Task experience and children's working memory performance: A perspective from recall timing

Towse, John N. (Department of Psychology, Lancaster University)

Cowan, Nelson (Department of Psychological Sciences, University of Missouri)

Horton, Neil J. (Department of Psychology, Lancaster University)

Whytock, Shealagh (Department of Psychology, Lancaster University)

To appear in: Developmental Psychology.

Address for correspondence

Dr John N Towse

Department of Psychology

Fylde College, Lancaster University

Bailrigg, Lancaster LA1 4YF, United Kingdom

Tel: +44-(0)1524-593705 Fax: +44-(0)1524-593744

Email: j.towse@lancaster.ac.uk

Acknowledgements

We are grateful to the staff and pupils at the following schools for their cooperation in the research: Arkholme; Ellel St John's; Forton; Halton St. Wilfred's; Nether Kellet; Tatham Fells. The research was facilitated by the support of the ESRC (grant RES000220452 and RES000230859) and was written in part whilst the first author was a research visitor at the

University of Colorado at Boulder. The manuscript has benefited from reviewers' comments and discussions with Andrea Cheshire.

Task experience and children's working memory performance: A perspective from recall timing

Abstract

Working memory is an important theoretical construct among children, and measures of its capacity predict a range of cognitive skills and abilities. Data from 9- and 11-year-old children illustrate how a chronometric analysis of recall can complement and elaborate recall accuracy in advancing our understanding of working memory. A reading span task was completed by 130 children, 75 of whom were tested on two occasions, with sequence length either increasing or decreasing during test administration. Substantial pauses occur during participants' recall sequences and they represent consistent performance traits over time, whilst also varying with recall circumstances and task history. Recall pauses help to predict reading and number skills, alongside as well as separate from levels of recall accuracy. The task demands of working memory change as a function of task experience, with a combination of accuracy and response timing in novel task situations being the strongest predictor of cognitive attainment.

Key words: working memory, recall timing, development, practice, proactive interference

Task experience and children's working memory performance: A perspective from recall timing

Working memory refers to the dynamic interplay of systems responsible for the maintenance of transient representations, as well as their transformation into useful cognitive products. The concept of working memory occupies centre-stage in cognitive science as a psychological construct important in its own right (Baddeley, 1986; 1996; Cowan, 2005; Gathercole, 1999; Miyake & Shah, 1999) and a component embedded within a variety of real-world skills among adults and children. These skills encompass reading comprehension (Daneman & Carpenter, 1980; De Beni, Palladino, Pazzaglia & Cornoldi, 1998) early reading development (Leather & Henry, 1994), mathematics (Adams & Hitch, 1997; Hitch, Towse & Hutton, 2001) and indeed a broad range of school curriculum topics (Alloway, Gathercole, Willis & Adams, 2004; Alloway et al., 2005) in addition to more abstract, higher level cognitive functions (Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005). The present work contributes to our understanding of working memory development in two ways: by examining the effects of task experience on performance, and by examining response timing along with accuracy as a function of this experience. We describe the basis of these two contributions in turn.

Measurement of Working Memory and Task Experience

Reading span is probably the most widely used index of working memory capacity (for some exceptions, see Bayliss, Jarrold, Gunn & Baddeley, 2003; Cowan et al., 2005; Towse, Hitch, Hamilton, Peacock & Hutton, 2005; Turner & Engle, 1989). In the original form of reading span used by Daneman and Carpenter (1980), participants read a sequence of separate sentences and attempted to remember the final word from each. In tests for children, a popular approach is to ask children to read a set of incomplete sentences and to provide an appropriate word to complete each one, with these words forming the memoranda (e.g., Leather & Henry, 1994). This gives the processing task a purpose (working out what the missing words are) and usually keeps children focused on reading for meaning.

Single administrations of reading span tests among adults sometimes generate only modest levels of test-retest reliability over varying intervals (Waters & Caplan, 1996). Among 8- to 11-year-old children, Hitch et al. (2001) reported that an *averaged* score from two tests correlated extremely well across a 12-month interval (r=.71), although the correlation between performance on single tasks at each time point was lower (see also Towse et al., 2005). This shows that although performance can vary from test to test, it is possible to obtain stable measures of working memory and that these are reliably associated with cognitive ability. However, surprisingly little is known about what contributes to stability and change in children's working memory.

Thus, in the present work we ask the fundamental question: what happens to working memory processes as children gain experience with reading span trials? In particular, we consider whether the ability to orchestrate performance on a *new* complex task is especially relevant to the predictive ability of working memory (Ackerman, 1988). For example, it has been proposed that the deployment of controlled processes - those dealing with novel situations - are central to working memory task performance (Engle, Tuholski, Laughlin & Conway, 1999; Kane & Engle, 2002). Even brief exposure to and practice on reading span trials may allow children to develop or modify strategies and procedures that help them accomplish the task. Simultaneously, proactive interference can emerge with practice; previous trial episodes can persist and disrupt performance on the current trial.

The present study examines how task experience affects working memory performance in two important and potentially complementary ways. First is the comparison of recall performance from one testing occasion to another (although some children completed a single test only at the second assessment epoch so as to control for any developmental maturation, i.e., improvements attributable merely to being tested at a later point in the study). To our knowledge, this represents the first systematic analysis of the reliability and consistency of recall timing in a complex working memory span task. This unique aspect of our study permits an assessment of the appropriate use of chronometric analyses in experimental and developmental research.

The second facet of task experience focuses on the more proximal or local impact of some trial sets on others. Previous research has clearly established that response duration in both immediate serial recall and complex span varies with list length; the first interword pause increases when there are more subsequent words to recall in the sequence (e.g., Cowan et al., 1994; Cowan et al., 2003). This implies that interword pauses involve list-wise search processes. However, relevant data come from test administrations that use an incremental list length testing procedure; participants recall long sequences only after short sequences (contingent on recall success). Accordingly, characteristics of, for example, recall of lists with three items may be influenced by responses given from lists with two items. From the simple perspective that practice aids performance, immediately preceding trials would be expected to make the recall process more efficient, as optimal strategies are implemented and refined (e.g., Whitney, Arnett, Driver & Budd, 2001). In contrast, from the perspective of inhibition-based accounts of working memory (e.g., Kail, 2002, Lustig, Hasher & May, 2001; see also Bunting, 2006; Chiappe, Hasher & Siegel, 2000; Hedden & Park, 2003), there is more opportunity for the build-up of proactive interference, with residual activation of recalled information from earlier trials contaminating production of subsequent items. The present paper evaluates the contribution of both practice and interference processes for reading span.

We addressed the independence of current trials from past history by either successively increasing or decreasing recall sequence lengths from an extreme starting point. This provides the opportunity to compare, for example, recall of items from a list with two words undertaken either as the first or final set of trials. One might consider this issue in general terms as an examination of massed practice on performance, in contrast to the spaced practice that is explored through the test-retest analysis.

We also used individual differences to address this issue. There is evidence particularly from adult research that the deployment of participants' task strategies can dampen the relationship between working memory and external measures of cognitive skill. For example, when the pacing of working memory trials is controlled by the participants not the experimenter, performance is less strongly associated with ability (Friedman & Miyake, 2004; see also Dunlosky & Kane, in press; Lepine, Barrouillet & Camos, 2005; Turley-Ames & Whitfield, 2003). To the extent that natural strategies will develop with task experience and children are often thought to be initially less sophisticated in their deployment of strategies, we investigate whether novel working memory trials have different properties from those presented once the child has acquired experience. For example, the aspect of working memory that correlates best with cognitive aptitude might be how quickly a satisfactory strategy for the task can be developed, or perhaps how well the task can be carried out before a strategy has developed.

