The endurance of children's working memory: a recall time analysis

Towse, John N (Department of Psychology, Lancaster University)
Hitch, Graham J (Department of Psychology, University of York)
Hamilton, Z. (Department of Psychology, Royal Holloway University of London)
Pirrie, Sarah (Department of Psychology, Lancaster University)

To appear in: Journal of Experimental Child Psychology. Please consult the journal for a final authoritative version of the manuscript.

Acknowledgements. We are grateful to the Lancaster University small grant scheme for provided the funds that made possible the extraction of data, to the ESRC (grant R000222789 and RES000230859), and the children of Emmer Green school. The paper was prepared in part while the first author was a research visitor at the University of Colorado at Boulder.

Address for Correspondence
John N Towse
Department of Psychology
Fylde College
Lancaster University
Bailrigg, Lancaster LA1 4YF,
United Kingdom
Email: j.towse@lancs.ac.uk, Tel: +44-1524-593705

Abstract

We analyse the timing of recall as a source of information about children's performance in complex working memory tasks. A group of 8-year-old children performed a traditional operation span task in which sequence length increased across trials and an operation period task in which processing requirements were extended across trials of constant sequence length. Interword pauses were larger than is commonly found in immediate serial recall tasks, yet shorter than for reading span. These pauses increased with the demands of recall, decreased across the output sequence and were to some extent predictive of scholastic ability. Overall, timing data illustrate that recall in working memory tasks involve subtle processes of item access rather than simple read-out of information from an immediate store.

The endurance of working memory: a recall time analysis

Working memory refers to the limited capacity systems and processes responsible for both acting upon and retaining transient representations (Baddeley, 1986; Baddeley & Hitch, 2000). The concept has been used for a particular architectural model of largely domain-specific memory devices (Halliday & Hitch, 1988; Hitch & Halliday, 1983) and for a more general framework describing active maintenance phenomena occurring alongside other cognitive activities (Conway, Jarrold, Kane, Miyake & Towse, 2007). We focus here on the general concept of working memory rather than any specific model.

It might appear self-evident that working memory capacity should be assessed with respect to *how much information* can be successfully retained whilst carrying out cognitive operations (Daneman & Carpenter, 1980). Working memory capacity is typically tested by varying sequence length (of words, numbers, visual objects etc., all embedded within processing episodes). This yields working memory span, the number of items remembered when also completing some mental operations. Measurement of children's working memory spans confirms that capacity is sharply limited, but increases with age. Moreover, individual differences in span are reliably predictive of reading and mathematics attainment, as well as general ability or fluid intelligence. This focus on the *size* of working memory has become very much the dominant approach in the field, and has been demonstrably successful in offering insights into experimental and differential issues (Conway et al., 2007).

The present research explores an alternative and potentially complementary measure of working memory. The working memory period paradigm has been developed as an operational measure of the endurance or persistence of temporary representations in a complex memory setting (Towse, Hitch, Hamilton, Peacock & Hutton, 2005). Unlike span, participants always have a constant number of items to remember. Trials vary instead in the duration of the processing requirements that accompany each memorandum. Working memory period is assessed by progressively increasing the processing requirements so as to determine the persistence of memory items retained in the face of ongoing cognitive activity. Thus, working memory *period* involves variations in the extent of processing whereas working memory *span* involves variations in sequence length. Towse et al. (2005) established that working memory period, like span, is broadly predictive of children's reading and number skills, and that these two measures have unique as well as shared features (see Towse, Hitch & Horton, 2007).

We argue that working memory period is especially suited for analyzing certain system characteristics. Whilst it is possible to explore serial position effects productively with working memory span (Unsworth & Engle, 2006a; Hutton, 2005), the composition of, and differential between, primacy, recency and middle or intermediate portions of the list necessarily changes with the number of items. Analysis within the working memory period paradigm is more straightforward and has greater statistical power with a constant sequence length. In the present study we capitalize on this characteristic to investigate recall timing.

A body of research on recall timing has already contributed to our understanding of short-term memory (Cowan 1999; Cowan et al., 2003; Haberlandt, Lawrence, Krohn, Bowe & Thomas, 2005; Hulme, Newton, Cowan, Stuart & Brown, 1999). Findings highlight the idea that memory tasks involve not only item maintenance, but also the construction and assembly of incomplete representations and their sequencing (Towse, Cowan, Hitch & Horton, in press). For example, pauses between successive items during recall typically increase with sequence length suggesting list-wise search processes. Also, pauses are shorter among children who go on to recall correctly longer sequences, implying that recall accuracy is influenced by search efficiency. Such regularities constrain theoretical accounts of working memory (Towse & Cowan, 2005).

