Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Memory and Language 54 (2006) 346–362

Journal of Memory and Language

www.elsevier.com/locate/jml

A temporal–contextual retrieval account of complex span: An analysis of errors ☆

Nash Unsworth *, Randall W. Engle

School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170, USA

Received 23 May 2005; revision received 18 November 2005
Available online 24 January 2006

Abstract

Complex working memory span tasks have been shown to predict performance on a number of measures of higher-order cognition including fluid abilities. However, exactly why performance on these tasks is related to higher-order cognition is still not known. The present study examined the patterns of errors made on two common complex span tasks. The results suggest that the patterns of errors made on these tasks are an important indicator of the processes that underlie performance and point to the importance of temporal–contextual cues. Furthermore, the individual differences data suggest that low scoring participants do not make more of each type of error at all serial positions, but rather the variability is localized to a few theoretically meaningful positions for each error type. The results suggest that low scoring participants have less precise temporal–contextual cues which leads to an inefficient search of memory.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Complex span; Errors; Individual differences; Fluid abilities

Introduction

Performance on complex verbal working memory (WM) span tasks has been shown to be an important predictor of a wide range of higher-order cognitive abilities such as reading comprehension (Daneman & Carpenter, 1980), learning (Kyllonen & Stephens, 1990; Unsworth & Engle, 2005a), performance on the Scholastic Aptitude Test (SAT; Turner & Engle, 1989), and measures of fluid abilities (Ackerman, Beier, & Boyle, 2002; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Kyllonen & Christal, 1990; Unsworth & Engle, 2005b). These complex span tasks have also predicted performance on many measures of lower-order cognitive abilities such as attention control in tasks such as antisaccade (Kane, Bleckley, Conway, & Engle, 2001; Unsworth, Schrock, & Engle, 2004), and Stroop (Kane & Engle, 2003; Long & Prat, 2002), as well as basic memory abilities in tasks such as categorical fluency and paired associates learning (Rosen & Engle, 1997, 1998). These relations point to the important role that working memory capacity (the construct thought to underlie performance on these tasks) plays in cognition. Deciphering the nature of this construct is an important endeavor that that can aid in the understanding of processes across a number of different cognitive domains.

Despite the impressive number and range of correlations these tasks demonstrate with other measures of cognitive abilities, little is known about the actual

---

☆ We thank Rich Heitz, Mike Kane, Tom Redick, John Towse, and two anonymous reviewers for helpful comments.

* Corresponding author.
E-mail address: gtg039d@prism.gatech.edu (N. Unsworth).

0749-596X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2005.11.004

processes that determine performance on the tasks themselves. Although some inferences can be made by examining the correlations between performance on complex span tasks with performance on another task, a more complete task analysis is needed to better understand the processes that are important for performance on the complex spans. For instance, several studies have used latent variable analysis of complex span tasks and examined the relation with higher-order tasks, such as those that reflect fluid abilities, by examining the role of other latent variables in the relation such as processing speed (Ackerman et al., 2002; Conway et al., 2002), or short-term memory (Engle et al., 1999). These studies demonstrate that the variance shared by the complex span tasks is related to variance that is shared by fluid abilities tasks over and above the variance from a third construct. What these studies do not tell us however, is what processes are responsible for that relationship. That is, these studies do not tell us about the processes that lead to performance differences on the span tasks directly and those processes that covary with performance on other cognitive tasks. To gain a better understanding of the processes involved in the complex span tasks and their ability to predict higher-order cognition, analyses of the components within the tasks is essential. In the present paper, we examine error patterns in complex span task based on a framework that suggests individuals use temporal–contextual cues to guide a memory search process. Thus, we attempt to examine notions of temporal–contextual retrieval that has been shown to account for performance in immediate serial recall (e.g., Brown, Preece, & Hulme, 2000) and continuous distractor free recall (e.g., Glenberg & Swanson, 1986) in terms of performance on complex span tasks.

