BRIEF REPORT

Working memory complex span tasks and fluid intelligence: Does the positional structure of the task matter?

Miriam Debraise[1] · Nicolas Gauvrit[2] · Fabien Mathy[1]

Abstract
The complex span task used to evaluate working memory (WM) capacity has been considered to be the most predictive task of fluid intelligence. However, the structure of the complex span tasks varies from one study to another, and it has not been questioned yet whether these variants could influence the predictive power of these tasks. Previous studies have typically used either structures based on alternating processing-storage patterns or alternating storage-processing patterns. We present one experiment in which the participants were submitted to both the processing-storage vs. storage-processing types. After completing both types of complex span tasks, the participants performed a reasoning test (Matrix Reasoning of the Wechsler Adult Intelligence Scale - WAIS-IV). The results showed a significant difference in the WM spans between the two conditions, with higher spans observed in the processing-storage alternating structure, and different serial position curves. However, the correlations showed that both types of tasks remained equally predictive of performance in the reasoning test. These results are discussed in regard to the time-based resource-sharing model.

Keywords Working Memory · Complex span task · Higher order cognition

The complex span tasks used to assess working memory capacity are predictive of many aspects of higher-order cognition. However, the structures of these tasks vary from one study to another, and it has never been called into question whether these variations could influence their predictive power. Previous research has exclusively used two types of complex span task structures, either those based on alternating storage-processing patterns (e.g., the operation span task of Unsworth, Heitz, Schrock, & Engle, (2005)) or alternating storage-processing patterns (e.g., the operation span task used by Barrouillet and Camos (2001)). The difference between these two types of tasks relies on whether the task begins with a processing episode that captures attention (e.g., solving mathematical operations, reading digits aloud) or by a storage episode involving the maintenance of presented items (e.g., letters, words, digits, shapes). We believe that the potential effect of the various

structures of complex span tasks should not be overlooked for two main reasons.

First, complex span tasks are extensively used in correlational studies, for instance, in examining the relationship between the span and fluid intelligence. Depending on the material used in complex span tasks, variations from moderate to strong correlations have been found (e.g., Kane et al., 2004; Kanerva & Kalakoski, 2016; Lucidi, Loaiza, Camos, & Barrouillet, 2014; Unsworth, Redick, Heitz, Broadway, & Engle, 2009). Our hypothesis is that the structure of the task could also influence the strength of the correlations. It is unclear why complex span tasks are so predictive of higher-order cognition (Conway, Kane, & Engle, 2003; Kane, Hambrick, & Conway, 2005; Unsworth et al., 2009). According to the dual-component model of Unsworth and Engle (2007b) and Unsworth and Engle (2007a), attention control and secondary memory are crucial mechanisms for both working memory and fluid abilities. Attention control enables the keeping of relevant information accessible despite interference, and secondary memory allows access to long-term memory, and they both contribute distinctively to higher-order abilities (Unsworth & Spillers, 2010). In a recent article, Engle (2018) specifies that attention could operate differently on working memory and higher order

✉ Miriam Debraise
Miriam.DEBRAISE@univ-cotedazur.fr

[1] Université Côte d'Azur, CNRS, BCL, Nice, France

[2] Université de Lille, Lille, France

abilities. In working memory tasks, attention control primarily enables the maintenance of information while other irrelevant thoughts are being suppressed. In a reasoning task (e.g., Raven's Progressive Matrices (Raven, 1962); Matrix Reasoning of the WAIS-IV (Wechsler, 2011)) the main role of attention control is to allow participants to disengage from information once it turns out to be irrelevant. However, maintenance processes are also involved in reasoning tasks, but in a lesser degree. Although not completely incompatible with this previous account, the Time-Based Resource-Sharing model (TBRS, see Barrouillet, Bernardin, & Camos, 2004) supposes that the relation between working memory capacity and intelligence is best explained by the fact that both constructs require switching between processing and storage within the time constraints of the task at hand. In that sense, general purpose resource sharing is essential, rather than the combined effect of attentional control and secondary memory (Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Barrouillet & Camos, 2007; Barrouillet, Lepine, & Camos, 2008; Lucidi et al., 2014). Studying the structure of the span task could therefore shed light on the mechanisms underlying the correlations with intelligence.

Second, but relating to the first point, we think that manipulating the structure of the task could help decide between models when they are implemented in various versions (e.g., Gauvrit & Mathy, 2018; Lemaire & Portrat, 2018; Oberauer & Lewandowsky, 2011). It is possible that one of the structures weights more on a given process (i.e., storage vs processing) at different positions (i.e., potentially impacting primacy and recency effects; see Unsworth and Engle (2006)), and the structural variations could thus impact empirical correlations and the fit of alternative models. The comparison of the fit of alternative models is beyond the scope of the present study. Instead, we evaluate the possible impact of the structure of the task on working memory performance, and its relation to fluid intelligence.

Working memory complex span tasks

The operation span task of Unsworth et al. (2005) is an example of a complex span task based on a processing-storage repeating pattern, where the participant is invited to solve mathematical operations that are interspersed with items to remember. Crucially in this type of task, the task begins with a processing episode, and the task ends with a storage episode. The operation span task has also been used by Barrouillet and Camos (2001), but the structure of the task is reversed, as it starts with the storage episode and ends with the processing episode. This difference warrants attention because the complex span tasks may be the most popular task for studying working memory (Aben, Stapert,

& Blokland, 2012; Barrouillet & Camos, 2012; Barrouillet, Portrat, & Camos, 2011; Colom, Rebollo, Abad, & Shih, 2006; Dunlosky & Kane, 2007; Gathercole, 1999; Friedman & Miyake, 2005; Lewandowsky, Oberauer, Yang, & Ecker, 2010; Mathy, Chekaf, & Cowan, 2018; Miyake et al., 2000; Oberauer, 2009; Oberauer et al., 2018; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012; Ricker, Vergauwe, Hinrichs, Blume, & Cowan, 2015; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009; Wang, Ren, Li, & Schweizer, 2015), and this task offers interesting theoretical insight when its temporal structure is manipulated (Bailey, 2012; Loaiza & Camos, 2016; McCabe, 2010). However, the choice of one of the two structures is generally adopted without clear justification. For instance, Stone and Towse (2015) mention that "the *traditional* method of administering complex span tasks such as the operation span task involves using a processing-storage order of phases rather than storage-processing (...). This method is rather curious as the processing task serves the purpose of adding to the cognitive demands of storage by requiring processing of a task while trying to store stimuli."

