TITLE: Item accessibility in children's working memory: convergent evidence from 3 related paradigms

Hannah E. Roome[1], John N. Towse[2]

[1]School of Psychology, Newcastle University, Newcastle, UK

[2]Department of Psychology, Lancaster University, Lancaster UK

Corresponding Author: Hannah Roome, Hannah.roome@newcastle.ac.uk

Conflict of Interest: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding: This research was supported by a grant from the UK Economic and Social Research Council to Chris Jarrold and John Towse (RES 062 23 2467).

Acknowledgments: We are very thankful for the cooperation of staff, parents, and children at Sandylands Primary School, Morecombe, St Andrews Primary School Wigan, St Anthony's Catholic Primary School, Preston, and Chorlton CofE Primary School, Manchester who participated in the running of this project. Thank you to J. Patel and A. Shulpina for processing the audio recordings and generating the different response times for each task. Thank you to B. Rawlings for his comments on the manuscript leading up to the submission of this article. The underlying research materials and data are part of a PhD thesis

Data Sharing: The raw data from this study can be found here: https://osf.io/kyurz/

Abstract

The development of working memory is typically measured by accuracy. Whilst this helps to understand age-related changes in *capacity*, it is not necessarily so well adapted to reveal working memory *processes*. Chronometry provides the opportunity to pivot more towards a consideration of processes unfolding over time during participants test performance. We focussed on the idea that working memory items are fragile, and the ability to access and reconstruct memoranda is a key factor in both the development of working memory capacity, and more complex cognition (such as reading comprehension). This study deployed different memory tasks, all potentially requiring cue-dependent search and retrieval processes. These included storage-only paradigms and a complex span paradigm that interleaves processing and storage elements. Three different response times were extracted from participants responses. The preparatory interval is a reflection of the time taken to prepare to be recalled memoranda; the interword pause reflects the time between recalled items where participants access the next item to be recalled, and the word duration indexes the articulation speed of the spoken memory item. Here, we found accuracy on these tasks remained consistent. However, each task showed a nuanced chronometric profile. Nonetheless, children's preparatory interval, i.e., predicted both individual differences in working memory capacity and reading comprehension. In conclusion, the research highlights the importance of reconstructive retrieval processes in WMC and reading comprehension. It also further emphasises the significance of response time measures in understanding the nuances of memory tasks and their impact on understanding cognitive development.

Key words: response time, reconstruction, children, working memory capacity.

Introduction

Working memory is a cognitive architecture and system that enables functional flexibility so as to process, maintain and recall information (A. Baddeley, 1986; A. D. Baddeley & Hitch, 1974; Logie, 2023). Its central role in cognition is emphasised by research showing that individual differences in working memory performance systematically associate with a variety of complex cognitive activities (Conway et al., 2007). Complex span – requiring the simultaneous processing and retention of incoming information - has been the traditionally privileged measure of working memory capacity (WMC).This naturally brings into focus an emphasis on *processing* and *maintenance* operations (e.g., Daneman & Carpenter, 1980; Turner & Engle, 1989). However, this approach potentially obscures the ultimate assembly and production of information from working memory – *recall* – which is also recognised as contributing to WMC (e.g., Healey & Miyake, 2009; Hurlstone, 2024; Roome et al., 2019; Unsworth & Engle, 2007). In this paper, we take a developmental perspective, examining the accuracy and chronometric profiles of children's recall, to inform about working memory processes and their links to reading comprehension (as an applied cognitive ability). In doing so, we draw on evidence from a complex span task alongside complementary and convergent memory paradigms.

Information within working memory is transient and fragile, susceptible to disruption or loss of fidelity. Thus, reporting working memory contents is neither immediate nor unambiguous. Children can prioritise particular sequence items (Atkinson et al., 2021) yet these and other data show that they are heavily reliant on outputting the most recent information and find it very difficult to retrieve displaced memoranda (De Alwis et al., 2009; Hall et al., 2015; Jarrold et al., 2015; Roome et al., 2014). Yet, the lack of correct recall does not necessarily mean that they have no viable representations to draw upon (Roome et al., 2019). Some memory representations must be available even where these do not suffice for complete and accurate recall.

Unsworth and Engle (2007) propose that working memory requires (a) active maintenance of items within primary memory (PM), and (b) retention within secondary memory (SM). At the point of recall, PM items are active and accessible and thus recalled quickly and relatively easily. But the remaining SM items must be reported through a long-term controlled retrieval process (Healey &

Miyake, 2009; Kane & Engle, 2000; Miyake & Friedman, 2004), a more deliberate, demanding, competitive process that requires internal generation of contextual cues and search delimitation to access target memoranda (Unsworth & Engle, 2007; see also Cowan et al., 2024 for a review of links between memory and attention)..

As working memory demands increase - for example, through more time-consuming processing episodes - accuracy slows down. But revealingly, *the pauses* between successfully recalled memoranda also increases (Towse et al., 2008). This implies that working memory representations have become degraded and therefore require more time to be clarified or redintegrated, i.e., restored (Hulme et al., 1999). Importantly, measuring the timing of children's recall provides a potentially critical investigative opportunity. That is, the chronometry of recall can yield important regularities and act as a convergent source of evidence about the underlying working memory representations, alongside the much more commonly studied accuracy of recall (see also Cowan, 1992; Cowan et al., 1998, 2003; Cowan & Elliott, 2023; Hurlstone, 2019).

These two convergent sources of evidence, for the external support of item recall and the systematic internal dynamic temporal processes that result in successful recall, are consistent with the Recall Reconstruction hypothesis for complex span performance (Towse et al., 2008, 2010). This hypothesis emphasises that recall can emerge from multiple cues and that participants can, where the task structure permits, use more than the to-be-remembered item itself as the basis for redintegration and recall. For example, the integration of processing and retention elements of the complex span task can facilitate the use of semantic representations within processing to retrieve to be remembered items (Osaka et al., 2002; Tanaka et al., 2014).

Consistent with this account, (1) complex span recall is slower than STM task recall and (2) recall from reading span tasks performance is much slower than recall from counting span, where processing does not uniquely specify the memorandum (Cowan et al., 2003). Furthermore, recall is slower in reading span tasks where semantic links exist between reading and to-be-recalled items, than where no semantic links exist (Towse et al., 2008, 2010). Cowan et al., (2003) also showed that individual differences in recall pauses are associated with individual differences in working memory and adults' and children's scholastic achievement.