Temporal–contextual retrieval

Recently, we have advocated a dual-component view of performance in complex spans (Unsworth & Engle, in press). This view combines a short-term maintenance component (primary memory) with a cue-dependent search process of longer-term memory (secondary memory). We have argued that one key to successful performance on the complex span tasks is the ability to successfully retrieve items from secondary memory under conditions of interference. Specifically, we argued that complex span tasks primarily require retrieval of items that have been displaced from primary memory. These displaced items must then be retrieved via a cue-dependent search process of secondary memory. For example, in the operation span task participants have to solve math operations while trying to remember unrelated items in the correct serial order. In such a task, items are briefly held in primary memory, but are quickly displaced from primary memory due to the need to switch attention to solving the math operations. Thus, most of the items in complex span tasks have to be recalled from secondary memory. The last presented item in a complex span task, however, is likely recalled from primary memory because there is no processing task after the last presented item to displace it from primary memory. The key to retrieving items from secondary memory is the ability to use cues to guide the search process. These cues include temporal–contextual cues, semantic cues, and phonological cues. However, because complex span tasks are episodic memory tasks, composed of unrelated items, we might assume that temporal–contextual cues will be primarily important in recall from secondary memory. In terms of the dual-component view that we have advocated, the current work is primarily focused on retrieving items from secondary memory via the use of temporal–contextual cues.

In line with previous research (e.g., Brown et al., 2000; Glenberg et al., 1980, Glenberg, Bradley, Stevenson, Kraus, & Renzaglia, 1983), let us assume that several different contextual elements can be associated with each item and that these different contextual elements change at different rates. During the presentation of each item in a list, the item is associated with contextual elements that can change throughout the experiment. As with Glenberg et al. (1980) the different rates of contextual change are assumed to come from a continuum, but for simplicity we will only concern ourselves with three different levels of contextual change. That is, there is a hierarchy of contextual elements at the top of which are slowly evolving global contextual elements associated with things like the room the experiment is being conducted in. At the next level of the hierarchy are slightly faster evolving contextual elements that are associated with each list. Finally, at the third level of the hierarchy are rapidly changing contextual elements associated with each to-be-remembered (TBR) item. During encoding, contextual elements from each level of the hierarchy are associated with each TBR item. It is assumed that retrieval is based on a cue-dependent search process in which items compete for selection. During recall, cues are used to delimit a search set and items are subsequently sampled and recalled from that search set. Because retrieval is a competitive process, the more items in the search set, the lower will be the probability of selecting any given item (i.e., cue-overload; Watkins & Watkins, 1975). At retrieval the different contextual elements can be used as cues to activate items associated with a given level of context. Cues based on the global context will tend to activate many items, whereas list contextual cues will only activate items presented within a given list, and the lowest level of contextual cues will activate only those few words associated with the rapidly changing contextual elements. Thus, each successive level of the hierarchy reduces the amount of cue-overload present, which leads to higher levels of recall.

During recall the contextual elements associated with the recall period will be used to sample items based on the match between the context at encoding and context at recall. Items presented close to the recall cue will share more contextual elements with the recall period than items presented further away and thus, will be associated with a higher probability of recall. This is because only a few items will share contextual elements with the recall cue leading to only a few items being in the search set. Thus, as with the temporal distinctiveness theory (TDT) of Glenberg and co-workers (Glenberg, 1987; Glenberg & Swanson, 1986) the temporal–contextual components are used to define search sets from which items are sampled. Crucially, the number of items within the search set varies depending on the time frame one is searching. Searching for items from the recent past results in a fairly narrowly defined search set. Searching for items that occurred farther back in time requires a more temporally extensive search set and hence more items are present in the search set (i.e., cue-overload). Furthermore, context for items presented further back in time will also be more similar to the context from previous trials leading to previous list items being included in the search set (i.e., intrusions). Thus, it is beneficial to have narrowly defined search sets so that only a few items are represented in the search set increasing the probability of correct recall. Consistent with other views of temporal–contextual retrieval (e.g., Brown et al., 2000; Glenberg et al., 1980, 1983), there are three critical assumptions: (1) items are associated with a hierarchal representation of temporal-context at encoding, (2) items are sampled based on the match between context at encoding and context at retrieval via a cue-dependent search process, and (3) the more items that are associated with the same temporal–context, the lower the probability of retrieving a given item will be (i.e., cue-overload).