The current research investigates the timing of recall for an operation period task, alongside operation span data. In each case, children solve sets of arithmetic problems and remember each answer for subsequent serial recall. We use this dataset to address four main issues.

First, recall timing phenomena have not to our knowledge been explored in working memory tasks with arithmetic processing requirements. Cowan et al. (2003) concentrated on recall timing for reading span and listening span tasks. Indeed, Cowan et al. (2003) proposed that children used memory of the sentence context to help scaffold the recall process (for further evidence, see Towse et al., in press). This accounts for why interword pauses were much shorter for counting span, which involves memory for the count totals of visual arrays (Case, Kurland & Goldberg. 1982) since the similarity of processing would fail to yield many cues for the reconstruction of memoranda. Thus, we predicted that operation span and operation period will contain briefer interword pauses than reading span since, unlike sentences, arithmetic problems are not especially meaningful or distinct from each other

Second, we compare and contrast period and span performance. Insofar as both procedures are designed as working memory measures, we predicted that where configurations match most, their profile will overlap. Moreover, with information available about children’s scholastic attainment, we explore individual differences in operation period recall timing. In this way, analyses facilitate an understanding of the points of continuity or otherwise between paradigms. We focus on 8-year-olds since they are commonly recruited into studies of working memory development and wider cognitive skills (e.g. Hitch et al., 2001).

Third, we sought to determine the impact of retention difficulty on the chronometry of recall. Analysis from short-term memory recall shows that interword pauses typically increase with more output items (e.g. Cowan et al., 1998). This could be because memory search acts on all potential memoranda or because each additional item is less securely remembered, or both. Since sequence length and task demand are inherently conflated, it is hard to address this question satisfactorily. Working memory period combines a constant sequence length with varying task demand. Therefore we can determine whether processing demand on its own affects the (relative or absolute) accessibility of memory items, as evidenced by interword pauses. We predicted that recall would become more protracted as children approached their endurance limits for retention.

Fourth, we investigated whether the length of interword pauses changed across across output position. Some immediate serial recall paradigms display serial position effects in recall timing (Haberlandt et al., 2005, but see Cowan, 1992). If participants can restrict memory search to not-yet-recalled items only, then interword pauses will become shorter at later positions.

Method

Participants

Forty-seven children producing audio recordings of correctly recalled period sequences, formed a subsample of those described previously in Towse et.al., 2005. Operation span recordings were available for forty-one of these children. Mean age was 8 years 8 months (ranging 7;8 to 9;6). All children attended school in South East England.

Procedure

Towse et al. (2005) describe test administration details in full. All children undertook number skills (in a group setting) and word reading (individually) assessments from the British Abilities Scale (BAS II: Elliott, Smith, & McCulloch, 1997). Operation period and operation span tasks were administered in counterbalanced order with an interval of approximately one week. Children always received instructions and practice on the arithmetic sums before either task commenced. Emphasis was placed primarily on computational accuracy with recall accuracy being secondary. No explicit mention was made about the timing of recall.

Each operation period trial comprised a sequence of four arithmetic problems presented visually on an Apple Powerbook 5300c computer. Each verbally-produced answer formed a memorandum. Immediately following completion of the fourth problem, a visual and auditory signal cued serial recall. Once the experimenter had entered the child’s sequence, the computer provided feedback on recall accuracy. Speech was initially recorded onto analogue audiotape with an external stereo microphone and later converted into digital files.

The initial set of three trials involved relatively short arithmetic problems (e.g. “5+0”, “9-0”); subsequent sets of three trials involved more time-consuming operations (e.g. “4+1” at the second level, “7-1-1” at the third level). Task progression required successful recall of at least one of three lists at a particular level. Although there were six task levels available, we focus here just on recall performance at the first three levels.

Operation span trials involved a sequence of equivalent ‘processing plus associated memory item’ episodes. However, whilst the content of arithmetic problems was comparable the *sequence length* varied across successive sets of three trials. Initial trial sets comprised two-item sequences, and increased by a single item (provided at least one of the three lists was recalled correctly) up to a maximum of seven.

Results

Extraction of recall times

Since the focus of analysis here lies in the chronometry of recall, we do not dwell on recall accuracy, especially since this is detailed in Towse et al. (2005). On each operation period trial, recall comprised eight sections; the preparatory interval (the gap between the audio-visual recall cue and the start of output); the four spoken words (the durations of the digits being recalled) and the three inter-word pauses. Operation span trials comprised a varying number of sections, according to the sequence length, but included at least four measurements; the preparatory interval, two spoken words and an interword pause.