To understand further children's accessibility of working memory representations and the development of reconstructive retrieval processes, we administered three working memory tasks: listening span, conceptual span, and delayed cued recall. For simplicity we measure timing on correct sequence recalls only - with an incorrect sequence, additional processes (e.g., response omissions, repeats, attempts at recall recovery) will likely add additional noise to the chronometric signals that we recognise to be already subject to considerable potential variance. With respect to the partitioning of verbal recall timing, we follow previous work in separating the output sequence into three distinguishable, but related, phases. In order of appearance within the sequence, these are the preparatory interval (PI), word duration (WD) and interword pause (IWP). PI indexes the preparation of sequence information (along with for example the reaction time to the recall cue). WDs indexes articulation speed (and for example linguistic parameters such as word length and word complexity). IWPs indexes processes that include access to and redintegration of the target memorandum.

Reading span is a complex span task that has been at the cornerstone of testing WMC across the lifespan (e.g., Conway et al., 2007). According to the dual memory framework, reading span memoranda are held initially in PM, but displaced into SM due to the requirement to engage in the subsequent processing component of the task. Given that already cited work exists to characterise recall timing in reading span, we use a verbal based variant, listening span, as a baseline paradigm, against which we can examine and integrate data from complementary and convergent procedures.

The *delayed cued recall* (DCR) task (Miyake & Friedman, 2004) involves a presentation of three memoranda along with an independent continuous distractor task, followed by a superordinate cue to just one of the three originally presented items. The structure is simple – a relatively low memory load, but the distractor actively is used to degrade the fidelity of the memory representations (Brown, 1958; Peterson & Peterson, 1959). Miyake and Friedman (2004) found that adults' speed of report was a significant and unique mediator of individual differences between working memory and cognitive ability. We assume here that providing children with an external cue should facilitate internally driven cue-dependent SM search and access process. Children show low accuracy accompanied by long response latencies on the delayed cued recall task (Roome et al., 2019), which underscores the conclusion that recall involves effortful and extensive search (Cowan et al., 2003).

We sought to conceptually replicate this finding while also investigating whether task performance is linked to other cued memory tasks.

The third paradigm, conceptual span, is a cued free recall task involving presentation of nine words presented associated with different superordinate categories, where superordinate labels are used to specify which three words should be recalled. This was first used as a measure of semantic short-term memory in adults (Haarmann et al., 2003) and active long-term memory (Woltz & Was, 2006). However, the task also measures recall selectivity and efficiency. That is, the recall cue categories provide external search criteria for the selection of representations from what is likely to be a supra-span list for children. Children show a very constrained capacity to retrieve memoranda that is out of immediate access (i.e. primary items) on a free recall task (Roome et al., 2014). By providing an external cue, the conceptual span task provides the opportunity to measures the ability to efficiently select and retrieve multiple target memory representations that are not in an immediate, accessible state.

We chose to investigate the predictive links between these complementary measures of working memory for reading comprehension. Reading comprehension requires the individual to maintain or store information that they have just read and use it to disambiguate the text and create a meaningful representation (Daneman & Carpenter, 1980; Daneman & Merikle, 1996). Working memory has been used as a successful explanation regarding individual differences in children's reading comprehension abilities, despite controlling for their decoding, word recognition skill and vocabulary knowledge (Swanson & Berninger, 1995; Yuill et al., 1989). In addition, Nouwens et al., (2017) showed the predictive role of conceptual span, (as opposed to phonological storage) in children's reading comprehension. Thus, understanding the mechanisms that underpin the working memory system will help to illuminate reasons why it contributes to the prediction of reading comprehension.

In summary, our key objective is to describe reconstructive retrieval processes that contribute to individual differences in reading comprehension – from a developmental perspective. To achieve this, we studied 7-10 year-olds. Specifically, we sought to establish how different measures of accuracy and chronometry (PI, IWPs, WD) affect our understanding of 1) the different retrieval profiles for storage-only and complex tasks; 2) the impact of cued recall and its ability to facilitate retrieval of long-term representations; 3) whether accuracy and response time measures predict individual differences in working memory capacity and, 4) reading comprehension.

Methods

Participants

The experiment was carried out in three different schools in the North-West of England, UK. Parental consent was gained for 104 children split into two age groups: 51 7-8 year olds (M = 8.01 years, range: 7.03 – 8.10 years, 29 females) and 53 9-10 year olds (M = 9.07 years; range: 9.00-10.07 years, 27 female). Five children did not complete the full testing sessions. One child was excluded from the chronometric analyses, as the experimenter did not gain permission to record the testing session.

Materials

The stimulus pool for the conceptual span and delayed cued recall tasks were reported by Roome et al., (2019). In total, 177 one- or two- syllable concrete nouns were taken from 27 semantic categories (Van Overschelde et al., 2004). The stimuli were chosen from the Morrison et al., (1997) corpus based on their age-appropriateness for the youngest age group (7 years old) and word association. Pilot data established common word associations; 40 adult participants wrote down the first three concrete nouns that came to mind from the target categories. Only items specified by less than 50% of participants were used as candidate stimuli. For example, for semantic category *fruit*, "apple" and "banana" were excluded, as they were produced 85% and 55% of the time respectively. Instead, "cherry", "grapes" and "pear" were included for this specific semantic category. Stimuli were recorded in a female voice and edited using Audacity. Stimuli were presented at a rate of 750ms per word followed by a 250ms interstimulus interval (ISI).

The listening span task, based on Roome et al., (2019), used a corpus of 88 sentences and 88 words. The sentences were taken from (Towse et al., 1998) and recorded in a female voice. The mean number of words in each incomplete sentence was 5.41 words, with a mean recording length of 1.71s, spoken at 3.09 words per second.

Design and Procedure

All participants were tested individually in a quiet classroom setting at their school. Two sessions needed to be completed, lasting approximately 30 minutes each, separated by one week. The first session involved the three memory measures (conceptual span task, delayed cued recall task and listening span task, Figure 1), which were run using PsychoPy software (Pierce, 2007, 2009), on a 15-inch MacBook laptop. The second session included reading comprehension (YARC) and vocabulary measures (BPVS III).

Memory Measures

Delayed cued recall task. This task followed the protocol described in Roome et al., (2019) adapted from Miyake and Friedman (2004) (Figure 1A). Participants heard three unrelated, concrete nouns, each presented for 750ms, with a 250ms interstimulus interval. After the stimuli presentation, a 15s distractor activity was implemented. Adults counted backwards aloud in threes from a random three-digit number shown on the computer screen. Children counted aloud target objects hidden in a visual scene. Different distractor activities were designed to occupy their attention in the verbal domain, whilst it not being so difficult that the activity caused undue fatigue (children) or being so easy that they were able to actively maintain the memoranda (adults). Afterwards, participants were cued by a semantic category to recall one target item. The experimenter pressed a button as soon as the participant started to produce their answer. This eliminated motor movement / button pressing as a chronometric variable. Participants completed 15 trials, separated into three testing blocks (five trials per block). The serial position of the target memory item was pseudo-randomised across trials, generating a total of five trials for each serial position. The number of correct responses and the total proportion of correct recall were recorded. The total proportion of correct recall was calculated by the total number of correctly recalled items divided by the total number of TBR items.