Using a similar conception of temporal–contextual retrieval Glenberg and co-workers (Glenberg, 1987; Glenberg & Swanson, 1986) have been able to account for the long-term recency effect typically found in the continuous distractor task under a variety of conditions. Specifically, TDT suggests that the recency effect occurs because the last items are recalled via a narrowly defined search set. With only a few representations within the search set, there is a higher probability of recalling a given item and hence a recency effect. Midlist items, however, are associated with a lower probability of recall because their temporally defined search sets are much wider leading to more representations within the search set (see also Bellezza, 1982). Thus, when using temporal cues to guide the search process, the temporal distinctiveness view gives a straight forward account of the data. This same view should also be able to account for performance on complex span tasks given the high degree of similarity between complex span and continuous distractor recall. Specifically, in both tasks, TBR items are interspersed with some form of processing activity such as solving math operations. The main differences are that complex spans typically require serial recall, whereas free recall is typically required in the continuous distractor task. Additionally, there is typically a zero second retention interval in complex spans (i.e., the time between the presentation of the last item and the beginning of the recall period), but a much longer one in the continuous distractor task.

Due to the general similarities between the tasks, a temporal–contextual retrieval account should also work for complex spans. In fact, previous work by Li and Lewandowsky (1993) has shown that serial position functions from continuous distractor serial recall with varying retention intervals were well fit by Glenberg's (1987) temporal–distinctiveness theory. Accordingly, complex span serial position functions should be quite similar to those obtained from the continuous distractor task with large recency effects and smaller primacy effects. In addition, a temporal–contextual retrieval account should be able to account for different types of errors that are made in complex span tasks. Below, we discuss previous work that has examined errors in memory span tasks and then we outline predictions for the complex span tasks and individual differences based on temporal–contextual retrieval.

Error analyses of memory span tasks

One important indicator of performance that can be examined in memory span tasks is the different types of errors that participants make during recall (Bjork & Healy, 1974; Conrad, 1964; Estes, 1972). Indeed, several studies examining simple span tasks have shown that an examination of error responses can be extremely important in constraining theories of immediate memory (e.g., Brown et al., 2000; Henson, 1998), as well as demonstrating important individual, developmental, and group differences (e.g.,Elvevåg, Weinberger, & Goldberg, 2001; Maylor, Vousden, & Brown, 1999; McCormack, Brown, Vousden, & Henson, 2000). In these studies there are generally three types of error responses, two of which can be considered item errors and one of which can be considered an order error (see Henson, 1998). The item errors are omissions, where no response is given, and intrusions, where an item is recalled that is not from the current set. The transposition, or order, errors occur when an item is recalled from the current set but in the incorrect position. An examination of these different error types by serial position provides insights into the processes involved in serial order recall. The studies that have examined these errors have been done with immediate serial recall. The typical pattern of results is as follows: transpositions are the most frequently occurring error, and they tend to show a bowed serial position curve with the majority of transpositions occurring in the middle of the list and fewer at the end points; omissions are the next most frequent error, and they tend to increase with serial position; intrusions are the least frequently occurring error, and they too increase with serial position (e.g., Maylor et al., 1999).

An examination of error responses can provide information not only on the underlying processes involved in a task, but also can provide information about individual and group differences. For instance, Maylor et al. (1999) examined the error responses of younger and older adults on immediate serial recall and found that the performance deficit that older adults show is partially due to the fact that older adults made more omission errors than younger adults in the last serial positions. Maylor et al. (1999) suggested that this indicated that older adults' performance deficits may be due to output forgetting (decay or interference during recall). Maylor et al., also found that the transposition gradients for older and younger adults differed. Transposition gradients provide a means of determining how far items have moved from their original positions. For instance, if a person is presented with the sequence ABCDE and recalls ABCED, then there would be transpositions at positions four and five. Furthermore, these transpositions have moved to immediately adjacent positions, and thus have an item separation of one. Previous research has shown (see Brown et al., 2000) that the majority of transpositions have an item separation of one. According to a hierarchal representation of temporal–context (e.g., Brown et al., 2000) this occurs because items presented closer together are more likely to share temporal–contextual elements than items presented further away and thus are more likely to swap positions when recalled. Maylor et al. found that older adults had slightly shallower transposition gradients than younger adults, indicating that older adults were more likely than younger adults to have item separations of two or more. This suggests that older adults may have problems using temporal–contextual cues to place items into their correct positions.

Maylor et al., modeled their data with the OSCAR model (Brown et al., 2000) and found that manipulating two parameters accounted for most of the data. Specifically, the authors found that varying both the context quality parameter and the output forgetting parameter captured the main trends in the data as well as the differential aging effects. The authors suggested older adults may have problems using temporal–contextual cues to access items, as well as problems that are associated with output forgetting (decay or interference based forgetting during output). Thus, a group that is known to score low on complex working memory span tasks demonstrated an inability to use temporal–contextual cues to guide retrieval.