Use of Timing Measures

Working memory span has been characterised as a “beguilingly simple outcome measure for a complex task” (Hitch et al., 2001). Accordingly, understanding working memory *per se* and characterising its relationship with other variables across development can be facilitated by considering multiple performance indices beyond the sheer number of items that can be simultaneously retained (Towse & Cowan, 2005; Towse et al., 2005). We suggest that the chronometry of recall represents one important source of evidence about memory representations. It forms an excellent ancillary measure because it can be derived from span trials; one does not require an additional task. Moreover, there is a growing body of research that studying recall dynamics offers an insight into memory and memory development that can be hard to obtain in other ways (e.g., Cowan et al., 1992; Cowan et al., 1998; Tehan & Lalor, 2000).

Cowan et al. (2003) showed how recall timing analyses could contribute to our understanding of reading span, reporting both similarities and differences between immediate serial recall and complex span paradigms. Interword pauses, that is the silent gaps between the articulation of each recall word, were much longer in reading and listening span tasks than commonly reported in short-term-memory measures. Preparatory intervals (the gaps before recall commences) were substantially longer too. Children were doing something different when it came to the assembly and production of an output sequence. On the other hand, recall times were quicker for counting span, an alternative working memory task, so it was the linguistic-based working memory measures that were particularly unusual in terms of output processes, rather than working memory span *per se*. Cowan et al. (2003) suggested that children may have been drawing on memories of the sentences they had read in order to help access and reconstruct the target recall answers (for evidence supporting this suggestion from adult data, see Towse, Cowan, Hitch & Horton, submitted). This position echoes other views about the overlap between sentence reading and recall (Copeland & Radvansky, 2001; Saito & Miyake, 2004). From this perspective, recall from reading span and listening span potentially involves more than memory search among activated candidate answers. It involves also the consideration of a diverse set of episodic (in the sense of verbatim or gist) information.

Cowan et al. also explored the relevance of recall timing variables for individual differences in working memory and cognitive ability. Overall response duration correlated with variance in recall ability, and recall processes were linked to wider achievement domains. The predictive value of recall timing is considered further in this paper, both at the level of the overall response and with respect to particular response components.

The preparatory interval is thought to involve processes relating to partial rehearsal, response planning and sequence preparation while word duration incorporates processes allied to articulation speed (see Cowan et al., 2003). The interword pauses necessarily reflect search through memory for the identification or specification of the next item to be recalled (although other processes are likely to be involved). Accordingly, interword pauses provide a more specific or focused measure of recall and item access than overall response length, though both measures are useful. Studies of recall timing in immediate serial recall have teased apart influences from preparatory intervals, interword pauses and word durations (Cowan, 1992; Cowan et al., 1998) supporting the contention that recall timing components reflect different memory process. See Table 1 for an overview of the dependent measures in this study.

Recall timing and task experience. Although Cowan et al. (2003) examined response timing in working memory tasks they did not examine effects of task experience. One reason to do so is to get a better understanding of the nature of the processes that change with task experience. Toward this end, we describe recall timing at the macroscopic level - response timing using the overall output duration that represents an amalgam of recall processes. We also consider particular recall timing at the microscopic level - the phases of recall that represent more specific sets of mental processes. Another important reason to examine response timing along with task experience is that timing measures may capture individual variance in responding to experience that eludes the working memory accuracy measure.

Method

Participants

We recruited 130 children who agreed to take part after parental consent had been obtained. Children attended a number of schools in the northwest of England, and there were 66 9-year-olds (M = 9 years 1 month, SD=3.62 months) and 64 11-year-olds (M=11 years 3 months, SD=3.46).

Apparatus, Stimuli and Procedure

Computer events were driven by an Apple Macintosh G4 ibook with 14-inch laptop screen (programmed using the “Revolution” language running under OS X) with response latencies measured in (1/60 s) ticks. Recordings were captured digitally on a minidisc player (Sony MZ-N710, with a Sony ECM-DS70P microphone).

Reading span. The experimenter provided a verbal overview of the task; children were asked to read aloud sentences on-screen and offer a suitable completion word. Following presentation of a set of sentences, each of the sentence completion words should then be recalled in serial order. As part of a practice phase, children were initially shown sentences to complete without any concurrent memory requirement.

Sentence completion words were mostly predictable and consistent across individuals (see Towse, Hamilton, Hitch & Hutton, 2000). The corpus was split into two equally-sized sets and counterbalanced in both test and retest situations. If a child produced a non-expected completion word, it was this item that they recalled.

Once the participant offered a completion word, the experimenter immediately tapped a computer key to initiate the next experimental event that followed after a 1 second interval (the keystroke also demarcated the completion of sentence reading). Participants were instructed to remain silent between the reading phases and to begin reading each sentence immediately. A visually-presented recall screen, contemporaneous with a brief auditory tone, cued children to report the memoranda in the appropriate order; a series of on-screen boxes signalled the appropriate number of responses. Children then received accuracy feedback on the sequence they had just produced.

For the children in the ascending sequence order condition (n=76), experimental trials commenced with three sets of two-sentence sequences. Provided at least one recall sequence was correct, three further trials were presented with the number of sentences in each trial increased by one, up to a maximum of five sentences. For the children in the descending sequence order condition (n=54), the first set of three trials comprised five-sentence sequences. Subsequent trial sets involved progressively shorter sequences down to the minimum of two sentences (unless children reached ceiling performance through correct recall of all three sequences before reaching this point). Most children were assigned the ascending condition so as to permit comparison with published findings that have used this format. In both conditions, children knew the list length prior to each trial.

All 130 children undertook the reading span test on at least one occasion. Seventy-five children completed two reading span assessments that differed only in the set of sentence stimuli, at time epochs t_1 and t_2, separated by approximately 10 weeks. Fifty-two children were tested at t_2 only. Three children were tested at t_1 only, being absent from school at t_2. Most children were tested twice because of the value of re-test data. Accordingly, it is possible to (a) examine session 1 performance (irrespective of exactly when that first assessment occurred), (b) compare, for a large subset of children, test and retest performance and (c) compare initial test against retest performance for the same time point (t_2). Condition (c) teases apart the impact of developmental change from any practice or experience-based change that may have occurred in (b).

Scholastic attainment. One-hundred and twenty-three children completed Word reading and Number Skills subscales from the British Abilities Scale II tests (BAS; Elliott, Smith & McCulloch, 1997); 7 children were unavailable for testing. Both tasks were completed individually at t_2, either before of after reading span assessment (varying with administrative convenience within the school timetable). The Number Skills test emphasizes written arithmetic. Children were encouraged to answer as many questions as they could, and their score signalled the total number of questions answered correctly. The Word Reading test involves the presentation of a card containing 90 words in ascending order of (normative) difficulty. Children read aloud as many words as possible, and the child's score represented the total number of words read correctly.

Results

Data Analysis

To achieve comparability in the increasing and decreasing order conditions, testing stopped when children reached floor or ceiling performance, respectively. Just as one typically assumes children at floor for a particular list-length will not recall longer sequences, children at ceiling in the decreasing order condition were assumed to recall shorter sequences correctly (this curtailment of trials affected only three children in session 1). Reading span was measured as the number of words recalled from completely correct sequences (see Conway et al., 2005, and Friedman & Miyake, 2005, for a discussion of different span scoring procedures).

From auditory computer files of all correct recall sequences, we segmented the speech waveform displays (using Sound Studio with Apple Macintosh OSX), co-referenced with the corresponding auditory signal, into contiguous intervals. In particular, we measured the length of the *preparatory interval*, the gap between the recall cue and the initiation of the response sequence, the *word duration* for each memorandum, and the *interword pause*, the length of the gap between words. The number of children with correctly recalled memory sequences is detailed in Table 2, which provides a general stratification of recall performance. Table 2 also specifies the number of children who provided correct sequences that were timed, which is necessarily smaller because (a) children occasionally produced extraneous non-recall words or re-started their recall sequence; (b) equipment failure led to the loss of data; (c) occasionally testing was terminated in the descending condition due to ceiling performance, as described above.