Recall durations were estimated for correctly recalled sequences only. A single observer partitioned each response into contiguous phases, guided by the auditory signal and its visual waveform. Subsequently, a more experienced observer checked and where appropriate adjusted every interval. Both coders worked from a set of recall protocols, having been trained on a set of example timings[1].

Occasionally, timing data were excluded because of disruption such as when a child changed their mind about the sequence order and started over. All the remaining timing intervals were translated into z-scores and screened for outliers (values were Winsorized, to the equivalent of z= 3.29, affecting 27 of 2206 period measurements and 13 of 1320 operation span measurements).

Recall timing in operation period and span

We are not aware of published data on recall timing from operation period or operation span. This makes it important to establish whether task performance is comparable (given the conceptual overlap between them) and compare data with other relevant memory tasks. Table 1 describes the chronometry of recall for operation span involving four-item sequences and the second level of the period task - where the arithmetical processing operations were most similar. Table 1 also includes two other datasets where children recalled four-item sequences; from reading span among 9- and 11-year-olds (Towse, Cowan, Horton & Whytock, in press), and digit span (Cowan et al., 2003) where children's mean age was 8 years 9 months.

The data permit several conclusions. First, operation period and operation span show highly similar patterns of recall timing. Second, as might be expected all four tasks converge with respect to the recall duration of words. Third, interword pauses in operation period and operation span last longer than in digit span whilst being much shorter than in reading span, with means well outside the standard errors for these other tasks. Finally, it is apparent that the preparatory intervals in reading span are substantially different (with respect to central tendency and spread) from those in the other three tasks, which were all similar.

We next address two specific questions about the chronometry of recall in the operation period task. Does recall timing change across the difficulty

---

[1] An overview of procedures for spoken response timing can be found at the URL: http://www.psych.lancs.ac.uk/research/TowseWM/. A blind sample of timings from an independent coder correlated with the present values, r(68)=.88.

levels of the period task, and does recall timing change systematically across the output sequence?

Figure 1 displays the mean level of recall performance at the first, second, and third levels of the period task, where most data reside. It should be borne in mind that we plot mean performance based on all the available data. As a result fewer children contributed to scores at the third level relative the first. This affects the size of the standard error bars, and comparisons between levels, in that more advanced task levels involve an especially able subset of children.

Repeated-measures analysis confirmed an increase in the length of the interword pause across the three test levels, $F(2,58)=3.27$, $p=.045$, $\eta_p^2=.101$. Silent intervals between the recall of each item increased when the processing task took longer to complete. This supports the suggestion that on these more demanding trials, items were less clearly represented in working memory or were less accessible in some way.

The second question concerns the timing of recall across output position. Do children recall items from the end of the list more efficiently than those at the beginning? The answer is yes. The length of the interword pauses became shorter as the output sequence progressed. This was reliable for the first and second test level [$F(2,88)=13.0$, $p<.001$, $\eta_p^2=.228$, and $F(2,92)=19.8$, $p<.001$, $\eta_p^2=.302$ respectively]. The trend at the third level approached significance, $F(2,58)=2.91$, $p=.080$, $\eta_p^2=.091$. Word duration also became shorter across the output sequence for each test level [$F(3,129)=16.0$, $p<.001$, $\eta_p^2=.275$, $F(3,138)=7.05$, $p=.001$, $\eta_p^2=.133$, and $F(3,87)=7.62$, $p<.001$, $\eta_p^2=.208$ for the first to third level respectively].

Contrasting with analyses of words and pauses, the preparatory interval durations did not change across period level, $F(2,58)=2.15$, $p=.140$ $\eta_p^2=.069$. This is consistent with other immediate memory studies where preparatory intervals do not vary consistently with list length (Cowan et al., 1998; Cowan et al., 2003; Towse, Cowan, Horton et al., in press).

Individual differences in performance

Finally, we consider individual differences in recall timing. Timing variables in operation period are internally consistent. We obtained reliable correlations between total recall durations for the first and second levels, $r(43)=.814$, $p<.001$, the second and third levels, $r(28)=.585$, $p=.001$ as well as the first and third, $r(28)=.612$, $p<.001$, a pattern obtained with specific recall components too [e.g., for interword pauses, $r(43)=.782$, $p<.001$, $r(28)=.604$, $p=.001$, & $r(28)=.510$, $p=.004$ respectively].