As explained above, the executive account of working memory (e.g., Engle, 2002, 2018; Engle, Kane, & Tuholski, 1999) posits that WM capacity is essentially defined by the ability to maintain controlled attention (in order to allow storage processes to take place) despite interference. Thus, the structural variations of the processing items (at least for medium and longer list lengths) should not influence working memory capacity because the demands in controlled attention should be equal in a processing-storage and a storage-processing complex span task. For instance, the processing-storage version of the task should not necessarily benefit the last item because the requirement to recall all of the items in order could sufficiently equalize the maintenance of information requirement.

In contrast, according to the TBRS model (Barrouillet et al., 2004) working memory spans are dependent of the cognitive load (CL) of the task at hand. Cognitive load is defined as the proportion of time during which a given task occupies attention during memory retention, preventing thus maintenance processes to occur. Several studies have found a linear relationship between the complex span and the cognitive load of the task (e.g., Barrouillet et at. 2004, 2007; Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009; Barrouillet et al., 2011). Thus, the greater the proportion of time during which attention is fully occupied (i.e., the higher CL), the lower the memory performance. Experimentally, cognitive load is manipulated by varying either the duration of the attentional capture, the number of processing steps and/or by changing the total amount of time available for the participant to perform the processing task. For instance, increasing the number of operations to solve in a complex span task while keeping the other

parameters constant increases the CL of the task at hand. Following this rationale, the processing-storage tasks should overestimate working memory capacity by minimizing the processing demand, as the first item of the task (i.e., solving a mathematical operation) is memory load free. Consequently, this first mathematical operation and its duration should be insignificant to the cognitive load of the task, whereas the same operation in the storage-processing version of the task should augment the cognitive load. In sum, the TBRS account can more easily predict differences between the two types of tasks than the above executive account.

Recently, a mathematical transcription (TBRS2) of the TBRS has been developed by Gauvrit and Mathy (2018), which enables the testing of the predictions of the simplest version of the verbal model. Contrary to other computational models (e.g., TBRS* of Lemaire and Portrat (2018) and Oberauer and Lewandowsky (2011)), the TBRS2 is minimalist in that it has only been developed to quantify the cognitive load using the minimal assumptions of the original TBRS model. Furthermore, the TBRS2 allows only a few parameters to vary to account for variations in the speed of the different WM processes (e.g., refreshment, decay). Figure 1 (top vs bottom) shows the predictions of the TBRS2 when the structure of the complex span task was manipulated. The top of each plot displays the structure of the task, with the events depicted horizontally as a function of time. Each new colored event on the top row of each plot represents a new to-be-remembered item. All of the black events represent the successive events of the concurrent task. The line entitled 'Focus' represents the letters that are successively at the center of attention. We can see that free time allows the different memory traces to be refreshed alternatively. The curves below show the level of activation of each of the to-be-remembered items. After 15 seconds (right side of the frame), the level of activation predicts the probability of recalling each of the items.

As seen from the numbers that are reported in the figure's note, the probability of the correct recall of the letters in the processing-storage type of task (top) is on average higher. These differences seem low using the default parameters, but these numbers could be better differentiated by using different sets of parameters. However, crucially, our prediction is that because the processing-storage type is always granted with a lower cognitive load, participants' spans should be higher. This is intuitive, as the last processing episode of the storage-processing type of task is detrimental to the recall process. In that respect, we can also
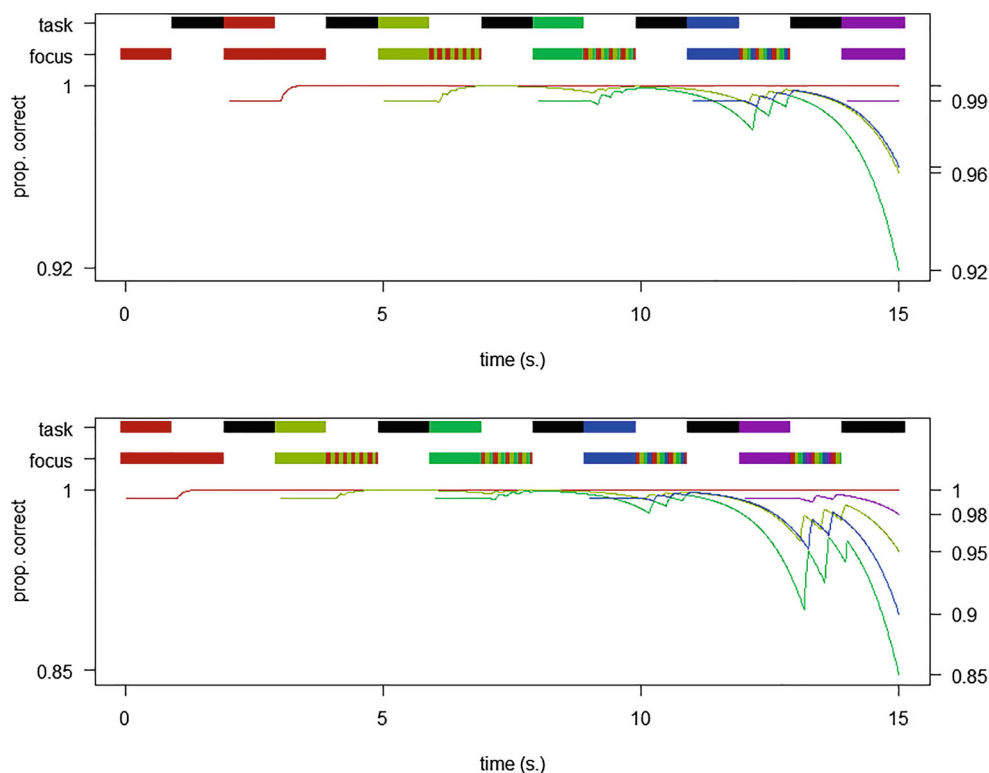