Two raters independently completed all timing measurements, after training on a different set of example sequences and with reference to common measurement guidelines (see Horton, Towse, & Cowan, 2007). In most cases (73% of responses), a single rater judged the timing of recall, while on the remaining occasions (i.e., 27% of responses) two raters examined the same file. A comparison of 94 sampled word and pause measurements showed that the two set of judgements were extremely closely correlated, $r(92)=.998$. This set of measurements contained 4 long intervals that affect the sample range (i.e. response outliers), but after excluding these values, the agreement between measurements was still very high indeed, $r(88)=.988$. Comparison of timings in absolute terms indicated close correspondence; mean pause lengths were within 20ms and word lengths 50ms of each other. Yet these differences were statistically significant, suggesting small biases or inconsistencies between raters in the location of word onsets / offsets. Accordingly, we ensured that the ratio in the number of timings used from each rater was approximately constant (1:3) across cells of the experimental design.

Correct recall times were then screened for outliers. For each recall time segment at each list-length, we examined the distribution of individual durations as z-scores. We set a conservative threshold of $z=3.29$; any larger values were curtailed back to this cut-off point (i.e., Winsorized). This affected 45 of 3352 durations (i.e., 1.3%) for lists with 2-4 words. Relevant trial data were then averaged together for subsequent analyses.

The following sections describe in turn the results for session one, for stability and change across sessions, and for correlations with aptitude measures. We draw upon data from the overall response duration, to provide a global measure of recall and to provide measures in line with analyses offered by Cowan et al. (2003). We also report data based on specific recall phases since they allow us to provide more focused accounts of recall processes. For all appropriate analyses, we report degrees of freedom adjusted for non-equal variances.

Accuracy and recall timing of reading span in session 1

We begin by considering children's data from their first assessment, as this provides the most direct point of comparison with previously published datasets. We will focus mainly on timing data but, first, Table 3 presents recall accuracy data. Reading span scores (measured as the number of words recalled from correct sequences) were 52.8% larger among older children $F(1,126)=19.4$, $p<.001$, $\eta_p^2=.134$. There was no reliable difference overall between ascending and descending sequence orders, $F<1$, $\eta_p^2=.001$, and so the results do not suggest a strong global effect from the build-up of proactive interference. This outcome is not an artefact of self-terminating test administration; even if one re-scored recall accuracy on the implausible assumption that children would fail to recall any of the non-presented easier lists in the descending sequence, reducing the scores for just a few children, the two sequence orders remain equivalent, $t(128)=.665$, $\eta^2=.003$. The interaction between age and sequence order was marginal, $F(1,126)=2.87$, $p=.093$, $\eta_p^2=.022$, but interpretation will be delayed until we report additional data from a second test session that clarifies this pattern of data.

Figure 1 shows the mean durations of recall components. Word durations are slightly longer but broadly similar to immediate serial recall data. However, as reported by Cowan et al., 2003, the preparatory intervals and interword pauses are considerably longer than is typically found for immediate serial recall. Appendix 1 provides detailed analysis of the characteristics of recall-timing including a comparison of sequences of different list length.

We focus on sequences with two items since they involve the most extreme contrast between ascending and descending sequences (being the first and last sets respectively) as well as yielding the greatest density of data. As indicated in Table 4, the length of recall was quicker for descending compared with ascending sequences, $F(1,114)= 13.1, p<.001, \eta_p^2=.103$ and 11-year-olds were quicker overall than 9-year-olds, $F(1,114)= 11.9, p<.001, \eta_p^2=.094$. Age and sequence order interacted, $F(1,114)= 6.80, p=.010, \eta_p^2=.056$ indicating younger children benefited most from a descending sequence.

Follow-up analyses showed that this pattern of main effects and interaction for the overall duration held true for the silent intervals in recall. With respect to the sequence order effect specifically, a descending sequence led to shorter interword pauses, $t(101.2)=3.19, p=.002, \eta^2=.091,$ preparatory intervals, $t(107.6)=2.90, p=.005, \eta^2=.0720,$ and initial word durations, $t(116)=2.22, p=.029, \eta^2=.040,$ although there was not a significant difference for the second word duration, $t(116)=1.15$, p=.251, $\eta^2=.011$. Thus, trial experience assisted multiple recall phases in terms of response speed, despite the absence of corresponding differences in recall accuracy.

Accuracy and recall timing from reading span as a function of task session

Evidence for stability. Among children who received two working memory assessments, both accuracy [$r(73)=.46, p<.001$] and the time taken to recall sequences with two items [$r(57)=.51, p<.001$] correlated across the two sessions (the accuracy correlation is very similar to other studies with a similar time interval; $r(54)=.47$; Towse et al., 2005, Expt. 2). Moreover, each specific recall component correlated across the two test sessions; preparatory intervals, $r(57)=.43$, the first word, $r(57)=.60$, the interword pause, $r(57)=.45$, and the second word, $r(57)=.44$, all $ps=<.001$. Intriguingly, the interword pause correlation between-sessions – where trials are separated by several weeks – is larger than that between the two item sequence pause and both the first interword pause and average interword pause with the three item sequence, where trials are separated by almost no time at all (r(40)=.13 and .25 respectively). This suggests that pauses at different list lengths might incorporate different processes. Figure 2 describes recall durations on the second session, and Table 3 compares accuracy across task session.

Evidence for change. Whilst the previous analyses establish performance stability, Table 3 and 4 also illustrate changes in, respectively, the accuracy and chronometry of working memory with experience. Analysis of variance on recall accuracy with session, sequence order and age as factors showed an increase in reading span from the first to the second session, $F(1,71)=6.18$, $p=.015$, $\eta_p^2=.080$. Since recall accuracy of children tested for the first time at t_1 and t_2 did not differ (M=9.21, SD=4.76 and M=9.45, SD=5.51, $t(128)=.27$, $\eta^2=.001$) this is not an effect of time of testing. Older children recalled more words, $F(1,71)=11.5$, $p=.001$, $\eta_p^2=.140$, but there was no overall effect of sequence order, $F(1,71)=1.19$, $p=.280$, $\eta_p^2=.016$.

These main effects were complemented by a significant 3-way interaction between age group, sequence order and test session, $F(1,71)=4.11$, $p=.046$, $\eta_p^2=.055$, which suggested an age-related proactive interference effect. With ascending sequences, 11-year-olds remembered more words than 9-year-olds at both the first and second session, $t(44)=4.29$, $p<.001$, $\eta^2=.295$, and $t(44)=2.50$, $p=.046$, $\eta^2=.124$, respectively. With descending sequences, the age difference was not reliable for either the first or second session, $t(27)=.71$, $\eta^2=.018$, and, $t(27)=1.58$, $\eta^2=.085$, respectively. Consequently, 9-year-olds initially recalled more words with descending sequences, but improved across sessions mostly with ascending sequences. In contrast

11-year-olds showed no initial advantage for descending sequences, but improved across sessions quite a bit on precisely those sequences. This suggests that among the older children, the benefits of practice can be offset by the impact of proactive interference, such that recall accuracy is facilitated by the descending sequence order in the second session. For younger children with some exposure to the task (i.e., at the second session) initial presentation of relatively easy sequences may have helped them optimise their performance.

Figures 1 and 2 describe the overall response duration for sequences with two items. Comparisons showed that recall was more rapid for younger children given descending sequences (M=3.82 s, SD=1.79, vs. M=2.14 s, SD=.45) yet older children were quicker with ascending sequences (M=2.12 s, SD=.24, vs. M=2.61 s, SD=1.10). This was confirmed by analysis of variance that yielded a marginal age effect [$F(1,55)=2.94$, $p=.092$, $\eta_p^2=.051$], non-significant sequence order and session effects [$F(1,55)=2.60$, $p=.113$, $\eta_p^2=.045$, and $F(1,55)=.243$, $p=.624$, $\eta_p^2=.004$ respectively], but a significant two-way interaction between age and sequence order, $F(1,55)=8.94$, $p=.004$, $\eta_p^2=.140$. This pattern reinforces the conclusion from accuracy data that practice (in this case, task experience within a session) may be particularly important for efficient performance among younger children. The three-way interaction was not significant, $F<1$, $p=.519$, $\eta_p^2=.008$.