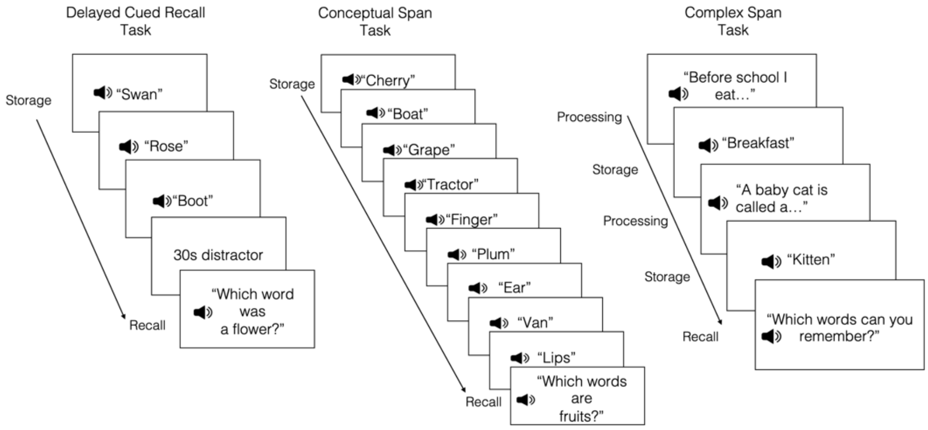
Conceptual span task. Prior to the administration of the task, children were familiarised with the semantic categories included in the task. To do so, the experimenter read each semantic category out loud and asked participants if they knew what each category was. If they did not, the experimenter provided them with a definition and an example item. This was carried out to try and ensure that participants were not disadvantaged at recall by their semantic knowledge.

For the task itself, nine-item lists were generated, incorporating three semantic categories, i.e., each semantic category contained three concrete nouns (Figure 1B). The task consisted of a total of twelve trials, broken down into three blocks, with four trials in each block. After completing a block, the participant was offered a break, to reduce any effects of boredom or fatigue that may negatively impact participants' overall performance on this task.

The trial structure corresponded to the non-clustered versions of the task used by Haarmann et al., (2003) and Nouwens et al., (2017). Participants listened to a nine-item list that was pseudo-randomised, a constraint being that items from the same category could not be presented consecutively. Participants then heard the question "*Which items were [semantic category]?*" Participants were instructed to try to recall the three to be remembered (TBR) items associated with that semantic category, in any order they wished. The cued stimulus category was also randomised for each participant. This ensured that participants had the opportunity to recall target memoranda from all serial positions, i.e., participants had four opportunities to recall a target memory item at each serial position.

Listening span task. Taken from Roome et al., (2019), and based on the integrated condition from Towse et al., (2010), participants heard a sentence and generated a semantically appropriate final word (Figure 1C). Once participants had completed all the sentences within a trial, they were required to recall their self-generated words in serial order. Within a trial, the task started with two sentences/TBR words, increasing incrementally up to five sentences/TBR words, creating four levels of difficulty. Participants had to complete three trials of each sentence/word length, which generated 12 trials overall, split into four blocks of testing. The experimenter recorded the number of correct items recalled and their serial position for each trial. The proportion of correct baseline recall was calculated for each participant by calculating the sum of correctly recalled items, in the correct serial position, divided by the total number of TBR items within the task. Recall of the correct item in the wrong serial position was marked as incorrect.

Figure 1: A schematic demonstration of the delayed cued recall conceptual span and complex span tasks.



Reading and Vocabulary Measures

York Assessment of Reading Comprehension (YARC, Snowling et al., 2009). All children read an age-appropriate reading passage and were asked eight questions about the information they had read. If they scored five or more, they read a second passage at a higher level. If they scored four or below participants read the passage at a lower age level. The children read their allocated stories out loud generating an accuracy score. Their answers to the questions generated a reading comprehension score.

British Picture Vocabulary Scale III (BPVS III,(Dunn & Dunn, 2009). Age-appropriate materials were used for each age group, using the set of items on the performance record that corresponded to the participants' age. The basal score was set when an individual made no more than one error in a set of 12 words. If an individual made more than one error, the experimenter administered the preceding set of items. Once the basal score was established, the experimenter continued with the test until a ceiling for an individual was determined. A ceiling score was set as eight or more errors within a set. A raw score was calculated by subtracting the number of the ceiling item (the last item in the ceiling set) by the total number of errors made by the individual.

Chronometry Analyses

Each correct report of sequences was coded in terms of three different response measures (Cowan et al., 1998, 2003); the preparatory interval (PI) inter-word pauses (IWPs) and word durations (WDs). The measurement of response was defined as follows: PI = the time period between the recall signal and the beginning of participant's recall; IWP = the time period between each recalled word; WD = The time taken to verbally articulate a memory item. Participants' performance on each of the memory tasks was audio recorded during the first testing session. The measurements were then extracted from the recordings using "Audacity" software (https://www.audacityteam.org). Each response time segment was checked through both the audio being listened to and the oscillography of the sound file. Response measures were only calculated for correctly recalled items, to ensure that the measured latencies reflected successful retrieval (Towse et al., 2010).

A second rater also recorded the three different response time measures for each task on 20 randomly selected participants across the age range (approximately 20% of the sample used for analysis). Table 1 shows the correlations between the two raters, with all measures showing high inter-rater reliability.

Table 1: The inter-rater correlations for each response time measure broken down by task.

	Conceptual Span Task			**Delayed Cued Recall Task**		**Complex Span Task**		
	PI	IWP	WD	PI	WD	PI	IWP	WD
Corr. Coefficient	.973***	.994***	.753***	.957***	.830***	.955***	.993***	.873***

Results

We focus in turn on three issues 1) the mean accuracy and response time across tasks; 2) the individual chronometric profiles of each task; and 3) individual differences in recall timing and the extent to which it predicts WMC and reading comprehension.