Fig. 1 *Top*: Predictions of the TBRS2 for an alternating processing-storage type of task. The probability of correct recall for the successive five letters are: 0.9999766, 0.961432, 0.9186392, 0.9642913, 0.9933071. *Bottom*: Predictions of the TBRS2 for the storage-processing condition: 0.9999683, 0.9484365, 0.8460564, 0.8965995, 0.9794474. Both simulations involved a sequence of five letters and the parameters were set as follows: Duration 80 ms, baseline 5, $d = 1.4$, $r = 9$

expect that the more demanding storage-processing type of tasks are better predictors of higher-order cognition.

Method

In the present experiment, the participants were submitted to both processing-storage and storage-processing conditions of an operation span task. After completing a simple span task (i.e., a span task without any concurrent events) and both the conditions of the complex operation span task (counterbalanced between the participants), the participants then performed a reasoning test (Matrix Reasoning of the WAIS-IV (Wechsler, 2011)). The two versions of the operation span task (OSpan) were adapted from Unsworth et al. (2005) and Barrouillet and Camos (2001), respectively. However, the timing of both conditions was based on that of Unsworth et al. (2005). Note that here, we only focus on the performance for the complex span tasks and the reasoning test, as the simple span task was intended to serve as a pretest for another study. We therefore consider that the simple span task was used herein as a warmup.

Participants

A total of 204 young adults (164 females, 38 males; mean age = 21.52 years, SD = 3.9) participated in this experiment. Most of the participants were second-year students of Université Côte d'Azur and received partial course credit for participating. The duration of the experiment was approximately 1 h.

Regarding power, we followed recommendation of Schönbrodt and Perugini (2013) to reach a sufficiently large sample of participants to obtain stabilized correlation values. After collecting data on $N = 83$ participants, we obtained the following estimates: $r = .409$ for the processing-storage condition and $r = .367$ for the storage-processing condition. Based on these values corresponding to the rounded estimate $\rho = .4$, we followed the authors' recommendation to collect data based on at least $N = 150$ participants for a corridor of stability of width $w = .15$ and a level of confidence equal to 95%.[1] Also, we made sure we could obtain a significant difference between the two estimates, with plausible correlation values. Based on the lowest value $r = .367$, we verified that we had reached a sufficient sample size using the test for two dependent Pearson correlations in G*Power (Faul, Erdfelder, Buchner, & Lang 2009). The computation showed that the second correlation should be at least .12 higher than the smaller

---

[1] The range of sample sizes was between $N = 342$ for $w = .10$ and $N = 84$ for $w = .20$, but the larger value exceeded our pool of participants who could obtain course credits in exchange of their participation the year we finished conducting the study in 2020.

one with $N = 180$ participants to reach a power of .80, which seemed a plausible difference to be observed between different measures of the span. Finally, we also computed the minimal sample size to reach a power of .80 in our mediation analysis using the function runGitHub in R, which indicated a sample size of $N = 170$.

Procedure and material

The simple span task consisted of three trials for each list of 3 to 9 letters randomly chosen without replacement from the following set of consonants: F, H, J, K, L, P, Q, R, B, S, V, X. Each trial started with a fixation cross displayed for 500 ms centered on the screen, followed by the to-be-remembered letters, each presented for 1000 ms. At the end of each trial, a matrix with all the consonants appeared at the center of the screen. The participants were then invited to recall the letters in the correct order by clicking on the letters. A feedback of the number of letters correctly recalled for the current set was presented on the screen before a new trial started. Five practice trials of a set size of two letters were administered before the simple span task began. The set of to-be-remembered consonants, their duration and the recall phase were identical in the complex span tasks that followed. Before the first complex span task, a training block of mathematical operations took place where the participant had to solve 16 mathematical operations as fast as possible. These trials began with a cross presented for 500 ms followed by a chosen mathematical operation (e.g., $(2 \times 6) - 4 = ?$). After solving the operation mentally, the participant was instructed to click on the mouse in order to pass to the next screen where a digit (e.g., 8) and the words 'VRAI' and 'FAUX' ('TRUE' and 'FALSE' in French, respectively) were displayed. The participant was requested to click on the correct answer, again as quickly as possible. In this example, the participant would have to click on the 'VRAI' box in order to be correct. After each operation was solved, the percentage of correctly solved mathematical operations was updated and displayed on the right corner of the screen. The participants were instructed that they had to reach an 85% success rate so that their results could be included in the study. The aim of this mathematical operation training was to familiarize the participants with the concurrent task but also to measure the average individual's response times (RT) to tailor the main task, as explained later. Similar mathematical operations (i.e., same kind of difficulty) were used in the complex span tasks. Both complex span tasks had similar structure in terms of cognitive load, distractors, and memoranda. Both complex span tasks used list lengths varying from 3 to 7 consonants and three trials per list length. The 15 trials (5 lengths × 3 trials) within each complex span task were randomized for each participant. Before each complex span
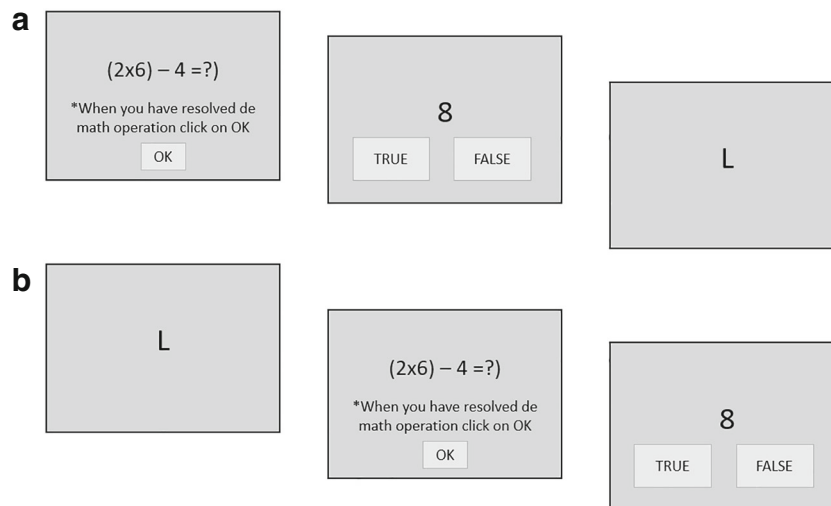
Fig. 2 Illustration of two types of structure in the operation span task. A = processing-storage task; B = storage-processing task. *Note*. Instructions were in French in the experiment

