



The domain-specificity of serial order working memory

Yingxue Tian¹ · Margaret E. Beier¹ · Simon Fischer-Baum¹

Accepted: 12 November 2021 / Published online: 27 December 2021
© The Psychonomic Society, Inc. 2021

Abstract

Making a turn while driving is simple: turn on the indicator, check for cars, then turn. Two types of information are required to perform this sequence of events: information about the items (e.g., the correct indicator), and the serial order of those items (e.g., checking before turning rather than vice-versa). Previous research has found distinct working memory capacities (WMCs) for item and serial order information in both verbal and nonverbal domains. The current study investigates whether the serial order WMC is shared for sequences from different content domains. One hundred and fifty-three participants performed sequence matching tasks with verbal (letters and words) and nonverbal (locations and arrows) stimuli. The accuracy of detecting mismatched item-identity and serial order information in sequences was used to operationalize item and order WMC. Using structural equation modeling analyses, we directly compared models that included either domain-specific or domain-general serial order WMC latent variables, finding that models with domain-specific serial order WMC latent variables for verbal and nonverbal materials fit the data better than models with domain-general latent variables. The findings support the hypothesis that there are separate capacities for serial order working memory depending on the type of material being ordered.

Keywords Serial order · Working memory · Domain-specificity · Individual differences · Structural equation modeling

Daily life is filled with activities that require holding sequences of items or events in working memory (WM). When we get someone's phone number or learn a new dance, we are faced with two separate challenges—the ability to remember the identity of the items (e.g., digits or dance moves) and the ability to remember the order in which those items occur. There is longstanding evidence that different sequence types (verbal versus nonverbal/visuospatial) rely on different WM systems (Hanley et al., 1991; Shah & Miyake, 1996; Vallar & Baddeley, 1984). Furthermore, there is growing evidence for a separation of item-identity and serial order working memory capacities (WMCs) for both sequences of verbal (for review, see Majerus, 2019) and nonverbal materials (Amiez & Petrides, 2007; Avons & Mason, 1999; Claessen et al., 2016; Hsieh et al., 2011; Saint-Aubin et al., 2007; Wansard et al., 2015).

The current study investigates the intersection of the verbal-nonverbal and the item–order dissociations within WM. We consider two hypotheses about the structure of WM, as depicted in Fig. 1. The domain-general serial order hypothesis (Fig. 1a) assumes that the classic verbal–nonverbal dissociation reflects a dissociation in item-identity WMC, but not in serial order WMC. The domain-specific serial order hypothesis (Fig. 1b) assumes that the verbal–nonverbal dissociation extends to both item-identity WMC and serial order WMC.

Previous research has been equivocal. Some support for the domain-general hypothesis comes from the similarity of behavioral effects in serially recalling verbal and nonverbal sequences—for example, similar serial position curves are observed regardless of item content (Hurlstone et al., 2014). However, there are also striking differences across domains (Gmeindl et al., 2011)—for example, the accuracy benefits of free recall compared with serial recall are greater in the nonverbal than the verbal domain. Dual-task studies have similarly mixed results. Some researchers have observed interference for embedded or concurrent order processing tasks both within-domain (e.g., an embedded digit task with a primary letter task) and cross-domain

✉ Yingxue Tian
yingxue.tian@rice.edu

✉ Margaret E. Beier

✉ Simon Fischer-Baum

¹ Department of Psychological Sciences, Rice University, Houston, TX 77005, USA

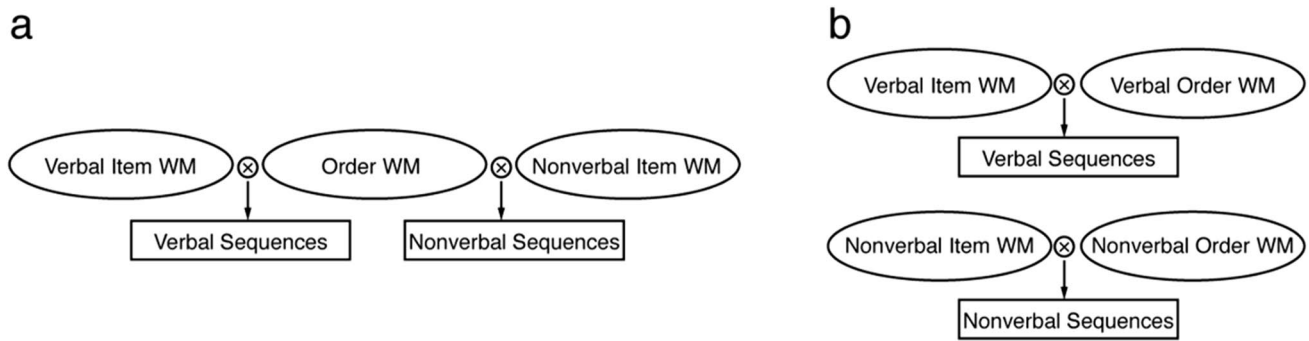


Fig. 1 An illustration of the domain-general (a) and the domain-specific (b) serial order hypothesis. Ovals reflect the different types of representations needed to support working memory, and \otimes is the

binding machinery that allows us to bind representations of items in a specific serial order. Arrows point to the output sequences after the binding of representations

(e.g., an embedded spatial task with a primary letter task; Depoorter & Vandierendonck, 2009; Jones et al., 1995; Vandierendonck, 2016), as would be predicted by the domain-general hypothesis; while others have reported negligible or asymmetric cross-domain interference that supports the domain-specific hypothesis (Morey & Mall, 2012; Soemer & Saito, 2016). Similarly, while some neuroimaging studies have found common neural substrates for order processing with verbal and nonverbal sequences (e.g., Majerus et al., 2010), those studies also identify regions that are selective to order processing in only one domain. Finally, while neuropsychological double dissociations have been reported with some patients showing difficulties on serial order processing tasks in a domain-selective manner (Kesner et al., 1994; McMackin et al., 1995), it is not clear whether the domain-wise dissociations of serial order WMCs in these reported patients was driven by differential demands to a domain-general serial order WMC, or by the spillover effects from differential impairments in domain-specific item-identity WMCs. The question of whether or not there is a domain-general serial order WMC has received a great deal of attention in the literature, but previous methodological approaches have yet to provide clear evidence.

In the current study, we use a latent-variable approach to test competing hypotheses about the relationship between serial order capacities for sequences of different content domains. The latent-variable approach has provided a powerful tool in understanding the relationship between interplaying cognitive constructs, like the structure of WMC constructs (Conway et al., 2003). Dissociations in domain-general studies are hard to interpret because they are usually confounded with task type and task difficulty. Instead of designing distinct but similar tasks for item–order or verbal–nonverbal contents, the latent-variable approach draws inferences about the underlying cognitive construct from the individual differences within the same set of data; we can explore the optimal cognitive architecture amongst

competing theoretical hypotheses by loading the observed data on different structural models in structural equation modeling (SEM) analyses and directly pitting nested models against each other. Another advantage of the latent-variable approach is the locus of its inference being intraindividual. Compared with the mean-based analyses, the latent-variable approach is not affected by the average performance in different conditions; instead, it assesses the structure of latent constructs based on how corresponding manifest variables covary. The challenge of balancing task difficulties in different domains or conditions could then be circumvented. Using the latent-variable approach, we asked two questions. First, can we replicate the dissociation between item-identity and serial order WMC¹ with this approach? Second, if we find an item–order dissociation in both domains, is serial order WMC domain-general or domain-specific? Three primary theoretical models were tested: (1) a two-factor model representing WMC only distinguishing verbal and nonverbal² capacities, (2) a four-factor model representing dissociated item and serial order WMCs in both the verbal and nonverbal domains, and (3) a three-factor model distinguishing verbal and nonverbal item-identity WMCs, but

¹ The distinction between WM and short-term memory (STM) is one of the major theoretical discussions in the field, and clearly identifying the boundary between these memory systems has been elusive. However, some researchers argue that these systems are distinct from each other. With this ontology in mind, it is worth noting that the tasks used in the current study—immediate recognition of whether two six-item sequences are identical or not—would be considered more traditional STM tasks than WM tasks.

² The current study aims to assess whether the same WMC is applied to support serial order information for verbal materials compared with other materials. Given that the nonverbal materials we used have differential spatial nature, we used the term nonverbal to avoid the implication of the spatial involvement in both material types. However, because of the vast set of nonverbal materials, we narrowed down the scope of our finding to visuospatial in the Discussion.

treating the serial order WMC as domain-general. In addition, based on the results of the study, we analyzed another model: (4) a three-factor model derived from the four-factor model but treating the item-identity and serial order WMCs as indistinguishable in the verbal domain. Inferences about WMC structures were drawn by comparing the fit of nested models while considering the model complexity. The first question—whether item and order WMCs dissociate—was tested with a comparison of the two-factor and four-factor model, with a better fit for the four-factor model supporting a dissociation between item and order WMCs. The second question—how serial order WMC relates across domains—was tested with a comparison between the four-factor (domain-specific) and the three-factor (domain-general) model. A reliably better fit of the four-factor than the three-factor model would provide clear evidence for the domain-specific serial order WMC hypothesis; otherwise, the result would support the domain-general hypothesis.

Method

Participants

We recruited 161 undergraduate students from Rice University. For the inclusion criteria, participants were required to (1) be at least 18 years old, (2) have normal or corrected-to-normal vision, (3) have no neural abnormalities, and (4) have no attentional disorders. All participants received extra course credit or \$25 for their participation. This study was approved by the Rice University Institutional Review Board.

The sample size was determined jointly by common rules-of-thumb (Anderson & Gerbing, 1988; Arrindell & van der Ende, 1985; Cattell, 1978; Ding et al., 1995) and by previous Monte Carlo simulation studies (Wolf et al., 2013). Rules-of-thumb afforded several prospective sample size estimates. The first rule-of-thumb was the absolute values, which suggested that a sample of 100 (Ding et al., 1995) or 150 (Anderson & Gerbing, 1988) could be sufficient for SEMs. The second rule-of-thumb was the $N:q$ ratio, where N is the expected sample size and q is the number of parameters being estimated. The current study had q ranged from 49 to 54 depending on the model complexity. Cattell (1978) suggested that the $N:q$ ratio larger than 3 would be sufficient, and thus a sufficient sample size would be within the range of 147 to 162. The third rule-of-thumb was based on the number of latent variables. Arrindell and van der Ende (1985) suggested a sufficient sample size being approximately twenty times of the number of factors. The largest model in the current study had eight factors, and thus a sufficient sample size would be 160. In a separate vein, Monte Carlo studies have been considered as well. Based on the findings in Wolf et al. (2013), when middle to strong factor

loadings and the indicator to latent variable ratio above 6 (4 or 8 in the current study) were expected, a sufficient sample size was less than 150. Taken together, we determined the expected sample size of 150.

Eight participants were removed from the analysis because they did not finish the experiment or did not respond with valid answers. Thus, the final sample consisted of 153 participants.

Procedure

The consent form and demographic questionnaire (including age, gender, education level, and ethnicity) were acquired for all participants upon arrival. Participants were seated individually in the test room and were tested for approximately 2.5 hours. All tasks were computerized with E-Prime software (E-Prime 2.0, Psychology Software Tools, Inc., Pittsburgh, PA, USA) and administered with Dell OptiPlex 9010 touchscreen PC (23-inch) at $1,920 \times 1,080$ resolution. After being instructed with the task description, participants completed four sequence matching tasks in the preassigned order corresponding to their participation entries to the study. The order of four tasks was counterbalanced across participants.

Materials

Four stimulus types were adopted to get converging evidence from two stimulus types in each domain. To form a sequence, six items were selected in a randomized order without repeated items in a sequence.