Description of key performance metrics

The accuracy and response time measures for all tasks are shown in Table 2. A 3 x 2 mixed factor ANOVA revealed 9-10yo recalled more items than 7-8yo, $F(1,101) = 12.976$, $p < .001$, $hp^2 = .114$. Further, the conceptual span and delayed cued recall tasks generated higher accuracy than the complex span task, $F(2,202) = 6.418$, $p < .001$, $hp^2 = .315$.

Both the PI and IWP responses did not show any effect of age. However, the PI duration for the complex span task was significantly longer than for the conceptual span task, $F(2,300) = 6.999$, $p = .001$, $hp^2 = .045$. This was the inverse for the IWP; the conceptual span task generates longer durations between the recall of items, $F(1,197) = 85.841$, $p <.001$, $hp^2 = .303$. The speed at which children articulated the recalled items showed a systematic decrease in duration from the conceptual span task, complex span task and delayed cued recall task, $F(2,303) = 155.815$, $p < .001$, $hp^2 = .507$. A task by age group interaction, $F(2,303) = 5.127$, $p < .006$, $hp^2 = 033$ revealed quicker WD in the 9-10 year olds for the conceptual span task, $F(1,102) = 19.802$, $p < .001$, $hp^2 = .163$.

These initial findings not only begin to show how difficult children find retrieving information outside of immediate access, reflected in their low task accuracy, but the response time patterns for each task are adaptive based on the task demands.

Table 2: The descriptive statistics of the response time and accuracy measures for each task.

	7-8yo		9-10yo	
Measure	M	SD	M	SD
Accuracy (Proportion correct)				
Complex span	.19	.13	.33	.17
Delayed cued recall	.35	.12	.38	.11
Conceptual span	.37	.12	.42	.12
Response Time: PI (s)				
Complex span PI	2.78	1.15	2.50	1.05
Delayed cued recall PI	2.29	1.27	2.33	1.88
Conceptual span PI	2.07	.81	1.92	.81
Response Time: IWP (s)				
Complex span IWP	1.29	1.00	1.36	1.03
Conceptual span IWP	3.80	2.06	2.99	1.94
Response Time: WD (s)				
Complex span WD	.61	.15	.56	.08
Delayed cued recall WD	.49	.09	.47	.08
Conceptual span WD	.82	.14	.70	.13

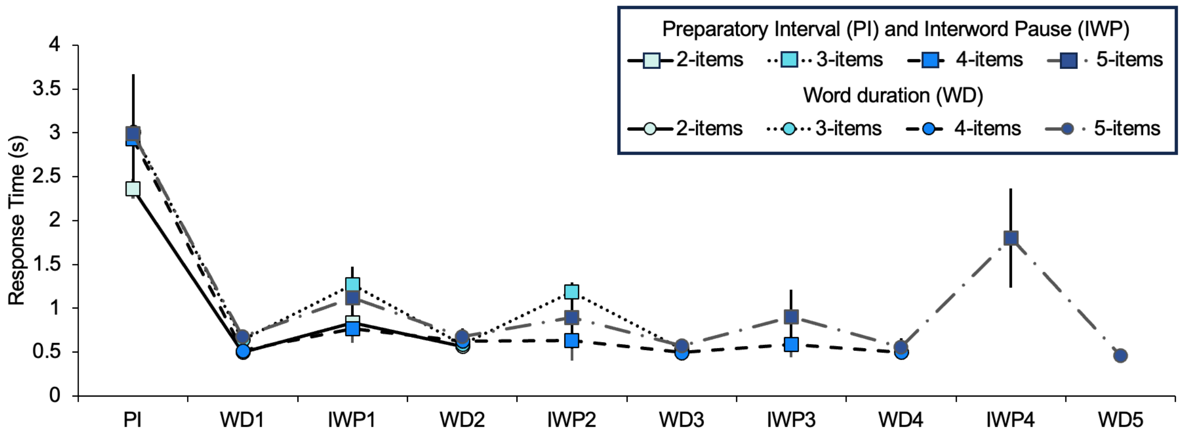
Patterns and regularities in response durations

Complex Span Task. As shown previously by Towse et al. (2010), when using an 'integrated' complex span task, the PI is much longer than the subsequent IWPs for 2- and 3-item list lengths. This version of the task incorporated lists of 2-5 items. Due to the differing levels of performance by participants, each response type was broken down by list length (Figure 2). To note, none of the response times showed an age effect.

Across all list-lengths, the PI was consistently longer than all IWPs (2-item: $F(1,95) = 183.598$, $p < .001$, $hp^2 = .659$; 3-item: $F(2,116) = 33.092$, $p < .001$, $hp^2 = .363$; 4-item: $F(3,90) =$

15.337, $p < .001$, $hp^2 = .338$; 5-item: $F(4,40) = 5.050$, $p = .002$, $hp^2 = .336$). Each list length showed subtly different WD profiles ((2-item: $F(1,93) = 4.438$, $p = .028$, $hp^2 = .046$, (3-item: $F(2,112) = 7.962$, $p < .001$, $hp^2 = .124$; (4-item: $F(3,78) = 4.756$, $p = .002$, $hp^2 = .155$; (5-item: $F(4,40) = 3.426$, $p = .017$, $hp^2 = .255$), showing both increased (2-items and 4-items) and decreased (3-items and 5-items) WDs as a function of position.

Figure 2: Mean time to recall correct sequences, broken down into the different phases of response-time analysis (the preparatory interval (PI), word duration (WD) and interword pause[s](IWP)) for the complex span task. The separate lines identify the different list lengths (2-5 items). Error bars indicate one standard error from the mean.