task, the participants practiced with two trials of length two (using either the processing-storage order or the storage-processing order).

The only variation between the two types of task was the order of the processing and storage events, as seen from Figure 2. The processing-storage task (Unsworth et al., 2005) consisted of presenting the mathematical operations before the consonants. Each trial began with a fixation cross presented for 500 ms followed by a mathematical operation. After solving the mathematical operation, a letter was presented on the screen for 1000 ms, followed by a new operation and so on, until the end of the trial and the recall phase. The participants were instructed that their available time to solve the mathematical operation would be limited to pace the task. As in Unsworth et al. (2005), we used the distribution of the individual RTs obtained during the practice trials to define the available duration based on their average RTs plus 2.5 SD. When the participants did not solve the operation within the available duration, the task resumed as if the participant had actually responded (the program skipped the screen displaying the response options 'VRAI' and 'FAUX'), and the program presented the next event (i.e., a consonant). Then, the trial continued its cycle as it was planned, but the missed mathematical operation was counted as an error. The storage-processing condition (Barrouillet & Camos, 2001) consisted of presenting the mathematical consonants before the operations. The procedure was identical to the processing-storage condition except that each trial began with a consonant. After both complex span tasks were completed, the reasoning matrix subtest of the WAIS-IV was conducted. This test is a nonverbal reasoning task consisting of 26 items including three practice items. For each item, a matrix of colored figures is presented, and the participant must find the missing figure among five response options. This test was not performed for all participants, because some of our participants had previously obtained the sufficient amount of course credits.

Results

The results from two participants were excluded from the analyses as they interrupted the experiment. No other participant was excluded from the analyses, although some did not achieve an 85% average solving rate on the mathematical operation. According to Unsworth et al. (2009), this exclusion criteria is unnecessary, as the processing accuracy is positively correlated to the storage accuracy. Thus, excluding participants with a low processing score may lead to a bias where low-span individuals are also excluded. Overall, the processing accuracy was at 89% (SD = 0.08) and 87% (SD = 0.08) for the processing-storage and storage-processing task, respectively. Among the 202 remaining participants, 178 participants completed the reasoning test.

The descriptive statistics for the spans and the results of the reasoning matrix test can be found in Table 1. The spans were calculated with the partial unit scoring. This score is calculated as the sum across the trials of the proportion of correct recall in the correct serial position within trials, divided by the number of trials per list length. For instance, using three trials per list length, if a participant failed only one trial at length 3 (because the second item was not correctly recalled) and then completely failed at all of the remaining trials of greater lengths, then the participant would be granted a span of $2 + (\frac{2}{3} \times 1 + \frac{1}{3} \times \frac{2}{3}) = 2.89$, with the first term 2 representing a span of 2 for the correct

Table 1 Descriptive statistics for span tasks ($N = 202$) and fluid intelligence ($N = 178$)

	Mean	SD	Range	Skewness	Kurtosis
Processing-storage	5.68	0.96	5	-1.12	1.16
Storage-processing	5.45	0.94	5	-0.85	0.54
Matrix reasoning	19.54	3.54	19	-0.87	0.52

responses across list lengths 1 and 2, and ($\frac{2}{3} \times 1 + \frac{1}{3} \times \frac{2}{3}$) representing performance at length 3 (or: $(1 + 1 + 1)/3 + (2 + 2 + 2)/6 + (3 + 3 + 2)/9 = 2.89$). This type of partial unit scoring essentially allows to equate the weight of the different set sizes (see. Conway et al. 2005; Kane et al. 2004). The scoring of the reasoning test was calculated as the sum of all the correct answers (one point for each correct answer).

We first analyzed the difference between the two types of tasks using both null hypothesis significance testing (NHST) and Bayesian analyses.[2] We provide the Bayes factors using equal priors for the considered alternative hypotheses to indicate their relative plausibility, knowing that a Bayes factors above 3 or below $\frac{1}{3}$) is substantial. Effectively, Bayes factors offer a simple expression of the degree to which data provide evidence for competing hypotheses across statistical tests (Morey et al., 2016).

The results showed that the mean span for the processing-storage task was found to be significantly higher compared to that of the storage-processing task ($t(201) = 4.46$, $p < .001$; $d = .31$; $95\%~CI = [.17, .46]$). Unsurprisingly, the Bayesian paired t test also showed evidence in favor of higher spans in the processing-storage condition ($BF_{10} = 911.3$).