Letter Twenty consonants (excluding *Y*) were the stimulus pool for one of the verbal materials. Letters were displayed in capital cases in bold Courier New font at the 18-point size in the center of the screen. Only consonants were included to avoid the formation of pronounceable segments (consonant–vowel–consonant structures) and the corresponding reduction of memory loads.

Word Twenty-five nouns were the stimulus pool for one of the verbal materials. Twenty-five nonrhyming nouns were selected from 95 monosyllabic, frequent (Kucera-Francis written frequency ranging from 10 to 30, average = 18.92), and imageable (Imageability rating above 350, average = 528.64) five-letter nouns. The word stimuli can be found in Appendix 1. The presentation format was identical to letters. Frequent, imageable words were used in the study to avoid the performance being biased by vocabulary knowledge.

Spatial location Ten spatial locations in the form of white squares with black borders were the stimulus pool for one of the nonverbal materials. The squares were at the size of 45×45 pixels with the border width of 1 pixel. Locations

were selected as a randomly distributed pattern on the screen with no clear verbal labels (e.g., exact position, collinearity, closure) available. The location stimuli can be found in Appendix 2.

Arrow Sixteen arrows characterized by two features (length and direction) were the stimulus pool for one of the nonverbal materials (Kane et al., 2004). Eight directions were used, pointing from the central fixation to each corner and edge midpoints of the screen. For each direction, two lengths (a quarter and half of the width of the screen) were included. The arrow stimuli can be found in Appendix 3.

Tasks

Four sequence matching tasks, each with a different stimulus type as described above, were administered. At the beginning of each task, instructions were given in the written form on computers. Participants were informed of the stimulus type of the task and the length of each sequence, and they were instructed to judge if two sequences were identical in each trial. After four practice trials to familiarize participants with the task requirements and the response equipment, 120 trials were tested in a fixed order in order to limit between-subject variance associated with the order of tasks (Goodhew & Edwards, 2019; Mollon et al., 2017). The fixed order of trials also assured that the potential proactive interference, if any, from previous trials would be consistent across participants. Each trial consisted of the presentation of two consecutive six-item sequences and a response page. Each sequence was presented after a central fixation cross (200 ms) with the speed of 600 ms per item and a 200 ms interstimulus interval (ISI). In the sequence of letters, words, and arrows, each item was presented in the center of the screen then disappeared, followed by a blank screen during ISI. In the sequence of visuospatial locations, the default background with ten static locations was presented throughout the entire presentation of a sequence, superimposed by one square turning to solid black for 600 ms then returning to the default state as the index of one item. Following two sequences, a response page was presented at the end of each trial, with the question “Are these two sequences identical?” and two response boxes (“Yes” and “No”). Participants responded by touching one of the two response boxes on the screen.

There were three trial types for each task based on the congruency of two sequences: identical, item-changing, and order-swapping (see also Gathercole et al., 2001, who refer to the combination of identical and order-swapping trials as the serial recognition task). In each trial, the first sequence was generated from the stimulus pool with positions associated with each item being counterbalanced. The second sequence was the variation of the first sequence according

to the trial type. An identical trial was when the second sequence was identical to the first sequence (Fig. 2a); the corresponding correct answer would be “Yes.” An item-changing trial was when the second sequence had five items in the same positions as the first sequence and one item was replaced by an extra-list item from the stimulus pool (Fig. 2b); the corresponding correct answer was “No.” The within-sequence position of the changed item was counterbalanced across trials. An order-swapping trial was when the second sequence contained all identical items as the first sequence, but two adjacent items swapped their positions (Fig. 2c); the corresponding correct answer was “No.” The within-sequence positions of the swapped pair of items were counterbalanced across trials. Trials with “Yes” and “No” as correct responses were of equal amount to avoid objective response bias (i.e., 60 identical trials, 30 item-changing trials, and 30 order-swapping trials for each task). Three trial types were preassigned in a randomized order for each task, and all participants received the same order of trials in each task. The randomized order of three trial types within each task allows participants to freely form their strategies without being specified to item or order, and in turn prevents any differentiated performance being induced by dissociations embedded in the task design.

Outlier screening

Accuracy and reaction time in four tasks (collapsed across three trial types) were calculated for the outlier screening. We defined outliers at the subject level, and two task-level criteria were considered collectively for each task to screen the subject-level outliers.

The “low” task-level outlier was defined as follows. For each participant, any task with an accuracy below 57.5% was defined as a univariate “low” task-level outlier; this chance level was determined as the accuracy of guessing performance on 120 trials regardless of trial type (i.e., with a binomial distribution of $N = 120$ and $p = 0.5$ at $\alpha = .05$). There were 41 participants with one or more “low” task accuracy values.³ These low values could either be caused by hasty guessing or struggles with the task (i.e., being at the lower end of the distribution).

The “fast” task-level outlier was defined as follows. For each participant, any task with log-transformed RT (logRT) below the lower hinge of the logRT distribution across all 153 participants was defined as a univariate “fast” task-level outlier. The lower hinge of the logRT distribution was defined as $Q_1 - 3 \times IQR$, where Q_1 was the first quantile of

³ The model comparison results did not change even when low performers were excluded, suggesting that the results are robust regardless of what data trimming method was applied.

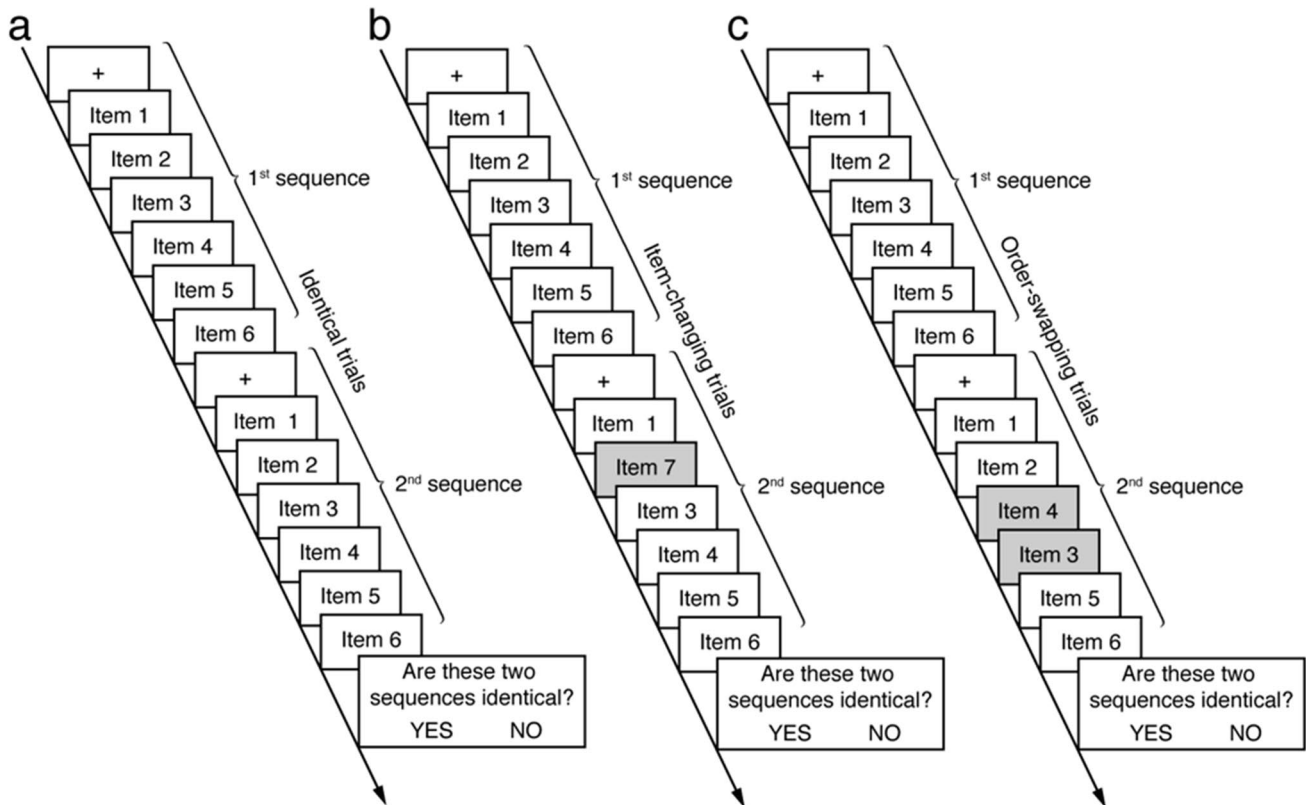


Fig. 2 Task schematics for (a) identical trial, (b) item-changing trial, and (c) order-swapping trial

the logRT distribution and IQR is the third quantile of the distribution (Q_3) subtracted by Q_1 . There was one participant exhibiting such “fast” task logRT value.

A participant outlier was defined when a participant had at least one task being both a “fast” and “low” task-level outlier. No participant met this conjunction criterion.

Note that we used this conjunction of criteria to detect near-chance performance made by participants who were merely guessing, instead of simply removing the effortful performance of individuals at the lower end of the distribution. Each task-level outlier criterion was based on the performance for a task/stimulus type in lieu of each trial type to ensure that any anomalous performance level observed in specific types of mismatched trials (e.g., item-changing or order-swapping trials) in the subsequent SEM analysis is not driven by the general anomalous performance for a specific task.

Data scoring

Accuracy measures⁴ were used as manifest variables for SEM. For each task, 30 item-changing trials and 30

order-swapping trials were aggregated into four parcels for the subsequent SEMs. Considering each trial as an item for SEM, we adopted a subset-item parcel approach to create parcels as manifest variables.

For each trial type, we split trials in alternation and assigned 15 trials to each parcel to take the advantage of parceling without manipulating the data too much. This way, for each stimulus type, the accuracy for 15 item-changing or order-swapping trials was calculated as a manifest variable of “item” or “order.” We yielded 16 parcels as manifest variables for SEM analyses, including four manifest variables (two “item” and two “order”) for each stimulus type.

For clarity and consistency with the tradition of experimental research, we reported descriptive statistics and inter-correlations based on eight all-item parcels (i.e., accuracy

⁴ An alternative approach, using d' discriminability rather than accuracy as manifest variables, is reported in the supplementary material. To address issues of independence between manifest variables, false alarm rates used for the calculation of d' s for item-changing and order-swapping trials were the proportion of incorrect trials from two independent parcels selected in alternation from identical trials, respectively. The choice of manifest variable had no impact on our model selection.

measures aggregated from 30 item-changing and order-swapping trials for each task, respectively).

Note that the reasons we used the parcel-based instead of the item-based approach were twofold. One was that the goal of the current study was to investigate the underlying structure of constructs rather than item-level properties. Two was that, compared with the item-based approach, parceling reduces model complexity, reach optimal reliability, and avoid violation of normality assumptions (Little et al., 2002; Yang et al., 2010).

Within the parcel-based approach, we chose the subset-item parcel approach over the all-item parcel approach. Compared with the subset-item parcel approach, the all-item parcel/scale approach (i.e., aggregating all 30 trials—the entire scale—into one parcel) tends to yield a higher model fit but also is prone to inflate the estimated path coefficients (Bandalos, 2002). Scale-based models are susceptible to measurement-error-related bias even in the absence of other misspecifications, have low power to detect measurement model misspecification, and their fit indices cannot reflect model misspecification (Coffman & MacCallum, 2005; Cole & Preacher, 2014; Rhemtulla, 2016; Yang et al., 2010). Hence, having a set of parcels provides a better solution over a scale-based model.