Reading span and the prediction of cognitive ability

Among both children and adults, reading span is typically a reliable predictor of cognitive performance. We combined the two BAS sub-scores to obtain a measure of scholastic attainment, which correlated strongly with reading span accuracy, $r(121)=.67$, $p<.001$. Cowan et al. (2003) reported that the duration of children's working memory recall was relevant to ability and here the correlation between ability and recall duration in two item lists was also significant, albeit modest, $r(109)=-.29$, $p=.002$

Inspired by Chuah & Maybery (1999) and following Cowan et al. (2003), we used sets of regression analyses to calculate the unique and shared components of variance in BAS performance that could be accounted for by three variables; age, the first assessment of accuracy at reading span, and interword pauses from sequences with two recall items. Cowan et al. used BAS Number Skills as a variance mediator in predicting just Word Reading performance. We have included age instead because the current sample varies on this dimension, and we have aggregated BAS scores to form an ability construct as the target variable. Cowan et al. used response duration as the recall timing measure, whereas here we focus on the more specific interword pause component, so as to target memory search and word identification processes.

Figure 3 reports the partitioned variance associated with children's ability. Each contributed significant unique variance (i.e., all variables yield significant ΔR^2 values, ps<.01). The analyses are important insofar as they (1) confirm that reading span is a strong associate of scholastic ability, sharing 45% of variance; (2) indicate that recall pauses are significant predictors of ability in their own right, sharing 14% of variance; (3) demonstrate that variance in recall pauses also overlaps with reading span, in that 18% of all the variance common to both reading span and ability is linked with pause length variation (that is, 45% of variance common to both reading span and BAS scores includes an 8% component that is linked also to pauses (8 / 45=17.8%)); (4) show that there are age-related changes in scholastic ability distinct from working memory changes. In other words, development across age involves more than the development of memory and recall ability. Figure 4 summarises the corresponding pattern of relationships among children who contributed pause data from sequences with 2 *and* 3 items. This reveals an even greater unique contribution from the interword pause in predicting ability. Additional consideration of the reliability of timing measures is provided in Appendix 1.

Further individual-difference analysis involving BAS scores complements the experimental evidence that working memory processes change as a function of task experience. Among the subset of children who were tested twice, the first assessment of reading span accuracy correlated strongly with ability, $r(57)=.65$, $p<.001$, which mirrors the finding for the whole sample. The second assessment of reading span accuracy yielded a more modest, albeit still highly reliable, association with ability, $r(57)=.37$, $p<.001$. These two correlations are significantly different, $z=2.82$, $p=.005$ (following Steiger's, 1980, computational recommendations). Recall accuracy is most predictive of cognitive skills when children have not been extensively exposed to the task.

The correlations between overall response duration and ability also showed the same pattern across session ($r(57)=-.32$ and -.27, both ps<.05) but the difference was not significant, z=.38. Nonetheless, recall timing analysis with respect to sequence order provided evidence of a task experience effect. When two item trials were presented first (i.e., with ascending sequences), scholastic ability correlated with both the preparatory interval, $r(67)=-.30$, $p=.013$, and the interword pause, r(67)=-.40, $p=.001$. When two item trials were presented last (i.e., descending sequences) this correlation was not significant, $r(40)=-.01$ and -.09 respectively. The difference in the size of the correlations between sequence order was significant for the interword pause, $z=2.85$, $p=.004$, although not for the preparatory interval, $z=1.49$, $p=.14$. Once again experience can modulate what it is that reading span measures. Finally, in regression analyses of ability scores, after entering age and the interword pauses for lists with two items, the pauses in three-item lists still yielded significant additional variance (ΔR^2=.07, *p*=.005). After entering age and pauses in lists with three items, the interword pause for two-item lists also yielded significant additional variance (ΔR^2=.03, *p*=.041). These effects support the conclusion that experience and task configuration lead to the emergence of different skills, such that interword pauses at different list lengths can represent partially separable variables.

Discussion

The present study involves a rich dataset, yet one that has the power to illuminate a number of important interrelated issues. It examines the effects of task experience on working memory, using both accuracy and response timing measures. We broadly consider each component of the results in turn, before we introduce more general issues.

Accuracy and recall timing of reading span in session 1

Working memory as measured by recall accuracy is a stable and predictive index of complex cognition, and is clearly a multifaceted construct that can be complemented using chronometric analysis. Recall from reading span is an effortful process that is far more protracted than is commonly found with immediate serial recall or indeed non-language based working memory tasks (Cowan et al., 2003). Moreover, the extra time to produce a sequence is not principally a function of the recall words, but the pauses surrounding them. Children take a relatively long time to initiate recall (up to 50% of the recall period is occupied with the initial gap before sequence production) and there are often long pauses between words. Each of these interword pauses increases as a function of the number of recalled items, indicating that the demands of individual item access increase when part of a longer list.

The present experimental design, in which children either recalled short sequence lengths followed by longer ones, or *vice versa*, permits an examination of the interplay between effects of list length and task experience. Within each session the sequence order affects the experience accrued before encountering either easier or harder trials. Sequence order did not affect accuracy overall but made a difference to response duration. The data allow one to gauge the relative importance of the build-up of proactive interference from previous trials (Bunting, 2006; Lustig et al., 2001) versus practice effects. It is clear that the two-word lists were recalled more quickly in the descending sequence order, as one would expect from a practice effect. Of course, logically this effect could co-exist with effects of proactive interference (PI) in that span itself could depend on opposite factors: practice for any particular sequence length and the reduction of PI where the longest lists are concerned. Nevertheless, perhaps because of these counteracting factors, we found no significant advantage of the descending order on span.

This is certainly not to say that PI is unimportant in children's reading span. Indeed, younger children were at a disadvantage in the standard, ascending sequence order, but there were no age differences with descending sequences. This replicates findings in the aging literature (Lustig et al., 2001; see also Chiappe et al., 2000). Among adults, there is some suggestion that there may need to be quite a few trials at each list length for PI effects to be observed when manipulating sequence order (see Lustig & Hasher, 2002, footnote 2). Whilst one might expect that children would be particularly sensitive to PI effects, we recognise that pinpointing the strength and characteristics of PI in children's working memory would involve a large series of convergent studies.

To summarize, there appear to be two potential interacting effects of trial repetition; the beneficial effect of practice (discovering how to perform a complex task efficiently) and the detrimental effect of interference from prior trials (alongside interrelated phenomena such as fatigue). We propose that the balance of these factors, and their time course, change with development. The build up of PI may contribute to age differences insofar as these are reliable only for the ascending sequence length format. Practice helps younger children adapt to the incremental demands of an ascending sequence length, while facilitating older children's adaptation to trials that begin as being supra-span. There is a complex dynamic between effects that contribute to experience and developmental change.

Accuracy and recall timing from reading span as a function of task session

Recall times show consistency from one test assessment to another and, crucially, they can indicate changes in task performance that are not evident from accuracy measures. Yet, practice can lead to important changes in what working memory measures. Our results suggest that both 9-year-olds and 11-year-olds sometimes benefit from reducing PI (i.e., from being tested with trial sequences that gradually get shorter) but that the role of PI changes across sessions. There may be more PI in older children when the second session is carried out. This could explain why the descending sequence order led to greater accuracy for older children in the second session (as shown in Table 3). In contrast, younger children showed more of a need for practice. They showed an increase in accuracy across sessions in the ascending condition, perhaps because the practice helps them to acclimatise to task requirements. Recall durations were also long in the ascending condition, when the short lists that were timed did not have the benefit of practice within a trial.