Then, we analyzed performance as a function of item position, as we expected variations of performance due to the structural difference between the tasks caused by the processing-storage shift at the first and last positions. From this shift, we at least expected both higher performance at the first and last position for the processing-storage order, since the processing event could not disrupt the following storage event. However, the TBRS2 predicts a more complex pattern of differential performance from the first item to the last item, depending of the type of task. Figure 3

shows a gradient of systematic differences depending on item position, as the TBRS2 predicts that differences progressively increase until mid-list and then decrease until the last item. Our results effectively showed a gradient of systematic differences depending on item position, but the differences tended to progressively decrease until mid-list and then increased until the last item. To capture the interactions between item position and type of task, the data was analyzed with a linear mixed-effects regression model in R using the lme4 package (Bates et al., 2015). We did not aggregate the data, and thus we used the binomial family option to account for performance at each of the three trials per list length, depending on whether the recall of the item was correct or not at each position. For each sequence length (from 3 to 7 items), we compared five different models, depending on which factor was entered in the model (intercept alone, item position alone, type of task alone, item position + type of task, and item position * type of task). The participants were included as random intercepts in all of the models. Different increasingly complex models were tested by comparing their Akaike Information Criterion (AIC) values. The most parsimonious model (i.e., the best model with the lowest AIC) was chosen using the aictab function (library AICcmodvag). This procedure ranked as best the model including both factors and the interaction term for $L = 4$, $L = 6$, $L = 7$ with $p < .001$, and $L = 5$ with $p < .05$. The selected model for $L = 3$ was the one that included item position and type of task without their interaction.

Finally, we analyzed the relation between the two types of tasks and our measure of fluid intelligence. The correlations between span tasks and the reasoning subtest are presented in Fig. 4. The span for the processing-storage condition correlated significantly with the reasoning test ($r = .230$, $p < .002$), as did the storage-processing task ($r = .259$, $p = 0.001$). The Bayesian correlations showed evidence for a correlation with intelligence higher than zero for the processing-storage condition ($BF_{10} = 10.73$) and for the storage-processing condition ($BF_{10} = 40.16$). The spans in both processing-storage and storage-processing tasks correlated significantly with each other ($r = .722$; $p < .001$). The Steiger's test for dependent correlations showed no significant difference between the correlation between the storage-processing complex span

[2]While the orthodox NHST remains a common tool for drawing statistical inference from a sample, it has been widely criticized over the years mainly because of risks of type I errors given the null hypothesis (e.g., Cohen 1994). More recent approaches advocate for more appropriate methods developed by Bayesian psychologists. The Bayesian statistics can assess the relative plausibility of the null and alternative hypotheses while avoiding the several drawbacks of the NHST paradigm (Dienes, 2011; Gallistel, 2009; Wagenmakers et al. 2018). All of the Bayesian statistics were run in JASP (retrieved from http://jasp-stats.org/) with default parameters.
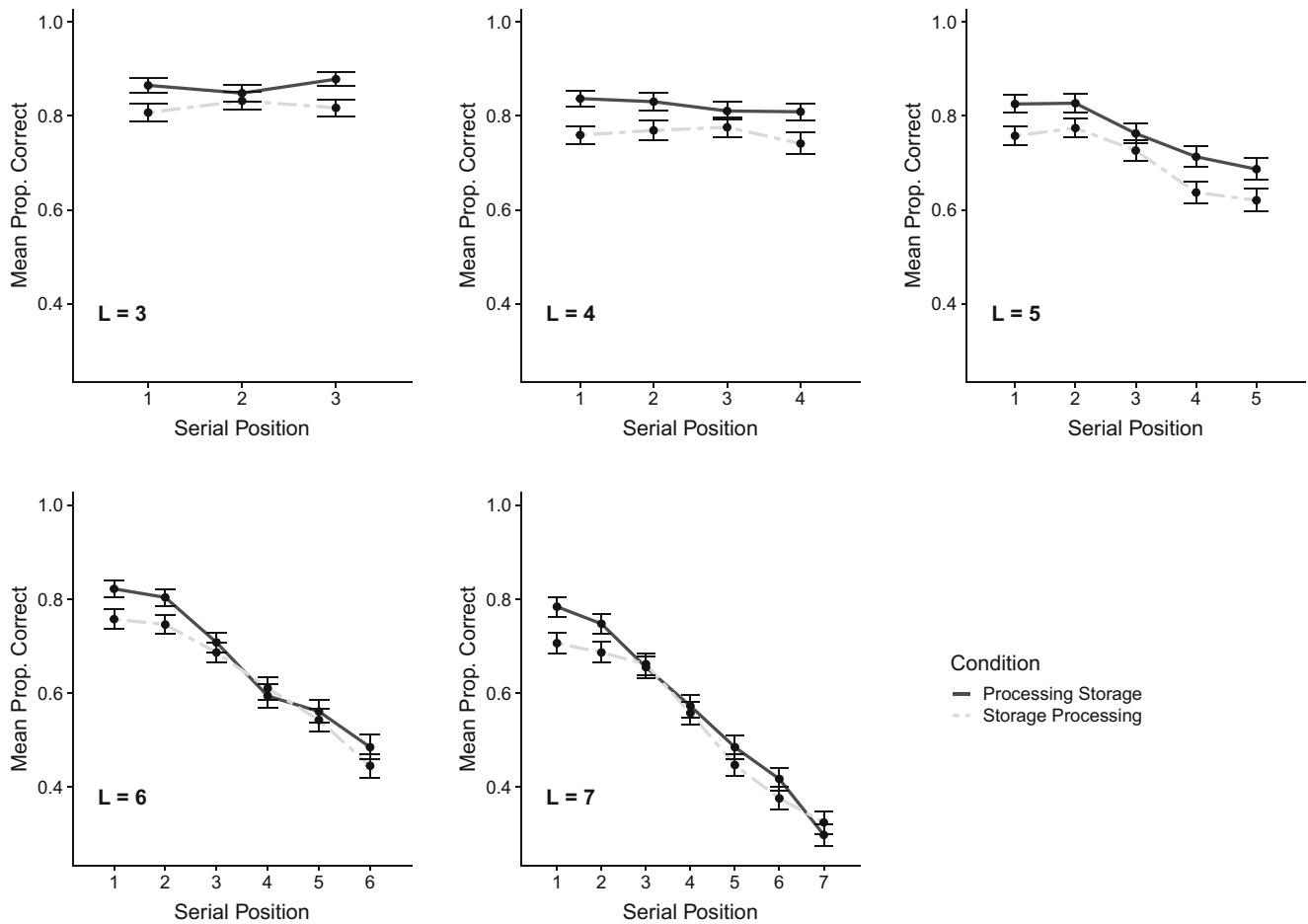
Fig. 3 Serial position curves as a function of type of task