Regarding the size of the subset-item parcel (i.e., how many trials to include in a parcel), although having fewer parcels (i.e., including more trials in each parcel) tends to increase the model fit (Bandalos, 2002), the extent of the change is minor as long as the number of parcels per factor is not excessive (Little et al., 2002; Marsh et al., 1998; Matsunaga, 2008; Rogers & Schmitt, 2004). Considering the balance between model fit improvement and coefficients estimation inflation, as well as the requirement of having at least three indicators for each latent factor in SEM (Little et al., 2013; Raubenheimer, 2004), we created two parcels for each trial type in each task, yielding at least four manifest variables for each latent factor.

Note that there are various schemes to create parcels (e.g., based on item properties, based on previous research, empirical approaches, systematically creating parcels in an isolated or distributed manner). Randomly assigning items to parcels could create bias favoring items with larger variances (Matsunaga, 2008), and using empirical approaches (e.g., factor analysis or correlational analysis between trials) to assign parcels might introduce unknown bias to the data structure. Therefore, we systematically created parcels for each trial type by odd–even splitting of items in a distributed way regardless of item content. To assure that our results do not depend on parcel allocation schemes, we used a bootstrapping procedure to randomly split each trial type into two parcels in 1,000 simulation runs. Our results regarding the

descriptive statistics, model fit indices, and model comparisons were robust to the parceling variability (see Table S7 and Table S8 in the Supplementary Materials). Our results do not depend on choosing the subset-item or all-item parcel approach either. We have conducted the SEM analysis with the scaled-based manifest variables, and the model comparison results did not differ between the scale-based and the parcel-based manifest variables (see Fig. S3 and Table S9 in the Supplementary Materials).

Statistical analysis

We report several fit indices for structural equation models. The absolute fit indicates how well the model reproduces the covariance matrix of the observed performance: chi-square (χ^2), relative chi-square ($\chi^2:df$), root-mean-square of error of approximation (RMSEA), and standardized root-mean-square residual (SRMR). A nonsignificant chi-square suggests a well-fitted model, indicating that there is no identifiable discrepancy between the model-implied and the population covariance matrices (M. W. Browne & Cudeck, 1992). To control for the possible inflated chi-square in complex models, the relative chi-square takes the number of free parameters into consideration and any value less than 2 reflects an acceptable fit (Hooper et al., 2008). RMSEA is calculated based on the residual matrix with a penalty for model complexity; SRMR is the standardized difference between the observed and model-implied correlation matrices with no penalty for model complexity. RMSEA or SRMR less than .08 suggests a good fit (Hu & Bentler, 1999). Additionally, we report the comparative fit indices to reflect the improvement in the fit of the model over the null model: normed fit index (NFI), nonnormed fit index (NNFI; also as Tucker-Lewis index [TLI]), and comparative fit index (CFI). Any comparative fit index larger than .90 indicates a good fit (Hooper et al., 2008). Moreover, we report the parsimonious fit indices to reflect the tradeoff between improvement in fit and loss of degrees of freedom: Akaike's information criterion (AIC) and Schwarz's Bayesian information criterion (BIC).

The model comparison was implemented by the chi-square difference test between two nested models. If the $\chi^2_{\text{difference}}$ is significant, the increase in the fit index (namely, the drop in the χ^2) by adding latent variable(s) is “worth” the loss of the degree(s) of freedom. Hence, if the comparison test was significant, the model with more latent variables is better at capturing the structure of WM constructs than the model with fewer latent variables. Moreover, we compare AIC measures between nested models, with the AICs difference less than 4 supporting the model with larger AIC (Burnham & Anderson, 2004).

Table 1 Descriptive statistics and reliability estimates for working memory accuracy measures

Variable	Mean	SD	Range	Skewness	Kurtosis	Reliability
Letter						
Item	.87	.12	[.43, 1]	−1.27	1.15	.76
Order	.84	.13	[.40, 1]	−0.87	0.20	.73
Word						
Item	.76	.15	[.40, 1]	−0.50	−0.51	.73
Order	.76	.17	[.33, 1]	−0.56	−0.47	.81
Location						
Item	.86	.14	[.30, 1]	−1.39	1.53	.80
Order	.73	.17	[.27, 1]	−0.40	−0.71	.79
Arrow						
Item	.72	.16	[.20, 1]	−0.56	−0.07	.73
Order	.63	.16	[.17, .97]	−0.24	−0.44	.70

Note. $N = 153$ for all tasks

Statement of transparency

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (Simmons et al., 2012).

Results

Demographic data for 153 participants were as follows: 92 females, 60 males, and one decline to report binary gender. The mean age was 20 years ($SD = 1.3$ years). Education levels ranged from freshman to postbaccalaureate in college.

Average reaction times (\pm standard deviation) across three trial types were 817 (± 336), 893 (± 384), 749 (± 341), 966 (± 448) ms for letter, word, location, and arrow tasks, respectively. Average accuracies (\pm standard deviation) across three trial types were 82% ($\pm 11\%$), 77% ($\pm 11\%$), 76% ($\pm 11\%$), 67% ($\pm 10\%$) for letter, word, location, and arrow tasks, respectively. Specifically, for identical trials, average accuracies (\pm standard deviation) were 78% ($\pm 19\%$), 78% ($\pm 14\%$), 71% ($\pm 15\%$), 68% ($\pm 16\%$) for letter, word, location, and arrow tasks, respectively. For a comprehensive view of the descriptive statistics and intercorrelations of the entire accuracy dataset (i.e., accuracy measures for all three trial types in four tasks), Table S4 and Table S5 are reported in the Supplementary Materials. The descriptive statistics of the log-transformed reaction time for all three trial types in four tasks are reported in Table S6 in the Supplementary Materials as well.

Eight manifest variables were used for descriptive statistics and intercorrelations. The descriptive statistics for the eight manifest variables are displayed in Table 1. Split-half reliability was derived for each manifest variable as an index of internal consistency. For each manifest variable,

corresponding trials were split into odd and even halves, and the Spearman correlation was calculated between the proportion-correct scores in two halves, then adjusted by the Spearman–Brown Prophecy formula. All manifest variables had adequate reliability (at or above .70). The skewness was less than 2 and the kurtosis was less than 7 in all measures, indicating that manifest variables were approximately univariate normally distributed (Ryu, 2011).

Note that there is no strict standard for the acceptable range of reliability, although the heuristics consider a reliability measure above .60 or .70 acceptable. In empirical works, this heuristic is applied more leniently; usually, around .50 is considered moderate as well (e.g., Alloway et al., 2004; Unsworth & McMillan, 2013). Reliability being too high ($>.95$) or too low is not ideal and it could reflect redundancy (Hulin et al., 2001) or random error in the measurement, correspondingly. For extreme values of reliability, they need to be interpreted with caution in the context of the measurement (e.g., how measurements were derived, the zero-order correlation of “unreliable” measure to other measures, how measurements are used for the research goal; see also Little et al., 2013, for a more in-depth discussion). Just like the distinction between significance and nonsignificance in null hypothesis significant tests has been criticized, simply using a cutoff to assess the reliability of measurements is not helpful either.

Correlations among eight manifest variables were uniformly positive ($.25 < r_s < .74$; see Table 2).⁵ The magnitudes of correlations were suited for the latent-variable analysis, given that they were not too high to reflect redundant indicators nor

⁵ The descriptive statistics and correlation table for 16 manifest variables is reported in Table S2 and Table S3 in the Supplementary Materials.

Table 2 Correlation matrix for working memory accuracy measures

	Item letter	Order letter	Item word	Order word	Item location	Order location	Item arrow
Order letter	.74						
Item word	.48	.54					
Order word	.48	.55	.73				
Item location	.34	.28	.30	.30			
Order location	.26	.25	.25	.30	.68		
Item arrow	.37	.37	.40	.36	.53	.60	
Order arrow	.36	.37	.32	.34	.45	.56	.66

Note. $N = 153$ for all tasks. All correlations are significant at the $p < .01$ level

too low to be the mere result of a positive manifold. Moreover, intercorrelations revealed two aspects of construct validity: manifest variables were correlated moderately strongly within each domain ($.48 < r_s < .74$ in verbal domain and $.45 < r_s < .68$ in nonverbal domain), and weakly across domains ($.25 < r_s < .40$). Concomitant convergence in both domains and divergence across domains indicated adequate validity for manifest variables (Campbell & Fiske, 1959).

As shown in Table 1, the difference between item and serial order WM performance was smaller in the verbal domain compared with the nonverbal domain. Indeed, Gmeindl et al. (2011) reported this pattern with a similar design to the current study (Experiment 2; sequence matching task with digits and visuospatial locations), but with a more traditional group-average statistical approach: they found that item changes and order swaps were equally difficult to detect for verbal tasks; but for visuospatial tasks, participants performed significantly better when detecting item changes than when detecting order swaps. Our experiment conceptually replicated this finding: We observed a significant interaction on the group-average performance in the two-way (i.e., item–order by verbal–nonverbal) repeated-measures analysis of variance (ANOVA), with letter and word conditions combined as the verbal condition and location and arrow conditions combined as the nonverbal condition, $F(1, 152) = 87.88, p < .001, \omega^2 = .36$, with fewer correct detections for the nonverbal order swap trials (68.1%) than the other trial types (Verbal Item: 81.5%, Verbal Order: 79.5%, Nonverbal Item: 78.9%). From the group-average performance, we could not tease apart whether this domain divide in item–order discrepancy was driven by domain-specific serial order WMCs or by different difficulty–efficiency levels of adopting a domain-general serial order maintenance mechanism in two domains. However, the variation driven by the task difficulty only affects the mean of the performance distribution, but it does not pose many issues for a covariance-based analysis in the SEM approach. The SEM analysis draws inferences based on the relationship of how an individual’s performance differs from the mean performance in different cognitive processes;

when the task difficulty does not induce any floor or ceiling effect, the discrepancy between the individual and the average performance is not affected by the absolute value of the average level. If the discrepancies between the individual and average performance in manifest indicators tend to be comparable in two constructs across individuals, the SEM results would suggest that these two constructs reflect the same underlying cognitive process; similarly, systematic incomparable discrepancies between the individual and the mean performance across tasks would suggest dissociated cognitive processes. In addition to circumventing the confounds created by unbalanced task difficulties in a group-average based analysis, an SEM approach also affords the clarity of directly pitting two hypotheses against each other to elucidate the structure of latent constructs.

Is WMC a unitary construct in a sequence?

To examine whether our measures are sensitive enough to reveal the underlying constructs, the first analysis focused on a conceptual replication of the frequently reported dissociation between item and serial order WMCs (Amiez & Petrides, 2007; Avons & Mason, 1999; Claessen et al., 2016; Hsieh et al., 2011; Saint-Aubin et al., 2007; Wansard et al., 2015). This was done through a direct comparison between two nested competing models to determine the structure of the data. Specifically, two models were specified to determine how WMCs for item and serial order information in the verbal and nonverbal domains were related to one another. All structural equation models were operationalized using 16 manifest variables for reliable model estimation without correlated errors.

Model 1 (i.e., the two-factor model⁶ in Fig. 3a) tested the notion that WMCs are best conceptualized as a unitary

⁶ Four task-specific factors, each with loadings for all four manifest indicators in each stimulus type, were included in all models to represent task-specific variance induced by the blocked experimental setting and individual expertise for a certain stimulus type, and henceforth were not included into the count of factors for simplification. These four factors were not allowed to correlate with other factors.