McCormack et al. (2000) similarly examined error responses in immediate serial recall. However, this research focused on differences between children of various ages and adults. The results suggested that younger children's errors were more likely to be omissions and intrusions than older children and adults. The older children and adult's errors were more likely to be transpositions. Furthermore, as with the Maylor et al. (1999) findings, younger children tended to have much shallower transposition gradients than the adults. That is, younger children were more likely to move items farther distances from their true positions than older children and adults. Brown, Vousden, McCormack, and Hulme (1999) modeled these data with the OSCAR model and found that varying the learning-context parameter resulted in the best simulation of the data. Thus, the authors suggested that one possible reason for developmental differences are differences in the ability to use temporal–contextual cues and that this ability increases with age. As with the aging study by Maylor et al. (1999), these results suggest that temporal–contextual cues are important in memory span tasks and individual differences in those tasks. Both studies suggested that individuals who perform poorly on memory span tasks do so in part because they either forget items altogether, or they are more likely to move items away from their current positions.

Other studies that have examined error responses in memory tasks have shown differences in errors between schizophrenic patients and controls in immediate serial recall (Elvevåg et al., 2001), as well as finding that older and younger adults differ in the number of intrusions in a complex span task (Lustig, Hasher, & May, 2001), and in a memory updating task (De Beni & Palladino, 2004). Additionally, research has shown that individuals with reading disabilities demonstrate more intrusions in a complex span task than do skilled readers (Chiappe, Hasher, & Siegel, 2000). All of these studies point to the value of using error analyses to understand the processes that underlay task performance and to understand individual differences.

The present investigation

The goals of the present investigation were threefold. The first goal was to examine the pattern of errors in complex span tasks to understand the underlying processes involved (see Towse, Hitch, Hamilton, Peacock, & Hutton, 2005 for similar analyses involving children). According to the temporal–contextual retrieval account outlined previously, three main predictions should be met.

(1) Items presented further from the recall period should be the most likely to be forgotten (i.e., omission errors) due to increased breadth of search and cue-overload.

(2) In accord with the first prediction, items presented further from the recall period should be more susceptible to intrusions from previous lists due an increased breadth of search (due to shared/similar context) that includes items from previous trials. Hence, intrusions should be more likely to occur at the first few positions and should decrease for later positions.

(3) Even if an item is recalled correctly, it still needs to be placed in the correct position, thus when transpositions occur they should be more likely to move to adjacent positions than to move to further away positions. In this case, serial recall can be conceived of as a two-stage discrimination process (e.g., Tehan, Hendry, & Kocinski, 2001) in which current list items are separated from previous list items (i.e., delimiting the search set in the first two predictions) and then correct items are recalled and placed into the correct serial position. Hence, temporal–contextual cues are needed not only to define the search sets, but are also needed to determine relative positioning of items. Any noise in the temporal–contextual cue can lead to movement errors. Thus, complex spans should show steep transposition gradients somewhat similar to what has been found with immediate serial recall.

The second goal of the present investigation was to examine performance differences between high, medium, and low scorers on complex span tasks in terms of patterns of errors. Although the low scoring individuals by definition make more errors, it is possible that they make more of a particular type of error or that they make more errors at a given serial position than do high spans. Examining individual differences in the processes that are engaged in these tasks is important in understanding performance differences on the tasks themselves as well as their predictive utility to higher-order cognitive abilities (e.g., Carroll, 1988; Estes, 1982; Hunt, 1978). In particular, we have argued that low scorers on complex span tasks are poorer at using context cues to guide the search process of secondary memory than higher scorers (Unsworth & Engle, 2006). Thus, low complex span individuals should be more likely to forget items presented further away from the recall period due to noise in their contextual cues which leads to cue-overload and a higher incidence of intrusion errors. Furthermore, due to increased noise in their contextual retrieval cues, low complex span individuals should be more likely to make transposition errors than higher scoring individuals and these errors should move greater distances than transpositions for higher scoring participants. That is, similar to the aging and development studies reviewed previously, low complex span individuals should have shallower transposition gradients than high complex span individuals, which indicates greater noise in their temporal–contextual cues than high complex span individuals.