task and the reasoning test and the correlation between the processing-storage complex span task and the reasoning test ($r_{diff} = .05$, $z = 1.0255$, $p = 0.305$).

We also attempted to use an index of the relative difficulty of the tasks by selecting for each participant their average performance for the first item positions (i.e., how well they perform for the first item in the storage-processing minus the processing-storage condition). Again, we observed no significant correlation that could have indicated that the participants who were the most sensitive to the increased difficulty of the storage-processing complex span task tended to have a lower IQ.

To better approach the relation between the two types of complex span tasks, we conducted a three-step mediating analysis by considering that the correlation between the processing-storage type of task and intelligence is mediated by the storage-processing type of task. The hypothesis was that this type of analysis could reveal an existing structural relation between the types of tasks. The regression coefficient between the processing-storage task and intelligence (direct effect) was significant (.8430, $p =$

.002). The indirect effect of the processing-storage task on intelligence was $(.72531) \times (.7092) = .51$, the value .72531 corresponding to the significant regression coefficient ($p < 2e - 16$) between the predicting Processing-Storage task and the resulting storage-processing task, and the value 0.7092 corresponding to the regression coefficient ($p = .07$) between the storage-processing task and intelligence in the multiple regression (i.e., controlling for the processing-storage task).

The significance of the indirect effect was tested using the R function mediation. Confidence intervals for the unstandardized indirect effects were computed based on 2000 bootstrap samples. The 95% confidence interval ranged from -0.00765 to 1.07 (the average bootstrapped unstandardized indirect effect was .52). This indirect effect was short of significance ($p = .05$) and the proportion mediated was equal to 61%, resulting from the values .52 (the indirect effect) divided by .84 (the total effect). In contrast, the average direct effect coefficient was equal to .32. This result means that the processing-storage type of task has a lower direct effect on intelligence.
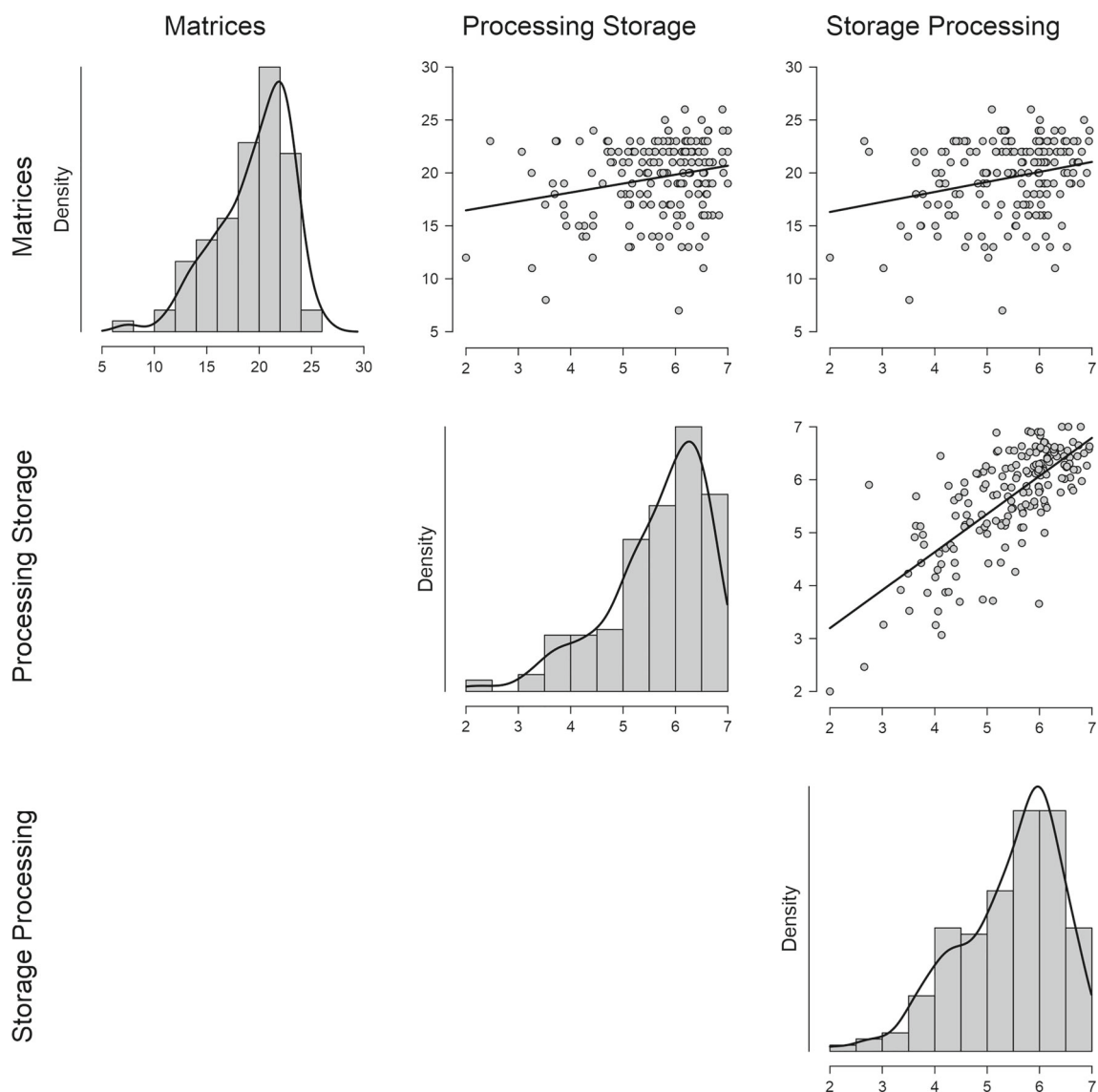
Fig. 4 Bayesian correlation chart for the complex span tasks (processing-storage or storage-processing) and the matrix reasoning subtest ($N = 178$)