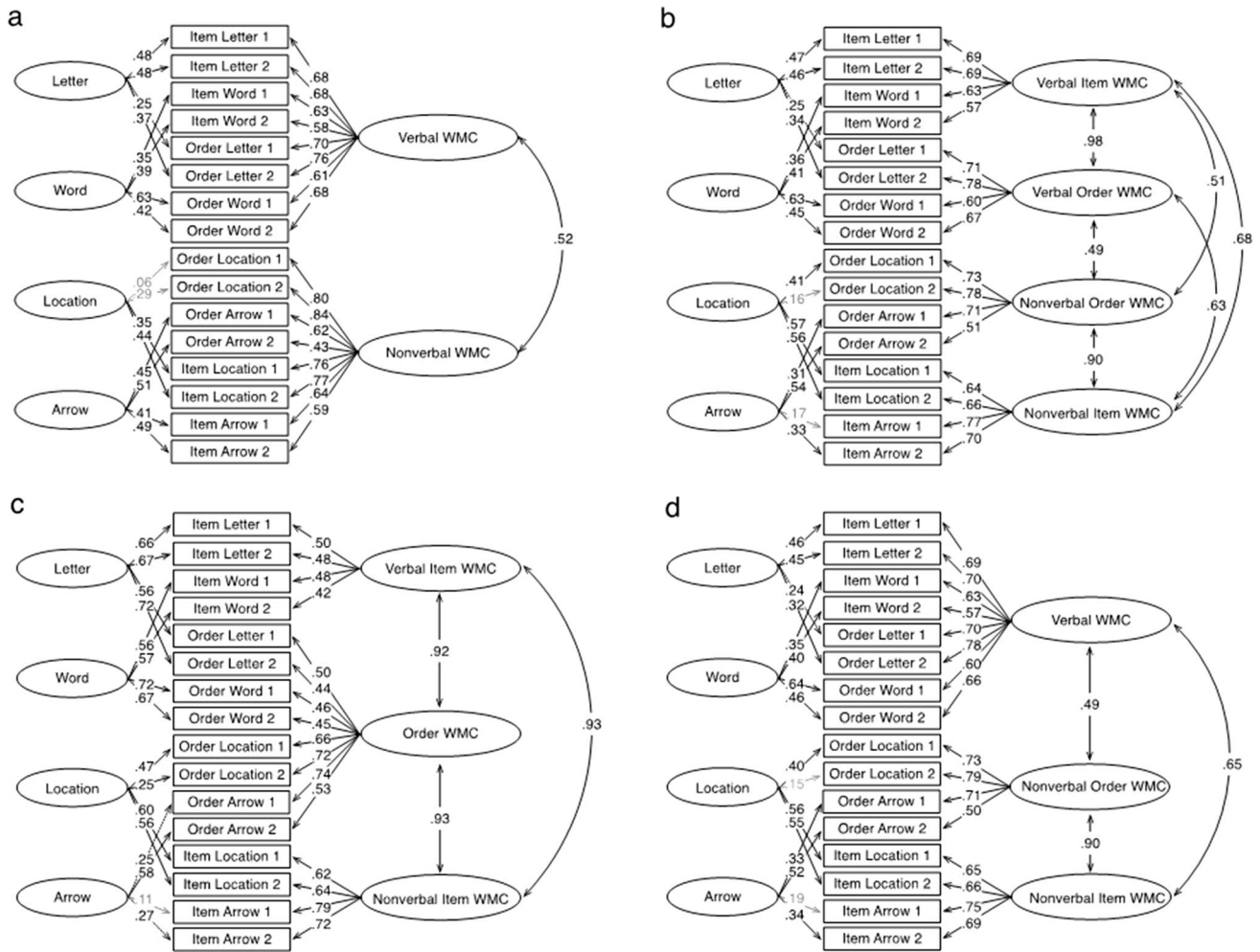


Fig. 3 Structural equation models consisting of (a) two latent factors (Model 1), (b) four latent factors (Model 2), (c) three latent factors (Model 3), and (d) another three latent factors (Model 4) for WMC in verbal and nonverbal domains. Numbers on the unidirectional arrows are standardized estimates of loadings for each manifest variable (rec-

tangle) onto its corresponding latent variable (oval). The numbers on the bidirectional arrow between latent variables represent the correlation between two latent variables. All black solid paths are with significance $p < .05$; black dashed path is with significance $.05 < p < .10$; gray paths are not significant ($p > .10$)

Table 3 Fit statistics for structural equation models

Model	χ^2	df	p	χ^2/df	RMSEA	SRMR	NFI	NNFI	CFI	AIC	BIC
1	101.02	87	.15	1.16	.03	.06	.93	.99	.99	-3117.41	-2969.24
2	84.72	82	.40	1.03	.02	.04	.94	.997	.998	-3125.81	-2960.52
3	133.79	85	.001	1.57	.06	.09	.90	.95	.96	-3082.56	-2928.33
4	85.31	85	.47	1.00	.005	.04	.94	1.00	1.00	-3130.31	-2976.09

Note. $N = 153$ for all models

construct in each domain. This two-factor model included a unified verbal factor with loadings for all verbal indicators, and a unified nonverbal factor with loadings for all nonverbal indicators. The two factors were allowed to correlate. Model 2 (i.e., the four-factor model in Fig. 3b) tested the notion that WMC was best captured as four dissociated yet

correlated factors representing item and serial order WMCs in both verbal and nonverbal domains. Thus, this four-factor model consisted of four WMC factors with separate loadings for item and serial order indicators in verbal and nonverbal domains. The four factors were allowed to correlate.

Table 3 shows the fit of the two aforementioned models. Model 1 and Model 2 both had good fit to the data. Critically, Model 2, which specified four WMC factors fit significantly better than Model 1, which specified two WMC factors, $\Delta\chi^2(5, N = 153) = 16.30, p = .006, \Delta AIC = 8.40$, suggesting a dissociation between item and serial order WMCs in both domains. Note that, although separated, the correlation between the item and serial order WMC factors was .98 (95% CI [.96, 1.00]) and .90 (95% CI [.85, .95]) in two domains in this model, indicating that item and serial order WMCs shared more than 81% of their variance. Taken together, the nested model comparison indicates that a parsimonious unified WMC is not sufficient to account for the capacity to remember sequences in both verbal and nonverbal domains; the structure of WM includes dissociations between item and serial order.

Is serial order WMC a unitary construct in different domains?

We next examined whether there are dissociated serial order WMCs for the verbal and nonverbal domains. We directly compared the fit of two competing models. Model 2 (i.e., the four-factor model in Fig. 3b) tested the notion that serial order WMC was best captured as two dissociated yet correlated factors representing domain-specific serial order WMCs in the verbal and nonverbal domains. Model 3 (i.e., the three-factor model in Fig. 3c) tested the notion that serial order WMC is best conceptualized as a single domain-general construct in the verbal and nonverbal domains. This three-factor Model 3 included a unified serial order WMC factor with loadings for eight serial order indicators in two domains, a verbal item factor derived from four verbal item indicators, and a nonverbal item factor derived from four nonverbal item indicators. The three factors were allowed to correlate.

As shown in Table 3, Model 2 and Model 3 both had good fit to the data. Critically, Model 2, which specified domain-specific serial order WMC factors, fit significantly better than Model 3, which specified a domain-general serial order WMC factor, $\Delta\chi^2(3, N = 153) = 49.07, p < .001, \Delta AIC = 43.25$, indicating a dissociation between verbal and nonverbal serial order WMCs. Note that, the correlation between serial order WMC factors in the two domains was .49 (95% CI [.41, .57]) in Model 3, indicating that the serial order WMCs shared around 24% of their variance across domain and that relatively large variances were of domain-specific origins. Moreover, the correlations between item and cross-domain serial order WMCs were .63 (95% CI [.55, .71]) and .51 (95% CI [.43, .59]), indicating that the interaction of item and cross-domain serial order WMC can account for less than 40% of the variance. This suggested that item information for one domain does not rely predominantly on the

serial order WMC for the other domain, and further implied that serial order WMCs are applied in a domain-selective manner. Taken together, the model comparison indicates that a parsimonious domain-general serial order WMC is not sufficient to capture the variances in WMCs among individuals when remembering serial order information in verbal and nonverbal domains; instead, our results favor domain-specific serial order WMCs.

Additionally, since such a strong correlation was observed between item and serial order verbal WMC factors in the four-factor model (.98, CI [.96, 1.00]), we compared Model 2 (the four-factor model) with another three-factor model (Model 4). Model 4 (i.e., the three-factor model in Fig. 3d) tested the notion that serial order WMC was best captured as two dissociated yet correlated factors representing domain-specific serial order WMCs, and in the verbal domain, the serial order WMC and the item WMC could be captured with a single factor. This model was also grounded in the previous literature. As Gmeindl et al. (2011) pointed out, item and serial order WMC appear to be more closely bound in the verbal than nonverbal domain. Model 4 consisted of three WMC factors, with a nonverbal item factor derived from four nonverbal item indicators, a nonverbal order factor derived from four nonverbal order indicators, and a verbal factor derived from eight verbal indicators. The three factors were allowed to correlate.

As shown in Table 3, Model 4 had a good fit to the data. Note that, in this three-factor model, the correlation between the verbal WMC factor and the nonverbal serial order WMC factor was .49 (95% CI [.33, .66]), indicating that nonverbal serial order WMC shared around 24% of its variance with the cross-domain WMC and that the relatively large variance in serial order WM performance was of a domain-specific origin. Model 2 that specified four factors did not fit the data significantly different from Model 4 that specified a single verbal WMC and two nonverbal WMC factors, $\Delta\chi^2(3, N = 153) = 1.61, p = .66, \Delta AIC = 4.4$, indicating that the dissociation between item and serial order WMCs was less distinct in the verbal domain compared with the nonverbal domain, but the variance in serial order WMC performance was still accounted for by separate latent constructs for each domain.

Finally, we compared Model 4 with Model 1 (which assumes that there is no separation between item and order processing in either the verbal or the nonverbal domain). Model 4 fit the data reliably better than Model 1, $\Delta\chi^2(2, N = 153) = 14.46, p < .001, \Delta AIC = 10.5$, which supports the hypothesis that item and order WMCs are dissociated at the latent level in the nonverbal domain.

Taken together, the primary result was the comparison between four models, which indicated that a parsimonious domain-general serial order WMC factor does a worse job of capturing the variance in serial order WMC performance

than models that assume a separation of serial order WMC in the verbal and the nonverbal domains. The worst model was Model 3 (Fig. 3c) in which a single latent construct was assumed for order memory performance across domains. Both of the best models—the four-factor Model 2 (Fig. 3b) that assumes a full separation between item and order in both the verbal and nonverbal domains, and the three-factor Model 4 (Fig. 3d) that assumes separated latent constructs of item and order in the nonverbal domain, but a single latent construct for both item and order in the verbal domain—separate the order capacity in the verbal from the order capacity in the nonverbal domain, and show some evidence for a dissociation between item and order WMC, at least in the nonverbal domain.