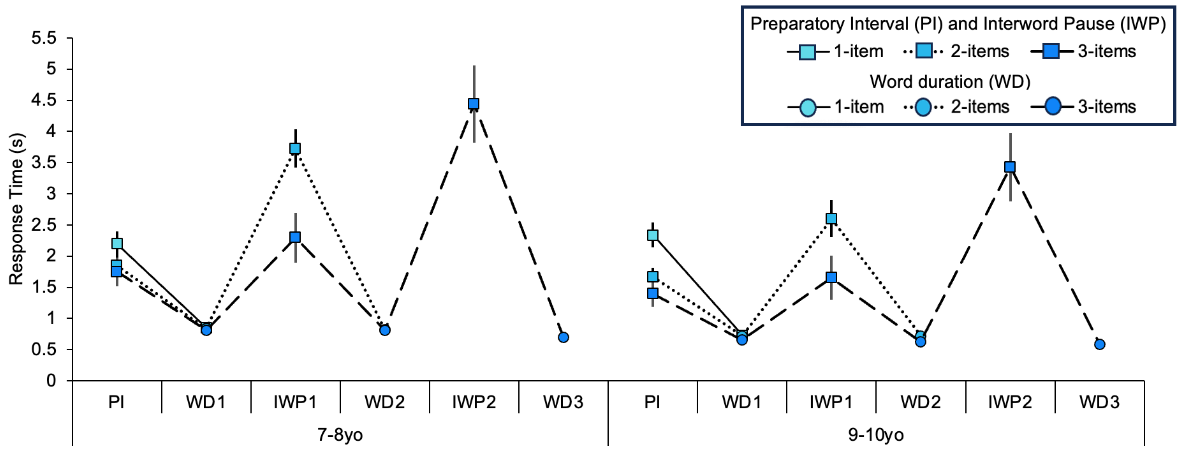


Delayed Cued Recall Task. This task required the recall of one item, and therefore only PI and WD were analysed. Both response types revealed equivalent durations for both age groups (PI: $F(1,102) = .013$, $p = .911$, $hp^2 = .000$; WD: $F(1,81) = 2.351$, $p = .129$, $hp^2 = .028$). The recall of the target item was equally distributed across three serial positions. Whilst the PI remained consistent across serial positions, ($F(2,162) = .011$, $p = .990$, $hp^2 = .000$; position 1: M=2.189s; SE=.334 ; position 2: M=2.153s; SE=0.196; position 3: M=2.139s; SE=.192), the WD for the first position was significantly slower than the third position, $F(2,162) = 35.737$, $p < .001$, $hp^2 = .306$ (position 1: M=0.543; SE =.014 ; position 2: M=0.442; SE=.012; position 3: M=0.436; SE=.010).

Conceptual span task. Participants attempted to recall three items on each trial. Due to the trial-by-trial response variability, we consider one-, two-, and three-item recall separately. For each list length, the PI and IWP showed no systematic age differences. As illustrated on Figure 3, the duration of the final item recalled is the longest. Focussing on two-item recall, the PI was quicker than the first IWP between the first and second spoken item, $F(1,94) = 44.042$, $p < .001$, $hp^2 = .319$. However, a significant interaction between the two factors revealed that whilst the PI showed no age effect, ($t(94) = .895$, $p = .373$), 9-10 year olds generated significantly faster IWP between the recall of the first and second items than 7-8yo ($t(96) = 2.463$, $p = .016$). The recall of three items revealed the IWP between items two and three, significantly slower than the prior PI and IWP, $F(2,104) = 22.987$, $p < .001$, $hp^2 = .307$

Figure 3 also shows participants word duration was quicker in the 9-10 year old group compared to 7-8 year olds for all list lengths (one-item: $F(1,101) = 10.280$, $p = .002$, $hp^2 = .092$; two-item ($F(1,93) = 11.111$, $p = .001$, $hp^2 = .107$; three-items ($F(1,51) = 11.322$, $p = .001$, $hp^2 = .182$). Whilst the WD did not differ when recalling two-items ($F(1,93) = .216$, $p = .643$, $hp^2 = .002$), the first item was verbalised significantly slower than the third item recalled ($p = .050$).

Figure 3: Mean time to recall correct sequences, broken down into the different phases of response-time analysis (the preparatory interval (PI), word duration (WD) and interword pause[s] (IWP)) for the conceptual span task. The separate lines identify the different number of items recalled (1-3 items). Error bars indicate one standard error from the mean.



Individual differences in performance

Inter-relations between the memory tasks

Table 3 shows the inter-relations of the chronometry and accuracy within and between tasks. The inter-correlations within tasks revealed no links between the chronometric measures (PI, IWP, WD) and accuracy measures on the conceptual span task or the delayed cued recall task. However, within the complex span task, all the chronometric measures significantly positively correlated with each other. Further, span performance on the complex span task was significantly negatively correlated with IWP. Thus, the shorter the pauses between items recall, the greater the span performance. The WD for each task significantly correlated with each other suggesting consistency in the WD duration across the three tasks (i.e., those that generated a longer WD for the complex span task, generated longer WDs on the delayed cued and conceptual span tasks).

Table 3: The inter-relations between the response timings and recall accuracy for each task.

Significance level *p < .05, **p < .01, ***p < .001

	Age	Conceptual Accuracy	Conceptual PI	Conceptual IWP	Conceptual WD	Complex Accuracy	Complex PI	Complex IWP	Complex WD	Delayed Accuracy	Delayed PI	Delayed WD
Age	1	.248*	-0.132	-.243*	-.435**	.499**	-0.071	-0.078	-0.115	0.202	-0.058	-0.107
Conceptual Acc.		1	-0.129	-0.030	-0.070	.501**	-0.021	-0.117	.252*	.306**	-0.067	0.119
Conceptual PI			1	0.105	0.105	-.243*	0.078	0.152	0.162	0.009	0.049	0.114
Conceptual IWP				1	-0.008	-0.111	-0.067	0.056	-0.083	0.026	0.054	0.034
Conceptual WD					1	-0.176	0.125	0.051	.280**	0.079	-0.032	.220*
Complex Acc.						1	-0.191	-.263*	0.016	.303**	-.230*	-0.123
Complex PI							1	.361**	.243*	0.129	0.003	.213*
Complex IWP								1	.358**	-0.107	.317**	0.176
Complex WD									1	0.094	0.044	.355**
Delayed Acc.										1	-0.187	0.083
Delayed PI											1	0.183
Delayed WD												1

Does the participants chronometry and accuracy on the delayed cued recall and conceptual span tasks predict WMC?

Listening span is the most widely investigated measure of WMC compared with conceptual span and delayed cued recall. Therefore, we employed a hierarchical multiple linear regression to determine the extent to which the reconstructive retrieval processes of the delayed cued recall task and conceptual span task predict individual variation in children's WMC (N=98). Age and vocabulary were entered at the first step; the second level included the predictor variables conceptual span task accuracy and PI; and delayed cued recall accuracy and PI. Overall, the regression model explained 52.3% of variance in children's WMC, Adjusted $R^2 = .523$, $F(6,91) = 18.757$, $p < .001$, revealing children's vocabulary, accuracy and PI on the conceptual span task as significant predictors (Table 4).

Table 4: Hierarchical regression for predicting variation in working memory capacity (WMC).