Discussion

The current study examined whether the structure of complex span tasks (i.e., based on processing-storage vs. storage-processing cycles) could influence the measure of WM capacity and whether these variants could impact the predictive power of higher-order abilities. Our hypothesis of a more difficult task in the storage-processing condition was derived from simulating the TBRS model. Our experiment confirmed our hypothesis that the structural pattern of the complex span task modulates working memory capacity, with greater spans being obtained when the task starts with a processing event instead of a storage event. The two conditions also influenced differential recall patterns as a function of item position, but not exactly the way predicted by the TBRS2.

One explanation for this discrepancy between the expected serial positions and the observed data might be due to the way the model handles refreshing. The original TBRS model does not make any assumption on the refreshing schedule nor on its duration. Gauvrit and Mathy (2018) discuss in their paper several variants of how refreshing could occur after an interruption. For the sake of clarity, we only presented in our introduction the prediction of the simplest variant (similar to the cumulative schedule of refreshing in forward order also implemented by Oberauer and Lewandowsky (2011)). In this variant, items are refreshed in their order of presentation, that is,

always starting with the first item after an interruption. Other variants of the TBRS2 model can account for other refreshing schedules thanks to which the model is able to keep track of the last refreshed item or the least activated item but our tentative simulations did not lead to satisfactory results as the model in its present form does not enable to restrict the parameter search. However, theoretically, in a complex span task such as the operation span task, the 'last refreshed item' variant would lead to a strong recency effect because refreshing would favor the last items of a list. In contrast, the 'least activated item' variant would predict no clear serial position effects, as all items should be somewhat equally activated. As seen in Fig. 3, except at very short list length, the data indicates a strong primacy effect and no recency effect. For this reason, it is unlikely that these variants could account any better for the data observed.

Another parameter of the TBRS2 concerns how long items are refreshed in working memory. Our predictions were based on a fixed refreshing duration (e.g., every item is refreshed for 0.08s., also implemented by Oberauer and Lewandowsky (2011)). However, the model has an alternative option, where refreshing goes on until a threshold of activation is met for a given item (or after a minimum duration if the threshold is already met for a hypothetical already highly activated item, for instance 0.01s.). This variant leads to greater variability in the final product of recall compared to a fixed refreshing duration, especially if the number of items to remember is higher than the individual span. However, although such a variant is even more hazardous when simulated, we argue that it is not psychologically plausible, as it would require an extra component allowing the participant to constantly scan the activation level of the items to decide whether the refreshing process must be pursued for a given item.

We presented here four optional versions of the TBRS2, which appear unlikely to better account for the observed patterns of recall than the default version. However, many other modeling choices could have been made regarding the refreshing schedule, including its total duration or a limited number of items being refreshed at once (see Lemaire, Pageot, Plancher, and Portrat (2018) for various computational implementations of refreshing schedules of the TBRS), but also the exact choices of the decay function. While it is likely that several other modeling choices combined could reproduce the data observed, again, finding the optimal set of parameters seems premature. It is also complicated to describe verbally exactly how these implementations would behave without conducting a thorough computational modeling study, as all these parameters interact with the design of the task such as its pace and list length (Lemaire et al., 2018). The fact that refreshing and decay processes are underspecified is an inherent limitation of verbal models in general. In contrast, computational modeling require that all parameters of a model are to be considered and well defined. This inevitably leads to a family of models that eventually predict different effects (Lewandowsky & Farrell, 2010), but discussing the predictions of every variant of the TBRS is beyond the scope of this article.

Concerning the prediction of higher-order abilities, the structure of the complex span task did not influence the predictive power of the tasks. The mediation analysis helped partition direct and indirect effect of the types of tasks, but it remains difficult to interpret. Nevertheless, this result appears reassuring given all of the studies which have been conducted so far to estimate the relation between working memory capacity and general intelligence.

A potential concern with our correlational analysis is a ceiling effect, as the lists were limited to seven items. However, only five participants obtained a span of seven based on the partial unit scoring system. We reran the analysis using a more stringent scoring method (i.e., absolute scoring) and we did obtain reduced performance and reduced skewness, but the overall pattern of statistical results did not change. Therefore, similar conclusions could be drawn. Our result could be limited in that the higher-order abilities were assessed by a single subtest, and because the two experimental conditions were limited to 15 trials (following Unsworth et al. 2005). Further tests should attempt to generalize our findings by using several measures of fluid intelligence and using a greater number of trials per condition to reach a more precise estimate of the individual's spans. Additionally, such a project would ideally involve a larger number of complex span tasks using more diverse material and different types of concurrent tasks.