A secondary result is that item and serial order WMCs in the verbal domain are strongly bound and the variance between them could be captured sufficiently by one latent construct. Due to the high correlation between item and order latent variables in the verbal domain, as well as previous studies suggesting that item and order are more closely bound in the verbal domain than the nonverbal domain (Gmeindl et al., 2011), we compared the three-factor Model 4 (Fig. 3d) with the four-factor Model 2 (Fig. 3b) and found that combining or separating item and serial order WMCs in the verbal domain did not yield any significant difference in the model fit. In the absence of any other information, we would support the more parsimonious model including three latent factors (Model 4) rather than four (Model 2). However, there is clear evidence from other methods suggesting that item and order information are dissociated in the verbal domain, from behavioral studies (Bjork & Healy, 1974; Murdock & Vom Saal, 1967; Wickelgren, 1965), neuroimaging studies (Henson et al., 2000; Kalm & Norris, 2014; Marshuetz et al., 2000; Rajji et al., 2017), and neuropsychological studies (Attout et al., 2012; Kesner et al., 1994; Majerus et al., 2007; Shimamura et al., 1990). Based on this converging evidence from other methodologies and the fact that we cannot statistically distinguish a model that differentiates item and order information in the verbal domain from one that does not, we endorse the four-factor Model 2 (Fig. 3b) that makes a distinction between item and order capacity in both the verbal and the nonverbal domains. A similar issue arose in the nonverbal domain, where the correlation between the latent variables for item and order capacity was still extremely strong (.90). However, both the three-factor Model 4 (Fig. 3d) and the four-factor Model 2 (Fig. 3b) had better fit compared with the two-factor Model 1 (Fig. 3a), indicating that despite this high correlation between the item and order latent variables, they are still statistically separable in the nonverbal domain. Converging evidence from other methodologies also support the item–order dissociation in the nonverbal domain (Adelstein et al., 1992; Amiez & Petrides, 2007; Avons & Mason,

1999; Bowler et al., 2016; Claessen et al., 2016; Dale, 1987; Hopkins et al., 1995; Hsieh et al., 2011; Kesner et al., 1994; Mammarella et al., 2006; Milner et al., 1991; Saint-Aubin et al., 2007; Wansard et al., 2015). Therefore, even with the high correlation between nonverbal item and nonverbal order, we would endorse the three-factor or four-factor models (Model 4 in Fig. 3d or Model 2 in Fig. 3b) over the two-factor Model 1 (Fig. 3a) and suggest that item and order are dissociated at the latent level in the nonverbal domain. In sum, our results favor domain-specific serial order WMCs, although the serial order WMC is hard to separate from item WMC in the verbal domain.

Finally, it is worth considering these results in the context of other studies using the latent-variable approach to examine the structure of WM. Previous studies with this approach have favored the generality over the classic dissociation between verbal and nonverbal WMCs (Ackerman et al., 2002; Kane et al., 2004). As reported in the Supplementary Materials, we compared the four-factor Model 2 (Fig. 3b) to two additional models: (1) a one-factor Model S1 representing a single WMC factor without any dissociations between either item and serial order or the verbal and nonverbal domains and (2) an alternative two-factor Model S2 with dissociated item and serial order WMCs but no dissociation between verbal and nonverbal domains. The results of these model comparisons again showed that the four-factor Model 2 (Fig. 3b), with the verbal–nonverbal dissociation and the item–order dissociation, had the best fit to the data.

Discussion

The current study addressed two questions: Are item and serial order WMCs identifiably distinct constructs? If so, is serial order WMC in verbal and nonverbal domains identifiably distinct? To answer these questions, we tested 153 young adults with college education using verbal and nonverbal–visuospatial sequence matching tasks to probe item-identity and serial order WMCs. We then used the SEM and nested model comparison to directly compare competing hypotheses about the nature of WMC constructs. The results support the item–order dissociation in WMC, although they support this dissociation more strongly in the nonverbal domain than the verbal domain. Furthermore, the results support a dissociation in serial order WMC between the verbal and the nonverbal domain, with models that include dissociated serial order WMCs for verbal and nonverbal domains clearly fitting the data better than the other models considered.

This experiment provides the first evidence using a latent-variable approach to support the dissociation between item-identity and serial order WMCs, with them being more dissociable at the latent level in the nonverbal domain than in

the verbal domain. It is worth noting that dissociated item and serial order WMCs still share a substantial amount of variance in both domains, suggesting that at least in the sequence matching task, item and serial order WMCs interact and covary closely when encoding and/or recalling a sequence. Specifically, item and serial order WMC latent variables correlated stronger (.98) in the verbal domain than the nonverbal domain (.90). Through model comparison, the endorsed four-factor Model 2 (Fig. 3b) was not significantly different from the three-factor Model 4 (with item and serial order WMCs loaded on a single factor in the verbal domain; see Fig. 3d), suggesting that verbal item and serial order WMCs are mostly isomorphic at the latent level in the verbal domain, as Gmeindl et al. (2011) suggested. On the other hand, this overlap in the verbal domain does not indicate that item and serial order WMCs are also isomorphic in the nonverbal domain, supported by the fact that this three-factor Model 4 (Fig. 3d) that separates item and order capacity in the nonverbal domain fits the data better than the two-factor Model 1 (Fig. 3a) that does not separate item and order in either domain.

The reason for the high correlation between item and serial order WMCs is manifold. Empirically, neither “item” nor “order” manifest indicators were collected as process-pure measures; they were selected with the attempt to capitalize on the WMCs to maintain item and serial order information and to detect corresponding changes in a sequence. Hence, indicators extracted from the sequence matching tasks would have correlations between item and order WMCs. However, if they only measure the ability to detect changes in a sequence without further differentiation, “item” and “order” measures would not have a difference at the group level and may have difficulty converging in the SEM when loading on separate latent constructs. Theoretically, the high correlations between item and serial order latent constructs in both domains could have multiple possible reasons. For example, item and order information might need to be closely bound to support performance in a sequence matching task (Gmeindl et al., 2011). An alternative reason is that item information might be maintained actively via subvocal rehearsal simultaneously as the order information when detecting order changes in the sequence matching tasks. From the literature on the order error generation (Farrell & Lewandowsky, 2004; Lewandowsky & Farrell, 2008; Oberauer et al., 2012; Page & Norris, 1998), order errors could either arise from the overlap of positional retrieval cues, or from item similarity in the memory for sequences (Henson et al., 1996). Hence, actively monitoring item similarity in WM to detect order mismatches could create closely correlated performance between item and serial order WM; that is, people who are better at subvocal rehearsal might be better at both item and serial order maintenance. In a similar vein, Camos et al. (2017) suggested that item and

serial order information are maintained by the combination of subvocal rehearsal and attentional refreshing mechanisms. Different levels of involvement of these two mechanisms might give rise to the dissociation between item and serial order WMCs at the latent level, but this common root of maintenance mechanisms could inevitably cause the high correlation between item and order latent constructs.

Using the latent-variable approach, we successfully replicated the item-order WMC dissociation supported by other methodologies and validated this approach for further probing the structure of serial order WM. The key result of this study is the comparison between nested competing models that endorses the model with serial order WM performance loading on separate latent constructs, and thus supports the domain-specific serial order WMC hypothesis over the domain-general hypothesis. This result adds new insight into the literature that has been marked with inconsistent results.

The clarity of this result may reflect the benefit of taking a differential psychology approach to understanding the organization of the cognitive system. In general, serial order information may be more explicitly required when we are processing verbal sequences than visuospatial sequences in the real world, and we might be more adept at recognizing changes in order information for verbal sequences than nonverbal sequences. In order to limit the domain-specific efficiency/difficulty, we attempted to design six-item sequences with different materials equivalently challenging by varying pool sizes of stimulus types in the current study. However, this manipulation could lead to other unequated confounds in task difficulty in two domains, like the exhaustivity of the pool or the proactive interference. When the pool size of a stimulus type was smaller, detecting a change would be easier given that fewer items might need to be monitored in total. Contrarily, with a smaller pool size, the proactive interference might be stronger for a stimulus type, given that a higher chance for the occurrence of overlapping items in the current and past memoranda. On the other hand, representational interference could also be another unintended confound in our current design. We had larger verbal than nonverbal pool in order to balance out the efficiency with verbal sequences from daily life. However, verbal stimuli differentiate at many representational levels (e.g., phonological, semantic, orthographic levels) and nonverbal stimuli do not have as many explicit dimensions to differ (e.g., horizontal and vertical axes for locations, length and direction for arrows). Hence, the verbal domain might have higher multidimension representational interference than the nonverbal domain and having a larger verbal pool could be detrimental to solving this representational interference. It is not feasible to identify and match all representation levels across domains to rule out the domain-specific interference. Hence, we encourage future work to address these potential confounding factors by matching the set size while identifying

other methods to match task difficulty across domains, for example by manipulating the item similarity within the stimulus pool or varying tasks for different domains. As discussed, various factors could affect objective task difficulty, and large discrepancies among individuals makes equating task difficulty across domains even more challenging. In approaches drawing inferences from the group-average performance, it is hard to compare across verbal and nonverbal sequences given that domain-specificity could easily be conflated by nonequivalent task difficulty in two domains. Indeed, with the analysis on the group-average performance, our experiment replicates the finding from Gmeindl et al. (2011): Performance in detecting item changes was significantly better than detecting order swaps in the visuospatial tasks, but not in the verbal tasks. The potential domain-wise difference in task difficulty poses a challenge to interpret this pattern: This group-average result could either be evidence for a separate serial order WMC for visuospatial sequences or evidence for a shared serial order WMC not being used efficiently with harder visuospatial tasks as with easier verbal tasks. However, this difficulty difference is not an issue for individual differences studies, as it only predicts a difference in means, but not a difference in covariation. For the current study, the independence of covariation in serial order WM performance for sequences in different domains across individuals provides strong evidence for the domain-specific serial order WMC hypothesis.

It is worth noting that the model endorsement decision was made based on the combination of the result of nested model comparisons and converging evidence from previous studies. This decision rule we adopted does not weaken the clarity of the SEM approach. Based solely on SEM results, models with serial order WM indicators loading on separate latent constructs (Model 2 in Fig. 3b and Model 4 in Fig. 3d) had significantly better fit compared with the rest of theoretical models, and combining or separating item and order constructs in the verbal domain did not have any significant difference. A common decision rule in SEM studies (especially in exploratory factor analysis studies) is parsimony, in which a model with fewer parameters being estimated is considered more sufficient in accounting for the observed data compared with a larger model. However, whether the parsimony decision rule is always optimal is debatable. It has been suggested that generalizability, parsimony, and precision need to be balanced when formulating the optimal model to account for the observed data (Blalock, 1986; Chin et al., 2008). Sometimes overemphasizing the parsimony decision rule can lead to the endorsement of misspecified but parsimonious models (Marsh & Hau, 1996; West et al., 2012). As Bollen and Long (1993) suggested, “These decision rules cannot replace sound judgment and substantive knowledge of the data.” Therefore, when no significant difference was observed between the two theoretical-driven

models, we chose to incorporate converging evidence from other methodologies rather than the parsimony as our decision rule for the current study.

There are three concerns that might be raised about our interpretation of the domain-specificity from the results. First, as with any study contrasting verbal and nonverbal WM, there are concerns that a verbal strategy could still be used in the nonverbal WM tasks. That is, nonverbal WM tasks are actually engaging the verbal WM system. However, we think that this is unlikely in our current study. To limit the confounding involvement of verbal WM, we selected stimuli that are hard to be verbalized with a simple label. For example, “top middle” is not sufficient to accurately describe a location. Moreover, a fine-grained description of six consecutive items would exceed the limit of verbal WMC (Cowan, 2001) and make the nonverbal tasks extremely difficult to complete, which is not our observation from the data. Even if it was possible for participants to use this verbal strategy, it would be hard to explain why we found that models with separated verbal and nonverbal WMC factors significantly outperformed those that assumed domain-free item and order WMC, as shown in the supplementary materials. We did conduct a follow-up questionnaire within a subset of our sample ($N = 53$), giving participants multiple-choice options and open-ended questions regarding the strategy they adopted to complete each task. The multiple-choice options included sequential encoding without any mental operations, within-domain chunking, within-domain mnemonics (e.g., expanding words to sentences or stories, remembering only the location of arrow tips to form a location sequence), and cross-domain mnemonics (e.g., associating each letter with a picture, labeling locations as number 1 to 10). Overall, only one participant solely and seven participants partially adopted verbal strategies in both nonverbal tasks, indicating that verbal strategy is not used dominantly for nonverbal WM tasks.