Reading span and the prediction of cognitive ability

As Figures 3 and 4 illustrate, the data reaffirm the robust link in children between reading span accuracy and scholastic ability (e.g., Hitch et al., 2001; Ransdell & Hecht, 2003; Towse et al., 2005). This result is to be expected given the evidence that working memory correlates with a raft of adult cognitive skills (e.g., Kane & Engle, 2002). Through the use of recall timing our study adds two notable dimensions. First, some though certainly not all of the predictive power of reading span is shared with recall pauses, meaning that recall processes are relevant to accounts of the link between working memory and cognitive abilities. This demonstrates that theories of working memory can be enhanced through a greater understanding of recall processes in addition to encoding and maintenance operations (see also Cowan et al., 1998; Towse & Cowan, 2005; Unsworth & Engle, 2006). Second, recall pauses offer a significant independent source of variance in the prediction of ability. Pauses in children's recall are associated with scholastic attainment not merely because such pauses are linked to memory, but also because they reflect independent processes. This finding supports theoretical arguments that reading span recall can draw on reconstructive processes that involve representations from the processing events (i.e., the sentences being read) (Cowan et al., 2003; Saito & Miyake, 2004).

One of the drivers for research into complex working memory is the attempt to understand why the relevant tasks share substantial amounts of variance with other cognitive processes (see Conway, Jarrold, Kane, Miyake, & Towse, 2007). Recall timing variables partly mediate the relationship between reading span and scholastic attainment and make an additional independent contribution to explaining attainment. On the basis of the overall pattern of data, we argue that recall involves the important process of converting incomplete representations into a suitably ordered output sequence (incomplete in terms of content and/or order). This conclusion resonates with arguments from adult data that working memory span can involve both immediate and longer-term memory processes (Miyake & Friedman, 2004; Unsworth & Engle, 2006) such that the task reflects embedded processes within working memory (Cowan, 1999). The present results using data from children offer a converging form of evidence for this view that different representational sources are involved.

Prediction and practice across sessions. The systematic variance in the pause between recalled words underlines the contention that recall involves specific and coherent mental processes, including item reactivation in the absence of continuous item maintenance during the retention interval. The interword pauses in particular both share variance between recall accuracy and ability and contribute unique variance to ability over and above that of span. Thus, specific components of recall timing as well as overall durations are predictive as well as reliable. Preparatory intervals were in general a less sensitive performance index. We account for this in terms of the multiplicity of processes incorporated in this measure, including post-sentence processing, sequence rehearsal and construction of the first recall word. Moreover, analysis of sequence order and session comparisons showed working memory performance, both for accuracy and recall duration, is most closely related to ability before children have had very long to learn how to do the task. At this early point, it appears to reflect the ability to orchestrate a complex and unfamiliar task (see also Rabbitt, 1997).

Several theories propose that working memory is linked most strongly with ability when reading span trials are novel and relatively unpractised, and that the deployment of memorial strategies can dampen the link between working memory and external measures (Dunlosky & Kane, in press; Friedman & Miyake, 2004; Lepine et al., 2005; Turley-Ames & Whitfield, 2003). Yet, the present data are, we believe, unique in offering direct and convergent evidence for this view among children. One interpretation of the data is that control of attention is relevant to reading span (Engle et al., 1999), but also that controlled attention requirements are not fixed but diminish with practice (see also Ackerman, 1988). In addition, Cowan et al. (2005) argue that working memory tasks measure a core capacity characteristic of individuals, but only until other procedures or strategies (e.g., grouping, chunking, rehearsal, etc.) develop that relieve the burden on this capacity (see also Cowan, 2001). We hasten to add that the evidence for controlled attention or core capacity views does not exclude other processes from contributing to the characteristics of working memory performance. Our findings highlight the importance of recognising the complexity of reading span and its potential malleability.

These findings therefore have both theoretical and methodological importance. They emphasise that working memory performance is not a unidimensional trait; exposure to complex span trials can lead to learning and change in terms of how the task is accomplished and the initial novelty of the task contributes to its links wider cognitive skills. From a methodological standpoint, the data indicate that benefits from collecting additional trial data need to be balanced against the risk that the task may no longer measure quite the same skill. In addition, apparently subtle details of how trials are administered can affect various aspects of performance (e.g., the association between pauses and ability is weaker with a descending sequence presentation, while age differences in recall ability are larger with an ascending sequence). Moreover, if a set of working memory tasks are administered in a fixed order, the association between later measures and ability may under-represent the link between them.

A new understanding of reading span in children

The importance of the current dataset arises in part from the demonstration of how recall from reading span involves both stability and malleability. On the one hand, individual differences in the chronometry of recall show stability, through both a significant test-retest correlation and associations with external measures of cognitive attainment. Indeed the reliability of both overall response durations and specific interword pauses was at least equivalent to the reliability of recall accuracy. Yet on the other hand, the length of pauses are not immutable; they change with age and task experience, and the strength of individual differences are modified by task experience too. Therefore, these contrasting outcomes are not actually incompatible with each other – they instead reflect the rich nature of the reading span task.

Both the chronometry of reading span recall, as well as its accuracy, are flexible and sensitive to experience. The overall response duration of two-item sequences becomes shorter when they follow longer sequences, an effect that occurs specifically for the interword pause as well as the preparatory interval. Thus, recent experience allows recall to take place more efficiently (we specify ‘recent’ since there is no corresponding advantage for a second session taking place some time after the first). In addition, there are changes in the predictive strength of recall performance with experience. As the absolute levels of recall increase from the first to the second administration, the strength of association declines between external measures of ability and both reading span and pause length.

Analysis indicates that, on the one hand, children achieve more accurate recall in a second session without systematic changes to recall duration. On the other hand, children recall short sequences more rapidly when they have already been exposed to longer sequences, while accuracy does not change. These performance dissociations demonstrate that whilst accuracy and recall time measures can be linked conceptually and empirically, they can be shown to be partially independent too. Each measure can yield separate and complementary evidence for cognitive processes in children’s memory (see also Cowan et al., 2006). Moreover, pauses in lists of two and three items appear to capture different aspects of individual differences. One possible explanation for this finding is that the contribution of primary and secondary memory to recall differs for these two sequence lengths (e.g., Unsworth & Engle, 2006). In addition, pauses increase with sequence length (see appendix). This implies that recall involves memory search through potential recall answers.

We argue that the gains from data analysis justify the investment of effort into the examination of recall dynamics, particularly as it can reveal aspects of performance that are not evident in accuracy measures. Our approach allows for a more complete behavioural picture, with multiple dissociations in the patterns of performance across experimental variables. Moreover, the present data amply confirm that working memory is not merely about the maintenance of memoranda. It is also about their production at the point of recall and their recovery from incomplete representations. The timing of recall changes with age and shows both consistency and flexibility. In both respects, there is evidence for coherence, which can be used to increase our understanding of working memory in children and its development.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. Journal of Experimental Psychology: General, 117, 288-318.

Adams, J. W., & Hitch, G. J. (1997). Working memory and children's mental addition. Journal of Experimental Child Psychology, 67(1), 21-38.

Alloway, T. P., Gathercole, S. E., Adams, A., Willis, C., Eaglen, R., & Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. British Journal of Developmental Psychology, 23(3), 417-426.

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. (2004). A structural analysis of working memory and related cognitive skills in young children. Journal of Experimental Child Psychology, 87(2), 85-106.

Baddeley, A. (1996). Exploring the central executive. Quarterly Journal of Experimental Psychology, 49A(1), 5-28.

Baddeley, A. D. (1986). Working Memory. Oxford: Clarendon Press.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The Complexities of Complex Span: Explaining Individual Differences in Working Memory in Children and Adults. Journal of Experimental Psychology: General, 131(1), 71-92.

Bunting, M. (2006). Proactive interference and item similarity in working memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 32(2), 183-196.

Chiappe, P., Hasher, L., & Siegel, L. A. (2000). Working memory, inhibitory control and reading ability. Memory & Cognition, 28(1), 8-17.

Chuah, Y. M. L., & Maybery, M. T. (1999). Verbal and spatial short-term memory: common sources of developmental change? Journal of Experimental Child Psychology, 73, 7-44.

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

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. Psychonomic Bulletin & Review, 12(5), 769-786.

Copeland, D. E., & Radvansky, G. A. (2001). Phonological similarity in working memory. Memory & Cognition, 29(5), 774-776.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24(1), 87-185.

