both cases.

with Raven's APM and with an alpha span task. Overall, the CCS seems to constitute an adequate task to measure domain-general working memory capacity in French-speaking samples. Despite including only half as many trials in total as the three classic computerized versions of the subtests (Unsworth et al., 2005), the CCS demonstrates similar psychometric properties. The satisfying qualities of the CCS indicate that short versions of complex span tasks may be used to provide an accurate measure of domain-general working memory, congruent with the conclusions of recent works (Foster et al., 2014; Oswald et al., in press). In other words, it is not necessary to have participants complete full versions of multiple complex spans to obtain a valid measure of their working memory capacity. In this respect, the CCS parallels the shortened working memory task developed by Oswald and colleagues, with the added benefit of being available for French-speaking samples and demonstrating the necessity of population-specific exclusion criteria.

It should be noted that the composite working memory score is more reliable and more valid than scores on the individual subtests; this reflects the fact that the CCS should be viewed as a unitary task assessing domain-general working memory, rather than as a task battery assessing working memory for different types of materials. Similarly, even though assessing domain-specific working memory capacity may be of interest, the CCS should not be decomposed into verbal and spatial subtests due to the limited psychometric value of individual subtests. With only one or two subtests per domain, it is also likely that this approach would yield task-specific rather than domain-specific estimates of working memory capacity; in order to obtain a valid measure of domain-specific working memory capacity, it would be recommended to use at least three tasks per domain (Foster et al., 2014).

The only major difference between the CCS and original versions of the three complex spans resides in performance of the participants on the processing tasks. Published versions of English-speaking complex spans typically recommend excluding participants who

score lower than 85% on the processing task (Conway et al., 2005; Unsworth et al., 2005), which results for example in about 15% of exclusions for the operation span in American samples (Unsworth et al., 2005). As can be seen in Table 3, applying the same criterion in our sample would result in excluding approximately 25% of participants on the reading span and symmetry span and more than 50% of participants on the operation span subtest. Why such a discrepancy? The instructions, the practice phases and the difficulty of the processing tasks are all identical in the CCS and in the original versions of the complex spans. The most likely explanation is a true difference between the samples; for the operation span subtest in particular, a significant portion of French psychology students come from Arts divisions and are ill-at-ease with mathematical operations. The fact that complex spans have reduced validity when the processing task is too difficult for participants (Turner & Engle, 1989) may be a cause for concern. However, most participants in our sample appeared to adequately carry out the processing tasks, the global CCS score demonstrated satisfying validity, and using the original exclusion criteria did not significantly alter the psychometric qualities of the task. For these reasons, the best solution is probably to retain the same processing task difficulty as in the original versions of the tasks for the sake of comparability, but to adopt less stringent exclusion criteria.

Interestingly, the fact that there exist significant sample differences in processing accuracy on complex spans, even between two populations of undergraduate university students, also suggests that the prescribed exclusion criteria should not be applied indiscriminately, even for the original version of the task. Indeed, it is likely that the proportion of participants achieving 85% accuracy in the task would be much lower in certain populations, such as clinical patients. In this respect, exclusion criteria should be adapted to the specific population being considered. The solution adopted here, discarding the data of a

subtest for participants who score in the bottom 5th percentile in the processing task, seems to be an adequate choice.

The CCS relies on the idea that combining working memory tasks related to different types of materials is a great way to eliminate content-specific variance and to obtain a domain-general measure of working memory capacity (Kane et al., 2004). However, all three subtests in the CCS use the same complex span structure; as a consequence, it is likely that performance in the CCS still includes method-specific variance. Complex span tasks are not the only adequate working memory measures: a wide variety of very different tasks can also yield useful estimates of working memory capacity, even tasks without clear processing and storage requirements (Oberauer, 2005). To obtain a truly general measure of working memory capacity, it may be desirable to combine complex span tasks with other working memory tasks (Redick et al., 2012). Since the alpha span is not a complex span task and demonstrates a significant correlation with the CCS, replacing the reading span subtest with the alpha span may partially solve this problem in studies where limiting method-specific variance is important.

In summary, the CCS constitutes a short working memory task suitable for obtaining a domain-general estimate of working memory capacity. Despite being shorter than classic complex span tasks, the CCS demonstrated satisfying psychometric properties in a large French sample.

Acknowledgments

The CCS and the alpha span task scripts are readily available from the authors or at the following address: <https://osf.io/bk7pm/>. We would like to thank Randy Engle and Martial van der Linden for kindly authorizing the use of material from their versions of the symmetry span and reading span, respectively.