	Predictors	B	Std. Error	β	t	p
1	Age	0.002	0.001	0.150	1.631	0.106
	Vocabulary	**0.006**	**0.001**	**0.572**	**6.221**	**0.000**
2	Age	0.002	0.001	0.124	1.460	0.148
	Vocabulary	**0.005**	**0.001**	**0.443**	**4.849**	**0.000**
	Conceptual Acc.	**0.331**	**0.109**	**0.243**	**3.048**	**0.003**
	Conceptual PI	**-0.032**	**0.015**	**-0.155**	**-2.168**	**0.033**
	Delayed PI	-0.015	0.008	-0.139	-1.942	0.055
	Delayed Acc.	0.057	0.110	0.040	0.523	0.602

Taking the unique contributions of vocabulary, accuracy and PI measures from conceptual span, it was calculated that out of the 52.3% of the total variance, the significant predictor variables accounted for 26.6% of that variance. Variance partitioning (Salthouse, 1994) further showed that all three measures significantly predicted listening span (Figure 4A). Thus, children's success with cued retrieval uniquely predicts their listening span, as does the length of the pause before word recall starts.

Does participants chronometry across the different tasks predict reading comprehension?

Reading comprehension as a complex cognition has been historically linked to working memory (e.g., Case et al., 1982; Daneman & Carpenter, 1980). We therefore examined whether the different working memory measures could explain individual differences in children's reading comprehension performance (N=91). We employed another hierarchical multiple regression analysis with reading comprehension as the dependent variable. At the first step, age, vocabulary and reading ability were entered; at the final level, all predictor variables that correlated with reading comprehension in Table 3 were entered.

Overall, the model explained 67.4% of variances in children's reading comprehension, Adjusted $R^2 = .674$, $F(9,81) = 21.694$, $p < .001$. At the final stage of the model, children's vocabulary and the PI for the conceptual span task uniquely predicted reading comprehension (Table 5). This suggests that's children's efficiency in accessing and retrieving semantically associated information is predictive of their reading comprehension.

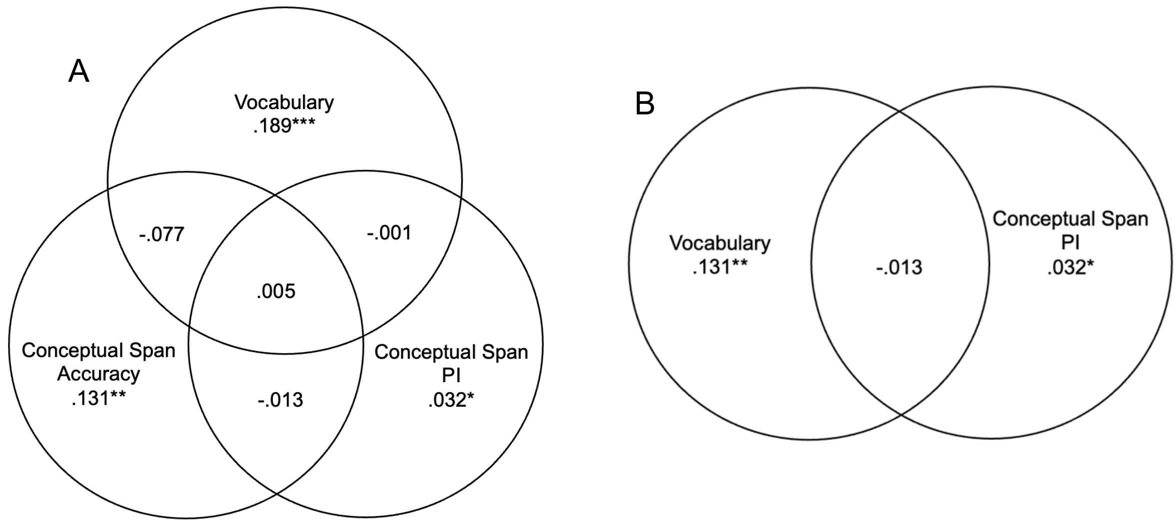
Table 5: Hierarchical regression for predicting variation in reading comprehension.

	Predictors	B	Std. Error	β	t	p
1	Age	0.047	0.058	0.064	0.816	0.417
	Vocabulary	**0.304**	**0.041**	**0.652**	**7.425**	**0.000**
	Reading Ability	**0.168**	**0.080**	**0.167**	**2.108**	**0.038**
2	Age	0.012	0.060	0.016	0.196	0.845
	Vocabulary	**0.261**	**0.044**	**0.559**	**5.982**	**0.000**
	Reading Ability	0.110	0.077	0.109	1.426	0.158
	Conceptual Acc.	4.762	4.460	0.077	1.068	0.289
	Conceptual PI	**-1.597**	**0.618**	**-0.162**	**-2.585**	**0.012**
	Conceptual WD	-2.069	3.639	-0.039	-0.569	0.571
	Delayed Acc.	4.166	4.378	0.062	0.951	0.344
	Complex IWP	-0.705	0.521	-0.087	-1.351	0.180

Complex Acc.	5.676	4.279	0.118	1.326	0.188

Out of the total variance accounted for in the previous model, the vocabulary and conceptual PI accounted for 14.3% of the variance, $\Delta R^2 = .143$, $F(1,81) = 35.783$, $p < .001$. Variance partitioning showed that out of the 14.3% variance, both variables significantly and uniquely contribute to reading comprehension above and beyond all variables accounted for (Figure 4B).

Figure 4: Venn diagrams representing A: the shared and unique variance accounted for by vocabulary, accuracy and preparatory interval (PI) gained from the conceptual span task that predicts variation in WMC from the complex span task; B: the shared and unique variance accounted for by vocabulary, and preparatory interval (PI) gained from the conceptual span task that predicts variation reading comprehension.



Discussion

A key objective in this paper is to investigate how children's reconstructive retrieval processes within working memory can be identified through patterns of recall chronometry. To measure working memory, historically there has been an emphasis on accuracy. Whilst this is suited to theories about capacity, it is potentially less naturally attuned to illuminating underlying processes. Chronometry provides the opportunity to focus, via time courses, on the processes involved in different types of memory tasks. This provides an important insight into the nuances of the retrieval process during development. Further, we can show how the use of storage-only tasks that require contextual cue-dependent search and retrieval mechanisms predicts the development of WMC and reading comprehension.

Complex span tasks have been central for advancing research into individual variation in WMC (Conway et al., 2007). The interleaved nature of the paradigm elements creates the environment in which representations can degrade as well as be refreshed (Barrouillet & Camos, 2020) and because memoranda are sometimes the product of the processing event, there are a range of processing cues to support recall. To understand the various processes better, we administered two additional tasks: delayed cued recall and conceptual span tasks. They lack a concurrent processing component, yet they involve cueing and selection processes. Children’s ability to retrieve memory items was below 50% across all tasks (range 19.3% – 42.1%). The relatively low performance on these tasks converges with the low SM estimates previously shown in children (De Alwis et al., 2009; Gibson et al., 2009; Roome et al., 2014; Roome et al., 2019). Yet the significant positive correlations across all three tasks is clearly consistent with the proposition that cued search forms a meaningful common attribute.