Cowan, N. (2005). Working memory capacity. Hove, U.K.: Psychology Press.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. Journal of Memory and Language, 31, 1-17.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. Cognitive Psychology, 51, 42-100.

Cowan, N., Elliott, E. M., Saults, J. S., Nugent, L. D., Bomb, P., & Hismjatullina, A. (2006). Rethinking Speed Theories of Cognitive Development. *Psychological Science, 17*(1), 67-73.

Cowan, N., Keller, T. A., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing cues to the mechanisms underlying age and word length effects. Journal of Memory and Language, 33, 234-250.

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.

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(2), 141-160.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

De Beni, R., Palladino, P., Pazzaglia, F., & Cornoldi, C. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. Quarterly Journal of Experimental Psychology, 51A(2), 305-320.

Elliot, C. D., Smith, P., & McCulloch, K. (1997). British Abilities Scale Second Edition (BAS II) Technical manual. Windsor, Berkshire: NFER-Nelson.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. Journal of Experimental Psychology: General, 128(3), 309-331.

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

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. Trends in Cognitive Sciences, 3(11), 410-419.

Haberlandt, K., Lawrence, H., Krohn, T., Bowe, K., & Thomas, J. G. (2005). Pauses and durations exhibit a serial position effect. Psychonomic Bulletin & Review, 12(1), 152-158.

Hedden, T., & Park, D. (2003). Contributions of source and inhibitory mechanisms to age-related retroactive interference in verbal working memory. Journal of Experimental Psychology: General, 132(1), 93-112.

Hitch, G. J., Towse, J. N., & Hutton, U. M. Z. (2001). What limits working memory span? Theoretical accounts and applications for scholastic development. Journal of Experimental Psychology: General, 130(2), 184-198.

Horton, N., Towse, J.N., & Cowan, N. (2007). *Response Timing in Working Memory: An Annotated Guide*. Retrieved 18 April 2007 from Lancaster University website: http://www.psych.lancs.ac.uk/research/TowseWM/.

Kail, R. (2002). Developmental change in proactive interference. Child Development, 73(6), 1703-1714.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in Working Memory capacity, executive attention, and general fluid intelligence: An individual difference perspective. Psychonomic Bulletin & Review, 9(4), 637-671.

Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. Journal of Experimental Child Psychology, 58, 88-111.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. Journal of Experimental Psychology: General, 130(2), 199-207.

Miyake, A., & Shah, P. (Eds.). (1999). Models of working memory. New York: Cambridge University Press.

Novick, M. R. (1966). The axioms and principal results of classical test theory. Journal of Mathematical Psychology(3), 1-18.

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), Methodology of frontal and executive function (pp. 1-38). Hove, East Sussex: Psychology Press.

Ransdell, S., & Hecht, S. A. (2003). Time and resource limits on working memory: Cross age consistency in counting span performance. Journal of Experimental Child Psychology, 86, 303-313.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing-storage relationship in reading span performance. Journal of Memory and Language, 50, 425-443.

Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. Intelligence, 27(3), 183-198.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. Psychological Bulletin, 87(2), 245-251.

Tehan, G., & Lalor, D. M. (2000). Individual differences in memory span: The contribution of rehearsal, access to lexical memory, and output speed. Quarterly Journal of Experimental Psychology, 53A(4), 1012-1038.

Towse, J. N., & Cowan, N. (2005). Working memory and its relevance for cognitive development. In W. Schneider, R. Schumann-Hengsteler & B. Sodian (Eds.), Young Children's Cognitive Development: Interrelationships among Executive Functioning, Working Memory, Verbal Ability, and Theory of Mind (pp. 9-37). Mahwah, NJ: Lawrence Erlbaum Associates.

Towse, J. N., Cowan, N., Hitch, G. J., & Horton, N. (submitted). The recall of information from working memory: insights from behavioural and chronometric perspectives. *Manuscript submitted for publication*.

Towse, J. N., Hamilton, Z., Hitch, G. J., & Hutton, U. (2000). Sentence completion norms among adults: A corpus of sentences differing in length (Technical Report No. CDRG 7): Royal Holloway, University of London.

Towse, J. N., Hitch, G. J., Hamilton, Z., Peacock, K., & Hutton, U. M. Z. (2005). Working memory period: the endurance of mental representations. Quarterly Journal of Experimental Psychology, 58A(3), 547-571.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? Journal of Memory and Language, 28, 127-154.

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

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. Quarterly Journal of Experimental Psychology, 49A(1), 51-70.

Whitney, P., Arnett, P. A., Driver, A., & Budd, D. (2001). Measuring central executive functioning: What's in a reading span? Brain and Cognition, 45, 1-14.

Table 1. Dependent variables and the processes that they are designed to measure.

Name	Measurement unit	Cognitive processes involved
Reading span	Recall accuracy (no. of serially ordered items)	Information retention alongside (& embedded within) ongoing processing activity
Preparatory intervals	Duration (seconds)	Response planning, sequence organisation and activation of the first recall item
Word duration	Duration (seconds)	Articulation speed
Interword pause	Duration (seconds)	Memory search and word reactivation, including redintegration and cue-based reconstruction

Table 2. Number of children credited with correct reading span sequence recall on first assessment.

Sequence length order	9-year-olds		11-year-olds	
	Ascending (n=37)	Descending (n=27)	Ascending (n=37)	Descending (n=27)
List-length-2	37 (37)	27 (20)	37 (37)	27 (24)
List-length-3	18 (14)	17 (12)	33 (29)	21 (21)
List-length-4	4 (3)	1 (1)	15 (14)	7 (7)
List-length-5	0 (0)	2 (2)	0 (0)	0 (0)

Note. Number of children with analysable sequences for recall timing in parentheses.

Table 3. Reading span scores, as the number of words recalled from correct sequences.

	First session	First of two sessions	Second of two sessions
Ascending sequence order			
9-year-olds	6.95 (4.83)	6.42 (3.99)	8.67 (3.64)
11-year-olds	12.0 (4.86)	11.9 (4.62)	11.5 (5.60)
Descending sequence order			
9-year-olds	8.04 (4.75)	9.00 (4.77)	9.94 (6.81)
11-year-olds	10.3 (4.05)	10.2 (3.69)	13.5 (5.08)

Note. Data from the first session involve 130 children, while 75 children completed two sessions, and their performance is reported at each assessment. Standard deviations in parentheses.

Table 4. Recall timing characteristics of correct list-length-2 sequences in children's first session.

Recall segment	PrepI	Wd1	Pause1	Wd2	Total duration
Recall segment timings (ascending sequence order)					
9-year-olds	2.49 (1.78)	.53 (.18)	.66 (.55)	.48 (.11)	4.16 (1.95)
11-year-olds	1.26 (.73)	.42 (.13)	.36 (.37)	.46 (.10)	2.50 (.90)
Recall segment timings (descending sequence order)					
9-year-olds	1.22 (1.56)	.45 (.15)	.31 (.14)	.47 (.10)	2.46 (1.63)
11-year-olds	1.11 (.70)	.39 (.09)	.31 (.21)	.42 (.11)	2.23 (.87)

Note. Standard deviations in parentheses. PrepI – preparatory interval, Wd1 = duration of first word, Pause1= interword pause between the first and second word, Wd2= duration of second word, Total Duration = Sum of all recall timing components.

Table 5. Recall time profile for children who successfully recalled 4-item sequences in reading span. Standard deviations in parentheses.

	Ascending sequence			Descending sequence		
	List-length 2	List length 3	List length 4	List-length 2	List length 3	List length 4
Preparatory intervals	1.213 (.509)	1.448 (.823)	2.135 (3.000)	1.345 (.999)	2.484 (2.461)	1.498 (1.518)
Interword pauses	.300 (.179)	.547 (.556)	.941 (.995)	.341 (.342)	1.051 (1.164)	.953 (.565)
Word durations	.469 (.095)	.511 (.123)	.557 (.160)	.431 (.087)	.539 (.134)	.560 (.170)

Figure 1. The duration of correct recall sequences as a function of list-length and sequence administration order, for the first assessment of reading span.