Second, an alternative interpretation of the domain-specificity in our results could be that there is a domain-general serial order WM, with domain-specific variability induced by individual expertise for different contents. Researchers have argued that item similarity could affect serial order maintenance (Henson et al., 1996), with it being more difficult to detect order swap errors when items are very similar. Besides from the average level of representational similarities across domains as we discussed previously, the extent to which items are mentally represented as similar might be influenced by individual’s experience and expertise in that domain. We addressed this concern by including four task-specific factors to model performance for a given item type.

Third, it might be argued that the domain-specificity was introduced by the specific task type. The tasks used in the current study were more similar to simple span short-term memory tasks than complex span tasks (e.g., operation

span). Previous latent-variable research has revealed the verbal–nonverbal distinctions with simple but not complex span tasks (Alloway et al., 2006; Kane et al., 2004), although the domain-specificity has been reported with complex span as well (Dang et al., 2012; Shipstead & Yonehiro, 2016). Also, Uittenhove et al. (2019) found the presence of domain-specificity in WM with probe-recognition tasks when it was negligible with recall tasks, but there have been some contradictory findings with recognition tasks for the domain-specificity in WM as well (Ricker et al., 2010; Ricker & Cowan, 2010). Provided that the complexity of the sequence matching paradigm used in the current study lands between simple–complex span and recognition–recall tasks, and the patterns of whether simple span and recognition tasks are prone to domain-specificity findings are mixed, future studies are encouraged to systematically gauge the task effect on the domain-specificity of serial order WM. It is also critical to tease apart whether a domain-general pattern emerges solely because of the nature of serial order WM or because executive control and other higher-level cognitive functions are tapped in complex, difficult tasks. Note that in some versions of WM conceptualization, the spectrum of task complexity coincides with the STM–WM distinction. Without a consensus view of the definition of WM, the current study adopted a broad working definition of WM as the capacity to temporarily maintain and process information. Most WM research does not differentiate WM and STM explicitly, using STM or short-term store to emphasize the storage portion of the WM concept (e.g., Baddeley, 2017; Cowan, 2008; Oberauer et al., 2018). However, sometimes the term WM is reserved for the processing portion of the concept or only for attention-related processes (e.g., Engle, 2002); under this restriction, WM and STM are usually operationalized as complex and simple span tasks, respectively (e.g., Kail & Hall, 2001; Unsworth & Engle, 2007). Although operationalized as distinct tasks, some studies still find that WM and STM reflect the same cognitive process (e.g., Cowan et al., 2003; Hutton & Towse, 2001; for opposite results supporting distinct WM and STM, see also Swanson & Kim, 2007). Given that the distinction between WM and STM could still be ambiguous even under the strict definition of WM, we chose to use WM as a broader term to describe the current study. However, we wanted to note that the tasks in the current study were close to STM tasks and our results on the domain-specificity of serial order processing might only hold for the organization of STM under the strict definition of WM.

In a similar vein, it might also be argued that the sequence matching paradigm taps episodic LTM above and beyond WM. The capacity limit of WM has been argued to be approximately four items (Cowan, 2001), whereas the current study required participants to maintain six or twelve items in memory. Some individual-differences-based WM

models (e.g., Unsworth & Engle, 2007) also suggested that the WMC depends on the capacity of cue-dependent search in LTM. Hence, we encourage future work to further disentangle the involvement of episodic LTM from WM, which could be achieved by varying lengths of sequences or using varying tasks to assess whether the distinct characteristics of LTM (e.g., differential involvement of familiarity and recollection) would be present in retrieval.

One potential limitation of the current study is that it was carried out in a rather restricted high-ability sample of students at Rice University, a highly selective private university. Previous work has suggested that individual differences studies that focus on these high-ability samples exhibit patterns prone to domain-specificity (Jensen, 1998; Kane et al., 2004). Due to the sample limitation, it is possible that the domain-generality of serial order WMCs has been underestimated in the current study. Furthermore, given the length of each testing session (approximately 2.5 hours per participant), we recruited a sample of approximately 160 participants. While this sample size is in line with some recent SEM studies in WM (e.g., Unsworth, 2016; Unsworth & Robison, 2017), it is somewhat small given the complexity of the models. It remains an open question for future research as to whether the same tasks with a different population would yield different results and whether the results would replicate with a larger sample.

Even with these limitations, the results indicate that individual differences in maintaining serial order in WM have domain-specific components of variance, and thus serial order WMC is not identical between the verbal and visuospatial domains. This finding suggests that there are domain-specific components in the mechanism of serial order in WM, but the exact manner of how the domain-specificity plays a role is unclear. Certainly, the domain-specific variance in individual differences in serial order WM performance does not rule out the possible domain-general component in serial order mechanism. For example, it is possible that there is a common set of serial order representations shared across domains (Farrell & Lewandowsky, 2004; Hurlstone & Hitch, 2015, 2018), but the mechanism for binding items and order is domain-specific (Hurlstone et al., 2014). It is also possible in the opposite direction: There could be a domain-general mechanism controlling the efficiency of two separate domain-specific order mechanisms in WM, with its capacity varying between individuals. Alternatively, another fine-grained alternative hypothesis is that cognitive processes required for verbal and visuospatial serial order WMCs are partially overlapping and partially distinct. For example, there could be multiple mechanisms for serial order WM (Camos, 2015) with only some of those mechanisms shared between domains. Some have argued that spatial processing plays a critical role in maintaining order information in verbal sequences (Abrahamse et al.,

2017), but this same spatial mechanism does not appear to contribute to serial order information processing for visuospatial sequences (Ginsburg et al., 2017). Therefore, using space to represent serial order might be a domain-specific serial order process only in the verbal domain, but not in the visuospatial domain, while other serial order processes might be shared across domains. These alternative hypotheses would be consistent with the pattern that we report here, and further follow-up is needed to test them.

Finally, it is worth considering the current study in light of debates about different WM models. Besides the WM models that explicitly proposed a mechanism to account for serial order maintenance (Botvinick & Plaut, 2006; Botvinick & Watanabe, 2007; G. D. A. Brown et al., 2000, 2007; Burgess & Hitch, 1999; Farrell & Lewandowsky, 2002; Henson, 1998; Page & Norris, 1998), most WM models do not have such specification but could account for serial order maintenance with the premise of the model. One strand of WM models is the multiple-component model (Baddeley & Hitch, 1974), in which the WM system is divided into two slave systems including the phonological loop for verbal information and the visuospatial sketchpad for visuospatial information. There are also a central executive and an episodic buffer to allocate attention and perform multi-modal computations; both constructs could operate in both domains. Another strand of WM models assumes no explicit domain-specific processes. These models include the embedded-processes model (Cowan, 1999, 2008, 2019) and the concentric model (Oberauer, 2002), in which WM is the activated portion of long-term memory (LTM) system, as well as the executive attention model (Engle, 2002) and the time-based resource-sharing model (Barrouillet et al., 2004), in which WM is the capacity to control or refresh attention. Our results suggest that there are independent processes for maintaining the serial order in sequences in two domains. The dissociation reported here is best accounted for in the framework of the multiple-component model (Baddeley & Hitch, 1974), which divides WM processing into

dedicated systems for verbal and nonverbal information. In the context of this model, serial order information could be maintained as an independent buffer from item information in each slave system, or as separate mechanisms in the central executive or episodic buffer. As different computational models suggest, this serial order information could exhibit domain-specificity by the context signal being applied separately to verbal and nonverbal items, in the form of the start-based gradient (Page & Norris, 1998), the end-based gradient (G. D. A. Brown et al., 2007), the both-edge gradient (Henson, 1998), or the distributed pattern of dynamic context (G. D. A. Brown et al., 2000; Burgess & Hitch, 1999; Macken et al., 2015). The domain-specificity could also be achieved by having distinct signals in two domains, as suggested by the perceptual-motor model (Hughes et al., 2016). In contrast, the current study might pose a challenge for attention- and activation-based WM models that assume no explicit domain-specificity (Barrouillet et al., 2004; Cowan, 1999; Engle, 2002; Oberauer, 2002). For these models, the verbal–visuospatial dissociation in item WMCs can be explained by differences in the LTM representations for different materials. However, they do not explicitly address how serial order information is maintained in WM; serial order is usually considered as one feature of items (e.g., Barrouillet & Camos, 2014). Given that these models treat WM as a unitary system across domains, they are hard to reconcile with the domain-specific serial order WMC hypothesis supported by the current study.

Appendix 1. Word stimuli

BLADE	BRICK	COACH	CROWN	DOUGH
DRAIN	FENCE	FLOCK	GLOBE	GRAPH
JUICE	LAUGH	LODGE	MOUSE	NERVE
PLATE	PRIZE	QUEST	RIDGE	SAINT
STORM	THEFT	TREAT	VERSE	WOUND

Table 4

Table 4 Properties of word stimuli

WORD	NSYL	NPHN	K-F-FREQ	IMG	FAM	CNC
BLADE	1	4	13	568	517	584
BRICK	1	4	18	574	529	610
COACH	1	4	24	560	509	561
CROWN	1	4	19	602	531	586
DOUGH	1	2	13	558	474	627
DRAIN	1	4	18	540	510	591
FENCE	1	4	30	611	526	597
FLOCK	1	4	10	516	434	477
GLOBE	1	4	13	583	477	535
GRAPH	1	4	17	535	524	553
JUICE	1	3	11	593	567	599
LAUGH	1	3	28	528	594	433
LODGE	1	3	19	464	429	538
MOUSE	1	3	10	615	520	624
NERVE	1	3	12	486	554	488
PLATE	1	4	22	527	556	595
PRIZE	1	4	28	517	508	474
QUEST	1	5	16	460	413	316
RIDGE	1	3	18	543	430	547
SAINT	1	4	16	394	463	458
STORM	1	4	26	587	555	527
THEFT	1	4	10	436	499	361
TREAT	1	4	26	360	534	399
VERSE	1	3	28	489	483	500
WOUND	1	4	28	570	474	561

Note. *NSYL* number of syllables, *NPHN* number of phonemes, *K-F-FREQ* Kucera-Francis written frequency, *IMG* imageability rating, *FAM* familiarity rating, *CNC* Concreteness rating (100–700). Imageability, familiarity, and concreteness are merged from three sets of norms: Pavio (unpublished), Toggia and Battig (1978), and Gilhooly and Logie (1980) as described in Appendix 2 of the MRC Psycholinguistic Database User Manual (Coltheart, 1981)

Appendix 2. Location stimuli

Fig. 4

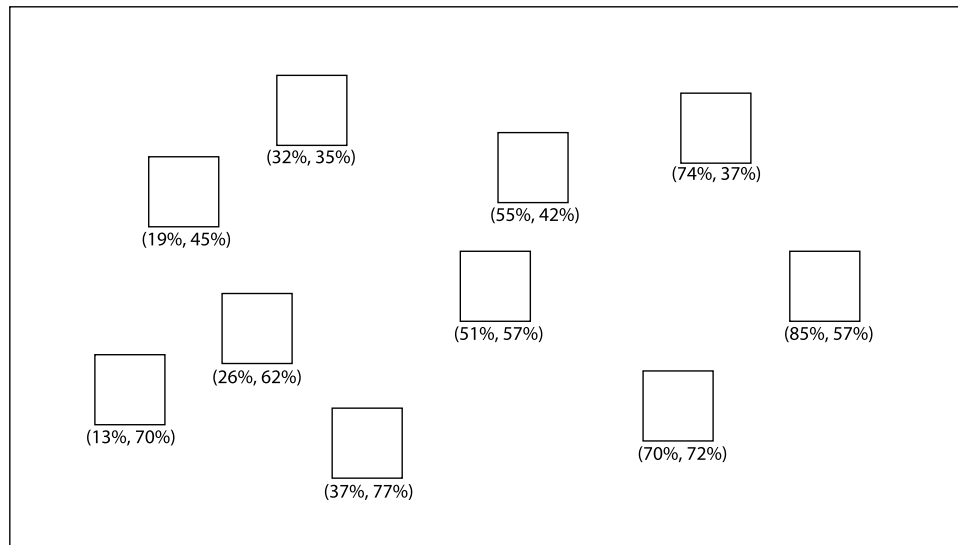


Fig. 4 Location stimuli were presented on a white background, and the coordinate is the position of the center of each location. The location materials consisted of ten randomly selected locations and were at a minimal distance (~4 cm) from one another. The display size was