Studying the chronometry of these tasks complements the understanding of the contextual cue-dependent search and retrieval process. All three tasks show that accurate recall is time consuming and supports recent work in adults (Towse et al., 2008) and children (Cowan et al., 2003; Towse et al., 2010). The PI extended over 2s, whilst an IWP could last up to 3s as children accessed and redintegrate degraded target memorandum. This contrasts with short-term memory span tasks, of which show much quicker – up to an order of magnitude – PI and IWPs in comparison to complex span tasks (Cowan, 1992; Cowan et al., 1994, 1998, 2003; Hulme et al., 1999).

Each task can also be differentiated by its chronometric profile. The simplest task was the delayed cued recall task, which required children to recall a cued single item after a 15s delay. This took participants around 2s to recall correctly; a much longer period than recorded for short-term memory tasks. Our version of the complex span task showed children generated PIs consistent in length with the delayed cued recall task, but longer than the conceptual span task. This task then required the addition of interleaving the processing and storage elements of the task (similar to Towse et al., 2008, 2010) to facilitate the use of semantic representations to retrieve memoranda (Osaka et al., 2002; Tanaka, Teppei et al., 2014). Here, we observed that children generated subsequently shorter IWPs, which was consistent across all list lengths. The increased PI duration is consistent with the reconstruction hypothesis in that the extra context provided by sentences in the task is used to help internally reconstruct the list of final words (Towse et al., 2008, 2010). Further, due to the duration differences between the PI and IWPs, it is possible that children use the PI to internally prepare the recall of *all* the items they can access, for that trial.

In contrast, after a cue is provided on the conceptual span task, children can recall the first item relatively quickly, reflective of the shorter PI in comparison to the complex span task, but still equivalent to the delayed cued recall task. This is likely due to semantic cues benefitting memory performance as it draws upon pre-existing semantic associations in long-term memory (Howard & Kahana, 2002). However, an additional complexity to this task relative to the delayed cued recall task, is the requirement to retrieve up to three items. With each item recalled, the duration of the IWP increased. This may specifically the increasing demand to redintegrate each cue-associated memory item individually. Thus, as demonstrated here, the nuances of recall timing observed across the different tasks potentially allude to the developmental improvement in the use of reconstructive contextual cues. But, more importantly, suggest that we must consider the cognitive demands in action on different working memory tasks that may quicken or slow down the process, and underlie processing speed.

Further, individual differences in response time correlated with children's overall WMC and reading comprehension. Specifically, participants accuracy and their PI on the conceptual span task was predictive of WMC, and the same PI uniquely predicted of individual variation in reading comprehension above and beyond all other measures. PI extracted from the delayed cued recall task has previously been linked to individual differences in children's WMC (Roome et al., 2019), and children's PI on complex span performance is predictive of scholastic aptitude (Cowan et al., 2003). Response times may reflect not just retrieval speeds (Kail & Salthouse, 1994), but the efficiency to organise memories, plan responses and retrieve target memoranda (Cowan et al., 2003). Thus, it may be these processes that heavily weigh on the access and retrieval of words and their meanings that significantly contribute to children's working memory capacity, as well as understanding of text in this case.

With the current data, vocabulary played a large role with respect to mediating WMC and children's reading comprehension ability. Automatic access to long-term vocabulary knowledge is key to be successful at this type of complex cognition. Based on the nature of our tasks, alongside the expected role of vocabulary in reading comprehension (Nouwens et al., 2017; Oakhill & Cain, 2012), it is not surprising to find vocabulary's role when contributing to semantically integrated listening span tasks and storage-only tasks that are reliant on the participants semantic understanding superordinate categories. Therefore, with hindsight, the use of vocabulary data as a predictor could be complemented with other measures such as semantic fluency or semantic inference (e.g., the anomaly judgment task used by Haarmann et al., (2003).

In summary, this current work supports and extends the hypothesis that working memory draws on search and reconstruction for recall. Across all three tasks it is evident that memory items are not all kept in a highly accessible state. Working memory recall is not spontaneous and requires time for search and production of the target memory. Data also provide converging evidence that participants are using more than just the memory item at recall, and that there is a cooperative relationship between processing and memory activities. This provides a further dimension to our understanding of the development of the working memory and highlights the need to consider the contributions of long-term retrieval mechanisms to WMC.

References

Atkinson, A. L., Allen, R. J., Baddeley, A. D., Hitch, G. J., & Waterman, A. H. (2021). Can valuable information be prioritized in verbal working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*, 747–764. https://doi.org/10.1037/xlm0000979

Baddeley, A. (1986). *Working Memory*. Oxford, England: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Barrouillet, P., & Camos, V. (2020). The Time-Based Resource-Sharing Model of Working Memory. In R. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory: The state of the science* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780198842286.003.0004

Brown, J. (1958). Some Tests of the Decay Theory of Immediate Memory. *Quarterly Journal of Experimental Psychology*, *10*(1), 12–21. https://doi.org/10.1080/17470215808416249

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386–404. https://doi.org/10.1016/0022-0965(82)90054-6

Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (2007). *Variation in working memory*. New York: Oxford University Press.

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language*, *31*(5), 668–684. https://doi.org/10.1016/0749-596X(92)90034-U

Cowan, N., Bao, C., Bishop-Chrzanowski, B. M., Costa, A. N., Greene, N. R., Guitard, D., Li, C., Musich, M. L., & Ünal, Z. E. (2024). The Relation Between Attention and Memory. *Annual Review of Psychology*, *75*(Volume 75, 2024), 183–214. https://doi.org/10.1146/annurev-psych-040723-012736

Cowan, N., & Elliott, E. M. (2023). Deconfounding serial recall: Response timing and the overarching role of grouping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(2), 249–268. https://doi.org/10.1037/xlm0001157

Cowan, N., Keller, T. A., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing clues to the mechanisms underlying age and word length effects. *Journal of Memory and Language*, *33*, 234–250. https://doi.org/10.1006/jmla.1994.1012.

Cowan, N., Towse, J. N., Hamilton, Z., Saults, 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

Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D., & Keller, C. V. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General*, *127*, 141–160. https://doi.org/10.1037/0096-3445.127.2.141

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433. https://doi.org/10.3758/BF03214546

De Alwis, D., Myerson, J., Hershey, T., & Hale, S. (2009). Children’s higher order cognitive abilities and the development of secondary memory. *Psychonomic Bulletin & Review*, *16*(5), 925–930. https://doi.org/10.3758/PBR.16.5.925

Dunn, L., M., & Dunn, D., M. (2009). *British Picture Vocabulary Scale—3rd Edition*. London: GL Assessment Ltd.