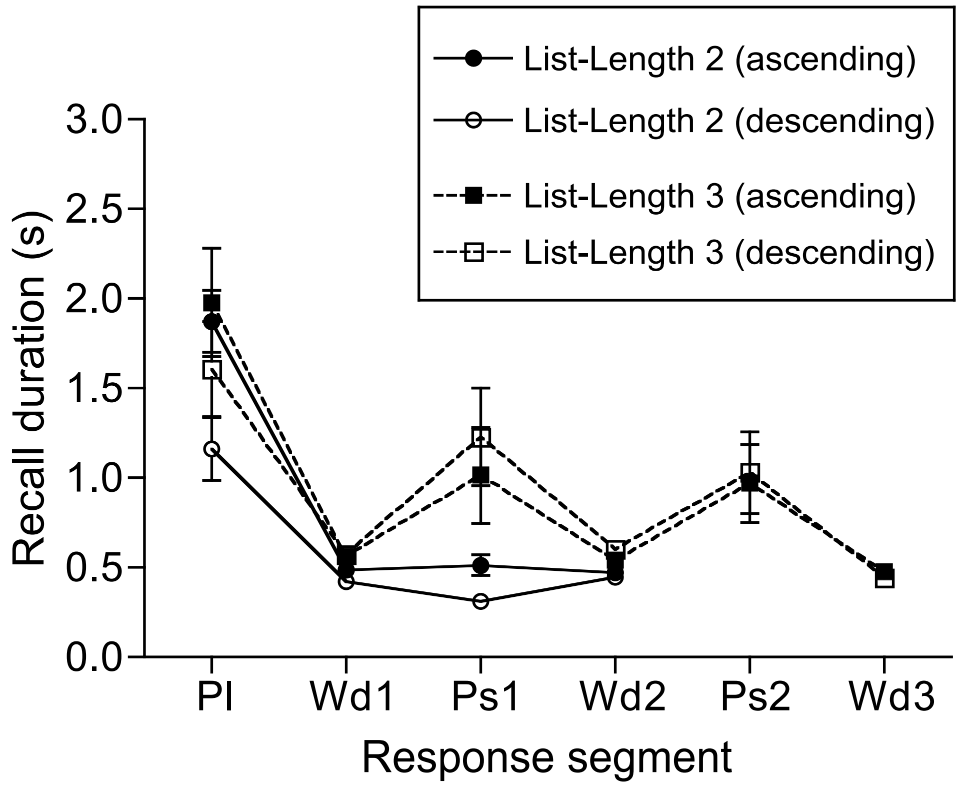


Figure 2. The duration of correct recall sequences as a function of list length and sequence administration order, for the second assessment of reading span.

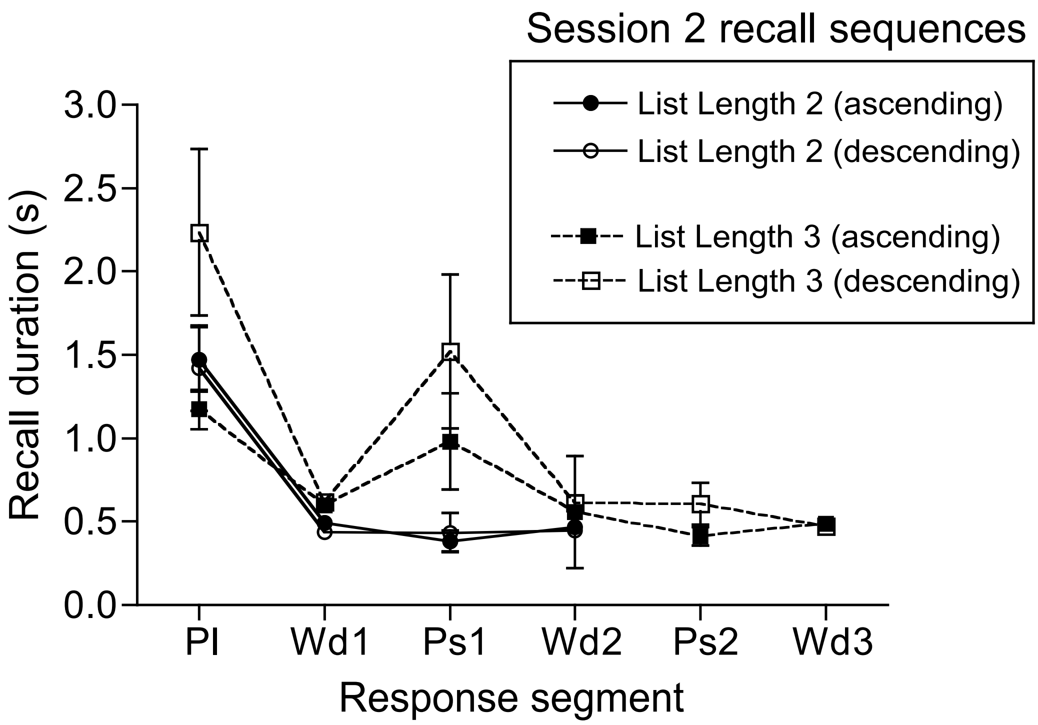


Figure 3. Schematic representation of the unique and common variance shared with the criterion skill measure from BAS performance, using children with list-length 2 recall pause data.

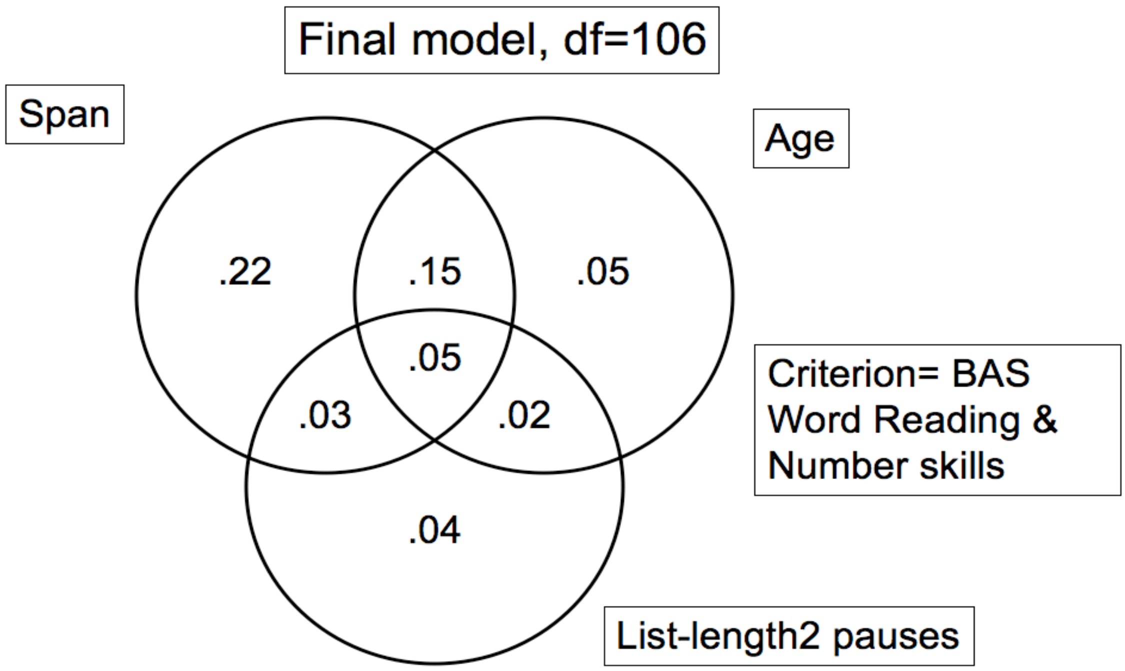
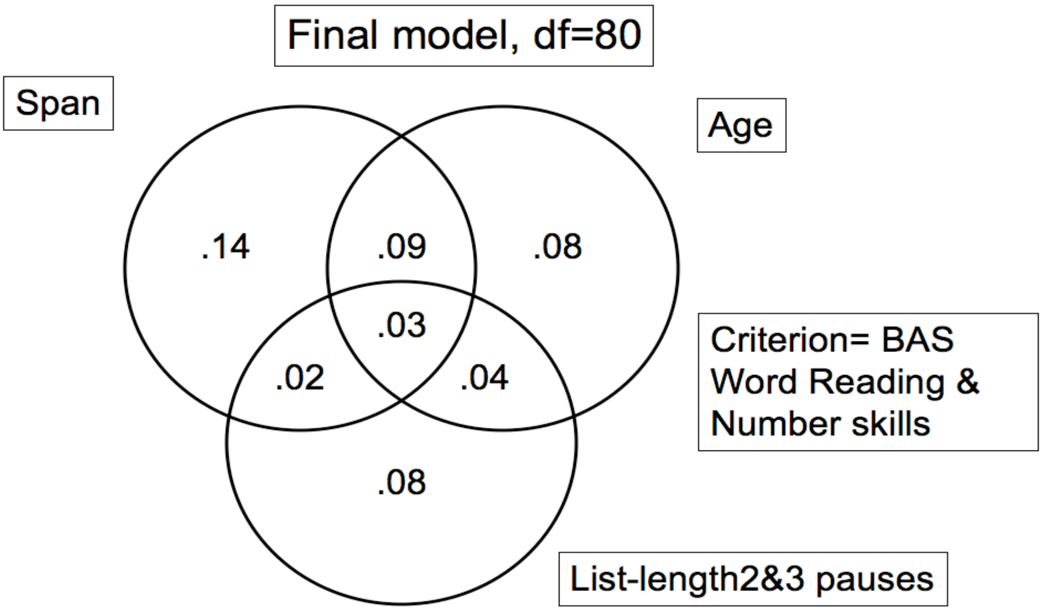