38.6 cm × 57.4 cm and all items had a visual angle of approximately 3.4 degrees, based on a distance of 42 cm between the participant and the screen

Appendix 3. Arrow stimuli

Fig. 5

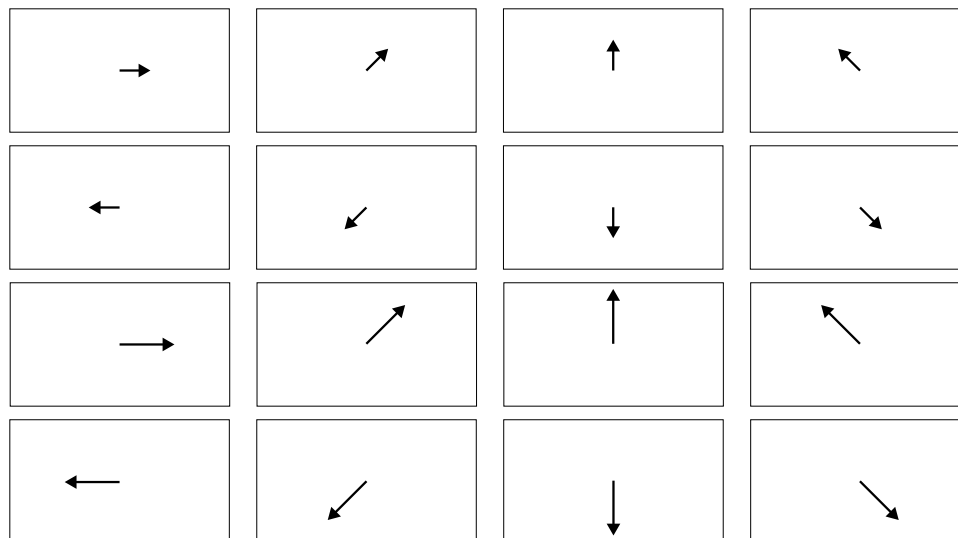


Fig. 5 Arrow stimuli were presented on a white background, starting from the center with varying lengths and directions. The arrow materials consisted of sixteen randomly selected locations. The display size was

38.6 cm × 57.4 cm and all items had a visual angle of approximately 2.7 to 5.4 degrees, based on a distance of 42 cm between the participant and the screen

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13421-021-01260-4>.

Acknowledgments We thank Bailey Brashears and Kristina Johnson for their help in data collection. This work was supported by the National Science Foundation (NSF BCS1752751) and The Institute for Rehabilitation and Research (TIRR) Foundation Mission Connect grant.

References

- Abrahamse, E. L., van Dijck, J. P., & Fias, W. (2017). Grounding Verbal Working Memory: The Case of Serial Order. *Current Directions in Psychological Science*, 26(5), 429–433. <https://doi.org/10.1177/0963721417704404>
- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, 131(4), 567–589. <https://doi.org/10.1037/0096-3445.131.4.567>
- Adelstein, T. B., Kesner, R. P., & Strassberg, D. S. (1992). Spatial recognition and spatial order memory in patients with dementia of the Alzheimer's type. *Neuropsychologia*, 30(1), 59–67. [https://doi.org/10.1016/0028-3932\(92\)90014-D](https://doi.org/10.1016/0028-3932(92)90014-D)
- Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, 87(2), 85–106. <https://doi.org/10.1016/j.jecp.2003.10.002>
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, 77(6), 1698–1716. <https://doi.org/10.1111/j.1467-8624.2006.00968.x>
- Amiez, C., & Petrides, M. (2007). Selective involvement of the mid-dorsolateral prefrontal cortex in the coding of the serial order of visual stimuli in working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 104(34), 13786–13791. <https://doi.org/10.1073/pnas.0706220104>
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>
- Arrindell, W. A., & van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, 9(2), 165–178. <https://doi.org/10.1177/014662168500900205>
- Attout, L., Van der Kaa, M.-A., George, M., & Majerus, S. (2012). Dissociating short-term memory and language impairment: The importance of item and serial order information. *Aphasiology*, 26(3/4), 355–382. <https://doi.org/10.1080/02687038.2011.604303>
- Avons, S. E., & Mason, A. (1999). Effects of Visual Similarity on Serial Report and Item Recognition. *The Quarterly Journal of Experimental Psychology*, 52A(1), 217–240.
- Baddeley, A. D. (2017). Exploring working memory: Selected works of Alan Baddeley. In: *Exploring Working Memory: Selected works of Alan Baddeley*. <https://doi.org/10.4324/9781315111261>
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *Psychology of Learning and Motivation*, 8(C), 47–89. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, 9(1), 78–102. https://doi.org/10.1207/S15328007SEM0901_5
- Barrouillet, P., & Camos, V. (2014). *Working memory: Loss and reconstruction*. Psychology Press.
- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133(1), 83–100. <https://doi.org/10.1037/0096-3445.133.1.83>
- Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, 13, 80–97.
- Blalock, H. M. (1986). Multiple causation, indirect measurement and generalizability in the social sciences. *Synthese*, 68(1), 13–36.
- Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models* (Vol. 154). SAGE Publications.
- Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, 113(2), 201–233. <https://doi.org/10.1037/0033-295X.113.2.201>
- Botvinick, M. M., & Watanabe, T. (2007). From numerosity to ordinal rank: a gain-field model of serial order representation in cortical working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(32), 8636–8642. <https://doi.org/10.1523/JNEUROSCI.2110-07.2007>
- Bowler, D. M., Poirier, M., Martin, J. S., & Gaigg, S. B. (2016). Non-verbal short-term serial order memory in autism spectrum disorder. *Journal of Abnormal Psychology*, 125(7), 886–893. <https://doi.org/10.1037/abn0000203>
- Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107(1), 127–181.
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3), 539–576. <https://doi.org/10.1037/0033-295X.114.3.539>
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230–258. <https://doi.org/10.1177/0049124192021002005>
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106(3), 551–581. <https://doi.org/10.1037/0033-295X.106.3.551>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Camos, V. (2015). Storing verbal information in working memory. *Current Directions in Psychological Science*, 24(6), 440–445. <https://doi.org/10.1177/0963721415606630>
- Camos, V., Lagner, P., & Loaiza, V. M. (2017). Maintenance of item and order information in verbal working memory. *Memory*, 25(8), 953–968. <https://doi.org/10.1080/09658211.2016.1237654>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>
- Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. Plenum Press. <https://doi.org/10.1007/978-1-4684-2262-7>
- Chin, W. W., Peterson, R. A., & Brown, S. P. (2008). Structural equation modeling in marketing: Some practical reminders. *Journal of Marketing Theory and Practice*, 16(4), 287–298. <https://doi.org/10.2753/MTP1069-6679160402>
- Claessen, M. H. G., Visser-Meily, J. M. A., Jagersma, E., Braspenning, M. E., & van der Ham, I. J. M. (2016). Dissociating spatial and spatiotemporal aspects of navigation ability in chronic stroke patients. *Neuropsychology*, 30(6), 697–708. <https://doi.org/10.1037/neu0000260>
- Coffman, D. L., & MacCallum, R. C. (2005). Using parcels to convert path analysis models into latent variable models. *Multivariate Behavioral Research*, 40(2), 235–259. https://doi.org/10.1207/s15327906mbr4002_4

- Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods, 19*(2), 300–315. <https://doi.org/10.1037/a0033805>
- Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547–552. <https://doi.org/10.1016/J.TICS.2003.10.005>
- Cowan, N. (1999). An embedded-processes model of working memory. In: A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). <https://doi.org/10.1017/CBO9781139174909.006>
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–114. <https://doi.org/10.1017/S0140525X01003922>
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*, 323–338. [https://doi.org/10.1016/S0079-6123\(07\)00020-9](https://doi.org/10.1016/S0079-6123(07)00020-9)
- Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin, 145*(8), 822–847. <https://doi.org/10.1037/bul0000199>
- Cowan, N., Towse, J. N., Hamilton, Z., Sauls, J. S., Elliott, E. M., Lacey, J. F., Moreno, M. V., & Hitch, G. J. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General, 132*(1), 113–132. <https://doi.org/10.1037/0096-3445.132.1.113>
- Dale, R. H. I. (1987). Similarities between human and animal spatial memory: Item and order information. *Animal Learning & Behavior, 15*(3), 293–300. <https://doi.org/10.3758/BF03205022>
- Dang, C. P., Braeken, J., Ferrer, E., & Liu, C. (2012). Unitary or non-unitary nature of working memory? Evidence from its relation to general fluid and crystallized intelligence. *Intelligence, 40*(5), 499–508. <https://doi.org/10.1016/j.intell.2012.05.002>
- Depoorter, A., & Vandierendonck, A. (2009). Evidence for modality-independent order coding in working memory. *Quarterly Journal of Experimental Psychology, 62*(3), 531–549. <https://doi.org/10.1080/17470210801995002>
- Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling: A Multidisciplinary Journal, 2*(2), 119–143.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23. <https://doi.org/10.1111/1467-8721.00160>
- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*(1), 59–79. <https://doi.org/10.3758/BF03196257>
- Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language, 51*, 115–135. <https://doi.org/10.1016/j.jml.2004.03.007>
- Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology, 54A*(1), 1–30. <https://doi.org/10.1080/027>
- Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395–427. <https://doi.org/10.3758/BF03201693>
- Ginsburg, V., Archambeau, K., van Dijck, J.-P., Chetail, F., & Gevers, W. (2017). Coding of serial order in verbal, visual and spatial working memory. *Journal of Experimental Psychology: General, 146*(5), 632–650. <https://doi.org/10.1037/xge0000278>
- Gmeindl, L., Walsh, M., & Courtney, S. M. (2011). Binding serial order to representations in working memory: A spatial/verbal dissociation. *Memory & Cognition, 39*(1), 37–46. <https://doi.org/10.3758/s13421-010-0012-9>
- Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition, 69*, 14–25. <https://doi.org/10.1016/j.concog.2019.01.008>
- Hanley, J. R., Young, A. W., & Pearson, N. A. (1991). Impairment of the visuo-spatial sketch pad. *The Quarterly Journal of Experimental Psychology Section A, 43*(1), 101–125. <https://doi.org/10.1080/14640749108401001>
- Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*(2), 73–137. <https://doi.org/10.1006/cogp.1998.0685>
- Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology, 49A*(1), 80–115.
- Henson, R. N. A., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia, 38*(4), 426–440. [https://doi.org/10.1016/S0028-3932\(99\)00098-6](https://doi.org/10.1016/S0028-3932(99)00098-6)
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*(1), 53–60. <https://doi.org/10.1037/1082-989X.12.1.58>
- Hopkins, R. O., Kesner, R. P., & Goldstein, M. (1995). Item and order recognition memory in subjects with hypoxic brain injury. *Brain and Cognition, 27*, 180–201.
- Hsieh, L.-T., Ekstrom, A. D., & Ranganath, C. (2011). Neural oscillations associated with item and temporal order maintenance in working memory. *Journal of Neuroscience, 31*(30), 10803–10810. <https://doi.org/10.1523/JNEUROSCI.0828-11.2011>
- Hu, L. T., & 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. <https://doi.org/10.1080/10705519909540118>
- Hughes, R. W., Chamberland, C., Tremblay, S., & Jones, D. M. (2016). Perceptual-motor determinants of auditory-verbal serial short-term memory. *Journal of Memory and Language, 90*, 126–146. <https://doi.org/10.1016/j.jml.2016.04.006>
- Hulin, C., Netemeyer, R., & Cudeck, R. (2001). Can a reliability coefficient be too high? *Journal of Consumer Psychology, 10*(1), 55–58.
- Hurlstone, M. J., & Hitch, G. J. (2015). How is the serial order of a spatial sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(2), 295–324. <https://doi.org/10.1037/a0038223>
- Hurlstone, M. J., & Hitch, G. J. (2018). How is the serial order of a visual sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning Memory and Cognition, 44*(2), 167–192. <https://doi.org/10.1037/xlm000440.supp>
- Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140*(2), 339–373. <https://doi.org/10.1037/a0034221>
- Hutton, U. M. Z., & Towse, J. N. (2001). Short-term memory and working memory as indices of children's cognitive skills. *Memory, 9*(4/6), 383–394. <https://doi.org/10.1080/09658210042000058>
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.
- Jones, D. M., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning,*