Gibson, B. S., Gondoli, D. M., Flies, A. C., Dobrzenski, B. A., & Unsworth, N. (2009). Application of the Dual-Component Model of Working Memory to ADHD. *Child Neuropsychology*, *16*(1), 60–79. https://doi.org/10.1080/09297040903146958

Haarmann, H. J., Davelaar, E. J., & Usher, M. (2003). Individual differences in semantic short-term memory capacity and reading comprehension. *Journal of Memory and Language*, *48*(2), 320–345. https://doi.org/10.1016/S0749-596X(02)00506-5

Hall, D., Jarrold, C., Towse, J. N., & Zarandi, A. L. (2015). The developmental influence of primary memory capacity on working memory and academic achievement. *Developmental Psychology*, *51*(8), 1131–1147. https://doi.org/10.1037/a0039464

Healey, M. K., & Miyake, A. (2009). The role of attention during retrieval in working-memory span: A dual-task study. *Quarterly Journal of Experimental Psychology*, *62*(4), 733–745. https://doi.org/10.1080/17470210802229005

Howard, M. W., & Kahana, M. J. (2002). When Does Semantic Similarity Help Episodic Retrieval? *Journal of Memory and Language*, *46*(1), 85–98. https://doi.org/10.1006/jmla.2001.2798

Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999). Think before you speak: Pauses, memory search, and trace redintegration processes in verbal memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 447–463. https://doi.org/10.1037/0278-7393.25.2.447

Hurlstone, M. J. (2024). Serial recall. Invited book chapter contribution to Kahana, M., and Wagner, A. *The Oxford Handbook on Human Memory* (Oxford University Press)

Hurlstone, M. J. (2019). Functional similarities and differences between the coding of positional information in verbal and spatial short-term order memory. *Memory*, *27*(2), 147–162. https://doi.org/10.1080/09658211.2018.1495235

Jarrold, C., Hall, D., Harvey, C. E., Tam, H., Towse, J. N., & Zarandi, A. L. (2015). What can we learn about immediate memory from the development of children's free recall? *Quarterly Journal of Experimental Psychology*, *68*(9), 1871–1894. https://doi.org/10.1080/17470218.2014.995110

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 336–358. https://doi.org/10.1037/0278-7393.26.2.336

Logie, R. H. (2023). Strategies, debates, and adversarial collaboration in working memory: The 51st Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, *76*(11), 2431–2460. https://doi.org/10.1177/17470218231194037

Miyake, A., & Friedman, N. P. (2004). *Where does the predictive power of the reading span test come from?* 45th Annual Meeting of the Psychonomic Society, Minneapolis.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of Acquisition Norms for a Large Set of Object Names and Their Relation to Adult Estimates and Other Variables. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 528–559. https://doi.org/10.1080/027249897392017

Nouwens, S., Groen, M. A., & Verhoeven, L. (2017). How working memory relates to children's reading comprehension: The importance of domain-specificity in storage and processing. *Reading and Writing*, *30*(1), 105–120. https://doi.org/10.1007/s11145-016-9665-5

Oakhill, J. V., & Cain, K. (2012). The Precursors of Reading Ability in Young Readers: Evidence From a Four-Year Longitudinal Study. *Scientific Studies of Reading*, *16*(2), 91–121. https://doi.org/10.1080/10888438.2010.529219

Osaka, M., Nishizaki, Y., Komori, M., & Osaka, N. (2002). Effect of focus on verbal working memory: Critical role of the focus word in reading. *Memory & Cognition*, *30*(4), 562–571. https://doi.org/10.3758/BF03194957

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193–198. https://doi.org/10.1037/h0049234

Pierce, J. W. (2007). Psychopy—Psychophysics software in python. *Journal of Neuroscientific Methods*, *162*(1–2), 8–13.

Pierce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics*, *2*, 10.

Roome, H. E., Towse, J. N., & Crespo-Llado, M. M. (2019). Contextual support for children's recall within working memory. *Quarterly Journal of Experimental Psychology*, *72*(6), 1364–1378. https://doi.org/10.1177/1747021818804440

Roome, H., Towse, J., & Jarrold, C. (2014). How do selective attentional processes contribute to maintenance and recall in children's working memory capacity? *Frontiers in Human Neuroscience*, *8*. https://www.frontiersin.org/articles/10.3389/fnhum.2014.01011

Salthouse, T. A. (1994). How Many Causes Are There of Aging-Related Decrements in Cognitive Functioning? *Developmental Review*, *14*(4), 413–437. https://doi.org/10.1006/drev.1994.1016

Snowling, M., Stothard, S., E., Clarke, P., Bowyer-Crane, C., A., Harrington, A., Truelove, E., & Hulme, C. (2009). *York Assessment of Reading Comprehension*. London: GL Assessment Ltd.

Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence*, *21*(1), 83–108. https://doi.org/10.1016/0160-2896(95)90040-3

Tanaka, Teppei, Saito, Satoru, & Kikuchi, Satoru. (2014). The role of sentence information in reading span performance: An examination of the recall reconstruction hypothesis. *Psychologia*, *57*, 164–176. https://doi.org/10.2117/psysoc.2014.164

Towse, J. N., Cowan, N., Hitch, G. J., & Horton, N. J. (2008). The Recall of Information from Working Memory: Insights from Behavioural and Chronometric Perspectives. *Experimental Psychology*, *55*(6), 371–383. https://doi.org/10.1027/1618-3169.55.6.371

Towse, J. N., Hitch, G. J., Horton, N., & Harvey, K. (2010). Synergies between processing and memory in children's reading span: Synergies between processing and memory in reading span. *Developmental Science*, *13*(5), 779–789. https://doi.org/10.1111/j.1467-7687.2009.00929.x

Towse, J. N., Hutton, U., & Hitch, G. J. (1998). Grass is coloured...red? Further sentence completion norms for children during a working memory reading task. *Technical Report CDRG3*, 1–26.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

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

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*(3), 289–335. https://doi.org/10.1016/j.jml.2003.10.003

Woltz, D. J., & Was, C. A. (2006). Availability of related long-term memory during and after attention focus in working memory. *Memory & Cognition*, *34*(3), 668–684. https://doi.org/10.3758/BF03193587

Yuill, N., Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology*, *80*(3), 351–361. https://doi.org/10.1111/j.2044-8295.1989.tb02325.x