Figure 4. Schematic representation of the unique and common variance shared with the criterion skill measure from BAS performance, using children with both list-length 2 and list length-3 recall pause data.



Appendix 1. Further analysis of recall timing

Performance as a function of list length. Among children with correct recall of two and three item sequences, analysis of the first word, with age, list-length and sequence order as factors, indicated that words were produced more quickly at the shorter sequence length, $F(1,69)=38.3$, $p<.001$, $\eta_p^2=.357$ and by older children, $F(1,69)=7.83$, $p=.007$, $\eta_p^2=.102$, while there was no main effect of sequence order, $F<1$, $\eta_p^2=.011$. The list-length by sequence order interaction was marginally significant, $F(1,69)=3.63$, $p=.061$, $\eta_p^2=.050$ as was the three-way interaction between age, sequence order and list-length, $F(1,69)=3.14$, $p=.081$, $\eta_p^2=.044$.

Previous analyses have suggested that children's preparatory intervals do not systematically change with list-length. Consistent with that view, analysis of preparatory intervals with age, list-length and sequence order as factors yielded a non-significant effect of list-length, $F<1$, $p=.449$, $\eta_p^2=.008$. Older children began their recall more promptly, $F(1,69)= 8.17$, $p=.006$, $\eta_p^2=.106$, and preparatory intervals were shorter with descending sequences, $F(1,69)=4.95$, $p=.029$, $\eta_p^2=.067$. The sequence order by age interaction was marginal, $F(1,69)=3.24$, $p=.076$, $\eta_p^2=.045$. Other interactions were not significant.

Analysis of the first interword pause with age, list-length and sequence order as factors showed in contrast with preparatory intervals that pauses significantly increased for three- compared with two-item sequences, $F(1,69)=16.24$, $p<.001$, $\eta^2_p=.191$, almost tripling in duration. Pauses were marginally shorter among older children, $F(1,69)=3.00$, $p=.088$, $\eta_p^2=.042$, but there was no effect of sequence order, $F<1$, $\eta_p^2=.011$. Interactions were not significant.

Analysis of output position effects can help to indicate, for example, whether recall processes involve list wide search or whether earlier items can be excluded from consideration (see Haberlandt, Lawrence, Krohn, Bowe, Thomas, 2005). Analysis with respect to two-item sequences showed that the length of the first and second words did not differ, $F<1$, $\eta_p^2=.002$, although older children articulated words more quickly, $F(1,114)=9.08$, $p=.003$, $\eta_p^2=.074$, and word duration was shorter with descending sequences, $F(1,114)=4.38$, $p=.039$, $\eta_p^2=.037$. Corresponding analysis for sequences with three items indicated that word durations did become significantly shorter at later positions, $F(2,146)=19.7$, $p<.001$, $\eta_p^2=.213$ along with an age difference, $F(1,73)=6.09$, $p=.016$, $\eta_p^2=.077$, but no sequence order effect, $F<1$, $\eta_p^2=.002$. There was an interaction between output position and sequence order, $F(2,144)=3.41$, $p=.036$, $\eta_p^2=.045$; the speeding up in recall was more pronounced from the second to the third word with descending sequences. A comparison of the two interword pauses at list-length-3 produced no effect of output position, $F<1$, $\eta_p^2=.012$, age, $F(1,68)=1.48$, $p=.227$, $\eta_p^2=.020$, or sequence order, $F<1$, $\eta_p^2=.001$. None of the interactions was significant.

Longer list lengths. By aggregating data across testing sessions, it becomes feasible to incorporate sequences of four items into recall timing analysis, with means reported in Table 5. It is worth bearing in mind that these data come mostly from older children (n=33 from a sample of n=130) who would correspond to a high-span group in an extreme-group design (where the upper 25% quartile is often selected). Analysis collapses across age group because of the small cell sizes. Preparatory intervals, analysed with list-length and sequence order as factors, showed no significant effect of length, $F(2,62)=1.60$, $p=.211$, $\eta_p^2=.049$, nor order, $F<1$, $\eta_p^2=.006$, nor an interaction, $F(2,62)=2.50$, $p=.114$, $\eta_p^2=.075$. Pauses increased with the length of sequences, $F(2.62)=7.12$, $p=.005$, $\eta_p^2=.187$, but there were no effect of sequence order, $F(1,31)=1.89$, $p=.178$, $\eta_p^2=.058$, nor an interaction, $F(2,62)=2.10$, $p=.146$,

η_p^2=.063. Words were spoken more slowly at longer sequence lengths, $F(2,62)=11.6$, $p<.001$, η_p^2=.272, but there was no effect of sequence order, $F<1$, η_p^2=.017, nor an interaction, $F<1$, η_p^2=.027.

Measurement reliability in recall timing. Although rarely discussed in detail, measurement reliability is highly important to interpreting recall timing. Variations in the temporal dynamics of word production are self-evidently constrained by intelligibility demands, in a way that does not apply to pauses. Pauses are also more open-ended in the mental activities that can produce them. Given that pauses for sequences of two and three items are not correlated, and since reliability can severely constrain the inter-relationships between variables (Schmidt & Hunter, 1999) we examined the reliability of pauses further.

There are different ways of formalising the reliability of each recall variable; here, we calculated the correlation between the times of the first two analysable recall trials used to calculate data for analysis (using the raw, pre-Winsorized data). This led to a reliability estimates of $r(46)=.76$, and $r(6)=.49$ for sequences of two and three items. The corrected correlation for pauses between two- and three-item sequences in the first test session became $r_c(71)=.26$, $p<.05$, yet the correlations between pauses and scholastic ability were $r_c(109)=-.41$, and $r_c(71)=-.55$, $ps<.01$ for sequences of two and three items respectively (using normative data for estimating the reliability of the BAS). Thus, even after compensation for unreliability, the correlations between recall pauses and BAS scores were higher than that between the pauses themselves. At one level, this is paradoxical: classical test theory suggests that test-retest reliability places an upper bound on any link between that variable and another (e.g., Novick, 1966). However, the pauses at each list length need not be measuring exactly the same construct, and indeed the evidence from the data themselves suggests that task experience produces changes in working memory processes.