The total duration of sequence recall ($n=30$ for children with data combined across levels) did not correlate significantly with the period score (the number of correct sequences), $r(28)=-.259$, $p=.166$, or ability as indexed by BAS, $r(28)=-.198$, $p=.294$. Nonetheless, across trial sets there was evidence of change in what recall timing variables measured. The interword pauses at just the first level *were* negatively associated with ability, $r(28)=-.369$, $p=.045$ such that pauses were shorter among children of higher ability. The strength of this association declined when pause measurements came from the second level, $r(28)=-.267$, $p=.154$, and declined again using pauses from the third, $r(28)=-.015$, $p=.936$. In other words, recall pauses reliably predicted cognitive ability, but only for the initial trials.

The group data indicates longer recall pauses at the more demanding task levels. Individual differences support this pattern; the time taken by an individual to answer the arithmetic problems across all task levels was correlated with their average interword pause in recall, r(28)=.443, p=.014. There was no corresponding relationship with word duration at recall, r(28)=.09, p=.639, so this is unlikely to be simply a global speed of processing effect; instead it suggests that period level affects item accessibility.

Discussion

The working memory period paradigm combines a fixed number of memoranda with systematic changes to the processing requirements that accompany them. Thus, in the operation period task, the participant calculates the answers to a constant number of multi-term arithmetic problems and remembers these answers for later recall. The arithmetic problems become progressively longer across sets of trials. The logic behind the task is that as the duration of the processing activity increases, so there is an increased opportunity for the working memory representations to become degraded to the point when recall is no longer successful (a logic that holds regardless of the causal mechanism for the loss of item accessibility). Of course, these changes in trial length are the result of manipulating what processing participants do, and the content of processing may contribute to the observed phenomena as well as its duration (see comments in Saito & Miyake, 2004; Towse et al., 2005).

The current analysis of the chronometry of recall helps to demonstrate that the fragility of memory increases as the processing component of the period task becomes more challenging. Children take significantly longer to produce the correct memory answers on more demanding period trials, consistent with the idea that the memoranda have become less accessible (individual-difference analysis also supports this finding) - this diminution in item accessibility could be conceived either in absolute or relative terms (ie loss of item integrity per se or changes in discriminability relative to others items). Moreover, other things being equal, one might expect recall to quicker as the task progresses, due to the benefits of practice (Towse, Cowan, Horton et al., in press). Thus the present findings probably underestimate the effect of increasing task demand.

Such evidence for changes in the (absolute or relative) accessibility of recall items supports the argument that working memory recall involves the revival of less-than-complete representations (Towse et al, in press). A related argument has been expressed by Unsworth and Engle (2006b). They propose that adults recall items either directly from (highly active) primary memory or indirectly from (cue dependent) secondary memory with the latter taken to involve memory search processes. Table 1 suggests that cued search in operation span and period is, on the one hand slower than digit span recall suggesting these tasks involve a greater contribution from secondary memory, whilst on the other hand search is much quicker than for reading span - which notably involves unique processing and memoranda events.

In short-term memory paradigms, interword pauses typically increase with the number of items to report (Cowan et al., 1994). Competition from extra response words offers one explanation for this effect. The present data suggest the operation of an additional factor; when the task is more difficult and each item less securely remembered, an item’s search and production process becomes slower.

More specific analysis of recall within the trial sequence shows reduction in the interword pauses and word duration as one progresses through recall (cf., Cowan, 1992). Adult interword pauses can also exhibit position effects (Haberlandt et al., 2005). Several factors may operate here. First, we suggest that children can at least partially restrict recall processes to act on those items not yet produced (i.e., implement a search and drop strategy, or sampling without replacement). Second, keeping sequence length constant in the period task may encourage such non-exhaustive search strategies. Third, later sequence items may be accessed to a greater extent through (quicker) primary memory processes (Unsworth & Engle, 2006a). Alongside pause effects, differences in word production suggest articulation is not modular, but also incorporates concurrent recall processes (Haberlandt et al., 2005).

Individual differences confirm that recall timing measures are reliable although the overall duration of recall does not predict separate ability measures. However, recall pauses from initial period trials did predict ability and recall accuracy. This replicates reading span data from Towse, Cowan, Horton et al. (in press) who demonstrated that task experience, independent of absolute difficulty can produce rapid changes in what working memory variables truly measure. They too found initial task performance was most highly predictive of ability and also noted changes in the properties of interword pauses with sequence length. Notwithstanding this specific correspondence, we recognize that the current sample size is modest for considering individual differences, and thus correlational outcomes should be treated with a certain caution.