The third, and final, goal of the present investigation was to examine the relative contribution of each error type to fluid abilities (gF). In the past complex span tasks have been shown to correlate quite well with measures of fluid abilities. These correlations typically rely on one index of performance on the complex span tasks, namely mean recall accuracy. However, it is unclear why a particular individual may have low recall accuracy. Is it because the individual could not remember any words (i.e., many omissions), because the individual recalled many nontarget words (i.e., intrusions), or because the individual correctly recalled many of the words but simply did not place them in the correct serial positions (i.e., transpositions)? Hence, examining error responses and their correlation with a measure of higher-order cognition should provide us with important information about the cause of the predictive power of complex span tasks. That is, an examination of error response provides a means of breaking up a single recall accuracy score into several subcomponents, which may have differential predictive utility.

We examined the error responses for 150 participants in two verbal complex span tasks (Ospan and Rspan). These subjects were a sub-sample from the Kane et al. (2004) study. For each subject the total number of omission, intrusion and transposition errors was determined for each list-length. Errors in list-length five were further examined in regard to their serial position. This allowed us to determine if a given type of error was committed at one serial position more than another. This information is important in determining the processing that constrains performance in these tasks. Participant scores on three fluid abilities tests were used to form a factor composite for correlational analyses.

Method

Participants

The data analyzed in the current study were a subsample of participants from a large correlation-based study (Kane et al., 2004). The sample consisted of 150 participants between the ages of 18 and 35. Participants were both college students and community volunteers from a combination of three universities and metropolitan areas (see Kane et al., 2004 for more details). None of the analyses or results reported in this paper were reported in the Kane et al. (2004) study.

Tasks

All participants completed a number of complex span, simple span, and reasoning measures. For the present investigation we analyzed data from two popular verbal complex span tasks (operation span and reading span) and three of the spatially oriented general fluid abilities measures.

Operation span (Ospan)

Participants solved a series of math operations while trying to remember a set of unrelated words. Participants saw one operation-word string at a time. For each trial participants were required to solve the operation and read the word aloud. Immediately after the participant read the word, the next operation-word string was presented. Three trials of each list-length (2–5) were presented, with the order of list-length varying randomly. At recall, words from the current set were written in the correct order. To ensure that participants were not trading off between solving the operations and remembering the words, an 85% accuracy criterion on the operations was required. Participants received three sets (of list-length two) of practice. For both of the span measures, items were scored if the item was correct and in the correct position. The score was the proportion of correct items in the correct position.

Reading span (Rspan)

Participants were required to read sentences while trying to remember a set of unrelated letters (B, F, H, J, L, M, Q, R, and X). For this task, participants read a sentence and determined whether the sentence made sense or not (e.g., "The prosecutor's dish was lost because it was not based on fact? M"). Half of the sentences made sense while the other half did not. Nonsense sentences were made by simply changing one word (e.g., "dish" from "case") from an otherwise normal sentence. There were 10 – 15 words in each sentence. Participants were required to read the sentence aloud and to indicate whether it made sense or not by saying either "yes" (makes sense) or "no" (does not make sense). After participants gave their response they said the letter aloud. At recall, participants wrote down the letters from the current set in the correct order. There were three trials of each set-size with list length ranging from 2 to 5. The same scoring procedure as Ospan was used.

Raven

Raven Advanced Progressive matrices (Raven, Raven, & Court, 1998) is a measure of abstract reasoning. The test consists of 36 items presented in ascending order of difficulty (i.e., easiest–hardest). Each item consists of a display of 3 × 3 matrices of geometric patterns with the bottom right pattern missing. The task for the participant is to select among eight alternatives, the one that correctly completes the overall series of patterns. Participants had 10 min to complete the 18 odd-numbered items. A participant's score was the total number of correct solutions.

WASI matrix reasoning

Each item presented a pattern of novel, colored figures, and most were arranged in a matrix with one figure missing. Nine items presented 2 × 2 matrices, 2 items presented 3 × 3 matrices, 2 items presented a missing piece from a continuous wallpaper-like design, and 1 item presented a missing piece from a linear sequence of 5 figures. Participants selected the one of five figures that would best complete the pattern. Participants had 7 min to complete 14 items that increased in difficulty. The items were 14, 16, 18, 20, 22, 24, 26, 28, 30, 31, 32, 33, 34, and 35 from the WASI test (Psychological Corporation, 1999).