As seen, the results revealed several interesting findings. First, the structure of the task has an influence on the *estimate* of working memory capacity. When the processing episode is at the beginning of the task, the estimate of the span is increased. This finding is in line with the TBRS model prediction, as the first processing event does not have an impact on the cognitive load of the task. Our second result is that the prediction of intelligence was not affected by the structure of the task when the tasks were compared separately (i.e., the correlations between the types of tasks and intelligence were comparable). However, a mediation analysis can reveal a relation between the types of tasks as the Processing-Storage type of task appears to have a lower direct effect on intelligence than the storage-processing type, but this relation is not easy to interpret theoretically. Further experiments are needed to better determine this apparent complex relation between the two types of tasks and fluid intelligence.

References

Aben, B., Stapert, S., & Blokland, A. (2012). About the distinction between working memory and short-term memory. Frontiers in Psychology, 3(301) pp. 1–9.

Bailey, H. (2012). Computer-paced versus experimenter-paced working memory span tasks: are they equally reliable and valid? *Learning and Individual Differences*, *22*(6), 875–881.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83–100.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental psychology: Learning, Memory, and Cognition*, *33*, 570–585.

Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language*, *45*, 1–20.

Barrouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory . In *The Cognitive Neuroscience of Working Memory*, (pp. 455,59–80).

Barrouillet, P., & Camos, V. (2012). As time goes by: Temporal constraints in working memory. *Current Directions in Psychological Science*, *21*, 413–419.

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: a time-based resource-sharing model account. *Developmental Psychology*, *45*, 477–490.

Barrouillet, P., Lepine, R., & Camos, V. (2008). Is the influence of working memory capacity on high-level cognition mediated by complexity or resource-dependent elementary processes? *Psychonomic Bulletin & Review*, *15*, 528–534.

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. Psychological Review, pp. 175–192.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Cohen, J. (1994). The earth is round (p<. 05). *American Psychologist*, *49*(12), 997.

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: a reanalysis of key studies. *Memory & Cognition*, *34*, 158–171.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: a methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*, 547–552.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*(3), 274–290.

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: a comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology*, *60*(9), 1227–1245.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.

Engle, R. W. (2018). Working memory and executive attention: a revisit. *Perspectives on Psychological Science*, *13*(2), 190–193.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, *37*, 581–590.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, *116*(2), 439.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. Trends in Cognitive Sciences(11), pp 410–419.

Gauvrit, N., & Mathy, F. (2018). Mathematical transcription of the "time-based resource sharing" theory of working memory. *British Journal of Mathematical and Statistical Psychology*, *71*, 146–166.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 66–71.

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*, 189–217.

Kanerva, K., & Kalakoski, V. (2016). The predictive utility of a working memory span task depends on processing demand and the cognitive task: the predictive utility of WM span task. *Applied Cognitive Psychology*, *30*, 681–690.

Lemaire, B., Pageot, A., Plancher, G., & Portrat, S. (2018). What is the time course of working memory attentional refreshing? *Psychonomic Bulletin & Review*, *25*(1), 370–385.

Lemaire, B., & Portrat, S. (2018). A computational model of working memory integrating time-based decay and interference. *Frontiers in Psychology*, *9*, 416.

Lewandowsky, S., & Farrell, S. (2010). Computational modeling in cognition: Principles and practice. SAGE publications.

Lewandowsky, S., Oberauer, K., Yang, L. X., & Ecker, U. K. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, *42*, 571–585.

Loaiza, V. M., & Camos, V. (2016). Does controlling for temporal parameters change the levels-of-processing effect in working memory? *Advances in Cognitive Psychology*, *12*(1), 2.

Lucidi, A., Loaiza, V., Camos, V., & Barrouillet, P. (2014). Assessing working memory capacity through time constrained elementary activities. *The Journal of General Psychology*, *141*, 98–112.

Mathy, F., et al. (2018). Simple and Complex Working Memory Tasks Allow Similar Benefits of Information Compression. *Journal of Cognition* 1(1): 31, pp.1–12.

McCabe, D. P. (2010). The influence of complex working memory span task administration methods on prediction of higher level cognition and metacognitive control of response times. *Memory & Cognition*, *38*(7), 868–882.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*, 49–00.

Morey, R. D., Romeijn, J.-W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, *72*, 6–18.

Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition*, *37*, 346–357.

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: a computational implementation of the time-based resource-sharing theory. *Psychonomic Bulletin & Review*, *18*, 10–45.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D., Conway, A., Cowan, N., et al. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: an interference model of complex span. *Psychonomic Bulletin & Review*, *19*, 779–819.

Raven, J. C. (1962). Advanced progressive matrices: Sets I and II. HK Lewis.

Ricker, T. J., Vergauwe, E., Hinrichs, G. A., Blume, C. L., & Cowan, N. (2015). No recovery of memory when cognitive load is decreased. *Journal of Experimental psychology: Learning, Memory, and Cognition*, *41*(3), 872.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory:, the gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1089.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612.

Stone, J. M., & Towse, J. (2015). A working memory test battery: Java-based collection of seven working memory tasks. *Journal of Open Research Software*, pp 3.

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, *54*, 68–80.

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104.

Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*, 1038–1066.

Unsworth, N., Heitz, R. P., Schrock, J., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: a latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635–654.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*, 392—406.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., et al. (2018). Bayesian inference for psychology. part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57.

Wang, T., Ren, X., Li, X., & Schweizer, K. (2015). The modeling of temporary storage and its effect on fluid intelligence: Evidence from both Brown–Peterson and complex span tasks. *Intelligence*, *49*, 84–93.

Wechsler, D. (2011). *Wais-iv Nouvelle Version de L'échelle d'intelligence de Wechsler Pour Adultes*. Montreuil: ECPA Pearson.