References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working Memory and Intelligence: The Same or Different Constructs? *Psychological Bulletin*, 131(1), 30-60. doi: 10.1037/0033-2909.131.1.30
- Conway, A. A., Kane, M. J., Bunting, M. F., Hambrick, D., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769-786. doi:10.3758/BF03196772
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450-466. doi: 10.1016/s0022-5371(80)90312-6
- Delaloye, C. C., Ludwig, C. C., Borella, E. E., Chicherio, C. C., & de Ribaupierre, A. A. (2008). L'Empan de lecture comme épreuve mesurant la capacité de mémoire de travail: Normes basées sur une population francophone de 775 adultes jeunes et âgés. = The Reading Span as a measure of working memory capacity: Norms based on a French speaking population of 775 younger and older adults. *European Review Of Applied Psychology / Revue Européenne De Psychologie Appliquée*, 58(2), 89-103. doi:10.1016/j.erap.2006.12.004
- Desmette, D., Hupet, M., Schelstraete, M., & van der Linden, M. (1995). Adaptation en langue française du 'Reading Span Test' de Daneman et Carpenter (1980). = A French version of M. Daneman and P. A. Carpenter's (1980) Reading Span Test (RST). *L'année Psychologique*, 95(3), 459-482. doi:10.3406/psy.1995.28842
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning & Motivation*(44), 145.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309-331. doi: 10.1037/0096-3445.128.3.309
- Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2014). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, doi:10.3758/s13421-014-0461-7
- Fournet, N., Roulin, J., Vallet, F., Beaudoin, M., Agrigoroaei, S., Paignon, A., & ... Desrichard, O. (2012). Evaluating short-term and working memory in older adults: French normative data. *Aging & Mental Health*, 16(7), 922-930. doi:10.1080/13607863.2012.674487
- Healey, M., Hasher, L., & Danilova, E. (2011). The stability of working memory: Do previous tasks influence complex span?. *Journal Of Experimental Psychology: General*, 140(4), 573-585. doi:10.1037/a0024587

- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. doi:10.1080/10705519909540118
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, 18(4), 394-412. doi:10.1080/09658211003702171
- Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(2), 336-358. doi: 10.1037/0278-7393.26.2.336
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, 133(2), 189-217. doi: 10.1037/0096-3445.133.2.189
- Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments & Computers*, 31(3), 429-432. doi:10.3758/BF03200722
- Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, 42(2), 571-585. doi:10.3758/BRM.42.2.571
- Lucidi, A., Loaiza, V., Camos, V., & Barrouillet, P. (2014). Assessing working memory capacity through time-constrained elementary activities. *Journal Of General Psychology*, 141(2), 98-112. doi:10.1080/00221309.2013.870121
- Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal Of Experimental Psychology: General*, 130(2), 199-207. doi:10.1037/0096-3445.130.2.199
- May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition*, 27(5), 759-767. doi:10.3758/BF03198529
- McVay, J. C., & Kane, M. J. (2012). Drifting from slow to “d'oh!”: Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 525-549. doi: 10.1037/a0025896
- Oberauer, K. (2005). The Measurement of Working Memory Capacity. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 393-407). Thousand Oaks, CA, US: Sage Publications, Inc. doi:10.4135/9781452233529.n22
- Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and individual differences*, 29(6), 1017-1045. doi:10.1016/S0191-8869(99)00251-2

- Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (in press). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*.
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Section 4, Advanced Progressive Matrices, 1998 edition*. Oxford, UK: Oxford Psychologists Press.
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal Of Psychological Assessment*, 28(3), 164-171. doi:10.1027/1015-5759/a000123
- Redick, T. S., & Engle, R. W. (2011). Integrating working memory capacity and context-processing views of cognitive control. *The Quarterly Journal of Experimental Psychology*, 64(6), 1048-1055. doi: 10.1080/17470218.2011.577226
- Redick, T. S., & Lindsey, D. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review*, 20(6), 1102-1113. doi:10.3758/s13423-013-0453-9
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145-154. doi:10.1007/s11336-008-9102-z
- Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal Of Experimental Psychology: General*, 136(2), 241-255. doi:10.1037/0096-3445.136.2.241
- St Clair-Thompson, H. (2012). Ascending versus randomised list lengths in working memory span tasks. *Journal Of Cognitive Psychology*, 24(3), 335-341. doi:10.1080/20445911.2011.639760
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2), 127-154. doi: 10.1016/0749-596x(89)90040-5
- Unsworth, N., Brewer, G. A., & Spillers, G. J. (2011). Variation in working memory capacity and episodic memory: Examining the importance of encoding specificity. *Psychonomic bulletin & review*, 18(6), 1113-1118. doi: 10.3758/s13423-011-0165-y
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498-505. doi:10.3758/BF03192720