In conclusion, while the working memory span paradigm is highly influential and we expect it to remain an influential vehicle for investigating complex memory, the present research shows that *working memory period* can be an informative and tractable procedure too. Span tasks focus on correct or failed recall and consequently one might conceive of the fate of encoded items as either remembered or forgotten. Whilst such a broad dichotomy is valuable at a general level, the present research indicates that memory representations can actually involve more subtle, graded forms (for a broader perspective, see Munakata, O’Reilly & Morton, 2007). There can be quantitative differences in the accessibility of remembered items, shown here with respect to processing duration and recall sequence position, and initial recall pauses correlate with ability measures. In sum, the chronometry of recall yields evidence about working memory processes that may be vital for a fuller understanding of this important system.

References

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

Baddeley, A. D., & Hitch, G. J. (2000). Development of working memory: Should the Pascual-Leone and the Baddeley and Hitch models be merged? *Journal of Experimental Child Psychology, 77*(2), 128-137.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short term memory span. *Journal of Experimental Child Psychology, 33*, 386-404.

Conway, A. R., Jarrold, C., Kane, M., Miyake, A., & Towse, J. N. (Eds.). (2007). *Variation in*

Working Memory. New York: Oxford University Press.

Cowan, N. (1999). The differential maturation of two processing rates related to digit span. *Journal of Experimental Child Psychology, 72*, 193-209.

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., et al. (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.

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

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.

Halliday, M. S., & Hitch, G. J. (1988). Developmental applications of working memory. In G. Claxton (Ed.), *New directions in cognition* (pp. 193-222). London: Routledge and Keegan Paul.

Hitch, G. J., & Halliday, M. S. (1983). Working memory in children. *Philosophical Transactions of the Royal Society of London, 302*(B), 325-340.

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*(2), 1-18.

Hutton, U. M. Z. (2005). *The similarities and differences between STM and WM tasks across development.* Unpublished doctoral dissertation, Royal Holloway University of London.

Munakata, Y., O'Reilly, R., & Morton, J. B. (2007). Developmental and computational approaches to variation in working memory. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake & J. N. Towse (Eds.), *Variation in working memory* (pp. 162-193). New York: Oxford University Press.

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.

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. (in press). The recall of information from working memory: insights from behavioural and chronometric perspectives. *Experimental Psychology*.

Towse, J. N., Cowan, N., Horton, N., & Whytock, S. (in press). Task experience and children's working memory performance: A perspective from recall timing. *Developmental Psychology*.

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.

Towse, J. N., Hitch, G. J., & Horton, N. (2007). Working memory as the interface between processing and retention: A developmental perspective. *Advances in Child Development and Behavior*, *35*, 219-251.

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

Unsworth, N., & Engle, R. W. (2006b). A temporal–contextual retrieval account of complex span: An analysis of errors. *Journal of Memory and Language, 54*(3), 346-362.

Table 1. The chronometric profile of correct recall for operation period (second task level) and four-item operation span trials (standard errors in parentheses). To facilitate comparison with other tasks, we have calculated and included comparable recall times for equivalent reading span and digit span trials. PI = preparatory interval; Wd1 = first word recalled; Pse1 = first interword pause; Wd2 = second word recalled; Pse2 = second interword pause; Wd3 = third word recalled; Pse3 = third interword pause; Wd4 = fourth word recalled. Digit span data derived from Cowan et al. (2003). Reading span data derived from Towse et al. (in press).

Recall segment								
	PI	Wd1	Pse1	Wd2	Pse2	Wd3	Pse3	Wd4
Operation period	.647 (.040)	.482 (.018)	.665 (.057)	.430 (.017)	.592 (.069)	.438 (.017)	.411 (.042)	.416 (.015)
Operation Span	.679 (.056)	.474 (.022)	.568 (.052)	.501 (.074)	.595 (.069)	.411 (.018)	.423 (.048)	.392 (.018)
Reading Span	1.935 (.443)	.591 (.037)	1.302 (.453)	.598 (.034)	1.067 (.184)	.593 (.041)	.655 (.137)	.464 (.026)
Digit span	.665 (.045)	.419 (.016)	.172 (.034)	.395 (.017)	.211 (.044)	.364 (.014)	.130 (.027)	.501 (.014)

Figure 1. Recall timing as a function of the period test level. Data are broken down for each recall segment, showing mean duration and standard error bars. PI = preparatory interval; Wd1 = first word recalled; Pse1 = first interword pause; Wd2 = second word recalled; Pse2 = second interword pause; Wd3 = third word recalled; Pse3 = third interword pause; Wd4 = fourth word recalled.