- Memory, and Cognition*, 21(4), 1008–1018. <https://doi.org/10.1037/0278-7393.21.4.1008>
- Kail, R., & Hall, L. K. (2001). Distinguishing short-term memory from working memory. *Memory & Cognition*, 29(1), 1–9. <https://doi.org/10.3758/BF03195735>
- Kalm, K., & Norris, D. (2014). The representation of order information in auditory-verbal short-term memory. *The Journal of Neuroscience*, 34(20), 6879–6886. <https://doi.org/10.1523/jneurosci.4104-13.2014>
- Kane, M. J., Tuholski, S. W., Hambrick, D. Z., 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. <https://doi.org/10.1037/0096-3445.133.2.189>
- Kesner, R. P., Hopkins, R. O., & Fineman, B. (1994). Item and order dissociation in humans with prefrontal cortex damage. *Neuropsychologia*, 32(8), 881–891. [https://doi.org/10.1016/0028-3932\(94\)90040-X](https://doi.org/10.1016/0028-3932(94)90040-X)
- Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *Psychology of Learning and Motivation—Advances in Research and Theory*, 49, 1–48. [https://doi.org/10.1016/S0079-7421\(08\)00001-7](https://doi.org/10.1016/S0079-7421(08)00001-7)
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1
- Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, 18(3), 285–300. <https://doi.org/10.1037/a0033266>
- Macken, B., Taylor, J., & Jones, D. M. (2015). Limitless capacity: A dynamic object-oriented approach to short-term memory. *Frontiers in Psychology*, 6, 293. <https://doi.org/10.3389/fpsyg.2015.00293>
- Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex*, 112, 122–133. <https://doi.org/10.1016/j.cortex.2018.04.016>
- Majerus, S., Van Der Linden, M., Braissand, V., & Eliez, S. (2007). Verbal short-term memory in individuals with chromosome 22q11.2 deletion: Specific deficit in serial order retention capacities? *American Journal on Mental Retardation*, 112(2), 79–93. [https://doi.org/10.1352/0895-8017\(2007\)112\[79:VSMIHW\]2.CO;2](https://doi.org/10.1352/0895-8017(2007)112[79:VSMIHW]2.CO;2)
- Majerus, S., D'Argembeau, A., Perez, T. M., Belayachi, S., Van der Linden, M., Collette, F., Salmon, E., Seurinck, R., Fias, W., & Maquet, P. (2010). The commonality of neural networks for verbal and visual short-term memory. *Journal of Cognitive Neuroscience*, 22(11), 2570–2593. <https://doi.org/10.1162/jocn.2009.21378>
- Mammarella, I. C., Cornoldi, C., Pazzaglia, F., Toso, C., Grimoldi, M., & Vio, C. (2006). Evidence for a double dissociation between spatial-simultaneous and spatial-sequential working memory in visuospatial (nonverbal) learning disabled children. *Brain and Cognition*, 62(1), 58–67. <https://doi.org/10.1016/j.bandc.2006.03.007>
- Marsh, H. W., & Hau, K.-T. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education*, 64(4), 364–390. <https://doi.org/10.1080/00220973.1996.10806604>
- Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33(2), 181–220.
- Marshuetz, C., Smith, E. E., Jonides, J., DeGutis, J., & Chenevert, T. L. (2000). Order information in working memory: fMRI evidence for parietal and prefrontal mechanisms. *Journal of Cognitive Neuroscience*, 12(Supplement 2), 130–144. <https://doi.org/10.1162/08989290051137459>
- Matsunaga, M. (2008). Item parceling in structural equation modeling: A primer. *Communication Methods and Measures*, 2(4), 260–293. <https://doi.org/10.1080/19312450802458935>
- McMackin, D., Cockburn, J., Anslow, P., & Gaffan, D. (1995). Correlation of fornix damage with memory impairment in six cases of colloid cyst removal. *Acta Neurochirurgica*, 135(1/2), 12–18. <https://doi.org/10.1007/BF02307408>
- Milner, B., Corsi, P., & Leonard, G. (1991). Frontal-lobe contribution to recency judgements. *Neuropsychologia*, 29(6), 601–618.
- Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision Research*, 141, 4–15. <https://doi.org/10.1016/j.visres.2017.11.001>
- Morey, C. C., & Mall, J. T. (2012). Cross-domain interference costs during concurrent verbal and spatial serial memory tasks are asymmetric. *Quarterly Journal of Experimental Psychology*, 65(9), 1777–1797. <https://doi.org/10.1080/17470218.2012.668555>
- Murdock, B. B., & Vom Saal, W. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology*, 74(1), 137–143. <https://doi.org/10.1037/h0024507>
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning Memory and Cognition*, 28(3), 411–421. <https://doi.org/10.1037/0278-7393.28.3.411>
- Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, 19(5), 779–819. <https://doi.org/10.3758/s13423-012-0272-4>
- Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, 144(9), 885–958. <https://doi.org/10.1037/bul0000153>
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105(4), 761–781. <https://doi.org/10.1037/0033-295X.105.4.761-781>
- Rajji, T. K., Zomorodi, R., Barr, M. S., Blumberger, D. M., Mulsant, B. H., & Daskalakis, Z. J. (2017). Ordering information in working memory and modulation of gamma by theta oscillations in humans. *Cerebral Cortex*, 27(2), 1482–1490. <https://doi.org/10.1093/cercor/bhw326>
- Raubenheimer, J. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology*, 30(4), 59–64. <https://doi.org/10.4102/sajip.v30i4.168>
- Rhemtulla, M. (2016). Population performance of sem parceling strategies under measurement and structural model misspecification. *Psychological Methods*, 21(3), 348–368. <https://doi.org/10.1037/met0000072>
- Ricker, T. J., & Cowan, N. (2010). Loss of visual working memory within seconds: The combined use of refreshable and non-refreshable features. *Journal of Experimental Psychology: Learning Memory and Cognition*, 36(6), 1355–1368. <https://doi.org/10.1037/a0020356>
- Ricker, T. J., Cowan, N., & Morey, C. C. (2010). Visual working memory is disrupted by covert verbal retrieval. *Psychonomic Bulletin & Review*, 17(4), 516–521. <https://doi.org/10.3758/PBR.17.4.516>
- Rogers, W. M., & Schmitt, N. (2004). Parameter recovery and model fit using multidimensional composites: A comparison of four empirical parceling algorithms. *Multivariate Behavioral Research*, 39(3), 379–412. https://doi.org/10.1207/S15327906MBR3903_1

- Ryu, E. (2011). Effects of skewness and kurtosis on normal-theory based maximum likelihood test statistic in multilevel structural equation modeling. *Behavior Research Methods*, 43(4), 1066–1074. <https://doi.org/10.3758/s13428-011-0115-7>
- Saint-Aubin, J., Tremblay, S., & Jalbert, A. (2007). Eye movements and serial memory for visual-spatial information: Does time spent fixating contribute to recall? *Experimental Psychology*, 54(4), 264–272. <https://doi.org/10.1027/1618-3169.54.4.264>
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4–27.
- Shimamura, A. P., Janowsky, J. S., & Squire, L. R. (1990). Memory for the temporal order of events in patients with frontal lobe lesions and amnesic patients. *Neuropsychologia*, 28(8), 803–813. [https://doi.org/10.1016/0028-3932\(90\)90004-8](https://doi.org/10.1016/0028-3932(90)90004-8)
- Shipstead, Z., & Yonehiro, J. (2016). The domain-specific and domain-general relationships of visuospatial working memory to reasoning ability. *Psychonomic Bulletin & Review*, 23(5), 1504–1512. <https://doi.org/10.3758/s13423-016-1021-x>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology*, 26, 4–7. <https://doi.org/10.2139/ssrn.2160588>
- Soemer, A., & Saito, S. (2016). Domain-specific processing in short-term serial order memory. *Journal of Memory and Language*, 88, 1–17. <https://doi.org/10.1016/j.jml.2015.12.003>
- Swanson, L., & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence*, 35(2), 151–168. <https://doi.org/10.1016/j.intell.2006.07.001>
- Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Erlbaum.
- Uittenhove, K., Chaabi, L., Camos, V., & Barrouillet, P. (2019). Is working memory storage intrinsically domain-specific? *Journal of Experimental Psychology: General*, 148(11), 2027–2057. <https://doi.org/10.1037/xge0000566>
- Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(1), 50–61. <https://doi.org/10.1037/xlm0000148>
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114(1), 104–132. <https://doi.org/10.1037/0033-295X.114.1.104>
- Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning Memory and Cognition*, 39(3), 832–842. <https://doi.org/10.1037/a0029669>
- Unsworth, N., & Robison, M. K. (2017). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(12), 1962–1987. <https://doi.org/10.1037/xlm0000421>
- Vallar, G., & Baddeley, A. D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 151–161. [https://doi.org/10.1016/S0022-5371\(84\)90104-X](https://doi.org/10.1016/S0022-5371(84)90104-X)
- Vandierendonck, A. (2016). Modality independence of order coding in working memory: Evidence from cross-modal order interference at recall. *Quarterly Journal of Experimental Psychology*, 69(1), 161–179. <https://doi.org/10.1080/17470218.2015.1032987>
- Wansard, M., Bartolomeo, P., Bastin, C., Segovia, F., Gillet, S., Duret, C., & Meulemans, T. (2015). Support for distinct subcomponents of spatial working memory: A double dissociation between spatial-simultaneous and spatial-sequential performance in unilateral neglect. *Cognitive Neuropsychology*, 32(1), 14–28. <https://doi.org/10.1080/02643294.2014.995075>
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. *Handbook of Structural Equation Modeling*, 1, 209–231.
- Wickelgren, W. A. (1965). Short-term memory for phonemically similar lists. *The American Journal of Psychology*, 78(4), 567–574. <https://doi.org/10.2307/1420917>
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 73(6), 913–934. <https://doi.org/10.1177/0013164413495237>
- Yang, C., Nay, S., & Hoyle, R. H. (2010). Three approaches to using lengthy ordinal scales in structural equation models: Parceling, latent scoring, and shortening scales. *Applied Psychological Measurement*, 34(2), 122–142.

Open practices statement The materials, data, and analysis R script for all tasks are available at Open Science Framework (<https://osf.io/px7nz/>).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.