BETA III

Each item presented a pattern of 3 novel, black-and-white figures arranged in a 2 × 2 matrix with one figure missing. Participants selected the one of five figures that would best complete the pattern. Participants had 10 min to complete 20 test items that increased in difficulty. The items were the 20 final items (numbers 6–25) of the BETA III test (Kellogg & Morton, 1999).

Results

The results are divided into three primary sections. The first section examined the total number of both output and input errors for all of the list-lengths as a function of span. The second examined both output and input errors only for list-length five as a function of both serial position and complex span and finally, the third section examined the unique and shared contribution of each error type in predicting fluid abilities. For all analyses, the two complex spans were combined, and the three fluid abilities measures were combined to form single composites (see the Appendix A for a breakdown of the errors for each span task).[1]

Errors were classified in terms of both output and input position (see Henson, 1998, Appendix A). Output errors consist of omissions, transpositions, and intrusions. Input errors can only be omissions or transpositions. An output omission occurs when a participant leaves a blank space during output. For instance, if presented with ABCDE, a participant may recall AFDE__. The blank space at position five would be considered an output omission. An output transposition occurs when an item from the current set is recalled in a different serial position than its correct serial position. For example,

---

[1] For the working memory span tasks, the composites were a simple average of probability correct for the two complex span tasks. For the error analyses each error type was averaged across the two complex span tasks. The gF composite was a factor composite of the three reasoning tasks.

D and E in the example are not recalled in their correct serial positions, but instead have moved inward. Finally, an output intrusion occurs when an item that is not from the current set is recalled. In the example, F at position two would be considered an intrusion because it is not an item from the current set. Omissions and transpositions can also be classified according to their input position. An input omission occurs when an item from the presented list is not recalled at all. In the example both B and C (positions two and three) would be considered input omissions. An input transposition is exactly the same as an output transposition except that the position has changed. That is, in the example D and E are transpositions but would be considered output transpositions at positions three and four but input transpositions at positions four and five.

Overall errors

The total number of each type of error for all list-lengths was computed and examined by complex span.[2] The upper quartile of participants, based on a *z*-score composite of the two complex span tasks, were deemed high spans ($N = 37$), the mid 50% of participants were deemed mid spans ($N = 76$), and the lower quartile of participants were classified as low spans ($N = 37$).[3] The goal of this analysis was to examine the relative contribution of each error type to the total number of errors and to determine if high and low scoring individuals differed on one error type more than another. By definition, the low spans will make more errors than either the high or the mid spans, but it is possible that the majority of low spans' errors are of one particular type, thus localizing the source of variability.

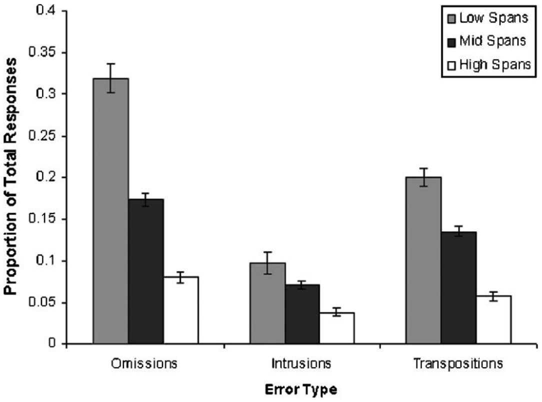

Fig. 1. Mean numbers of output omissions, intrusions, and transpositions as proportions of the total number of responses and as a function of complex span.

Output errors

As shown in Fig. 1, the results suggest that the majority of errors are output omissions, followed by output transpositions, and then intrusions. Note that the dependent variable for these analyses is proportion of total responses. Furthermore, low spans made more of each type of error, but they made disproportionately more omission errors than the other span groups, followed by transposition errors, and intrusions. Thus, the major difference between the span groups seems to be differences in the number of omission errors made.

These results were supported by a 3 (error type: omission, transposition, and intrusion) × 3 (complex Span: high, mid, and low) mixed analysis of variance (ANOVA) with error type as the within subjects variable. The analysis yielded main effects of error type, $F(2, 294) = 106.05$, $p < .01$, partial $\eta^2 = .42$, and, of course, complex span, $F(2, 147) = 326.82$, $p < .01$, partial $\eta^2 = .82$. Additionally, these two factors interacted, $F(4, 294) = 16.76$, $p < .01$, partial $\eta^2 = .19$. This interaction suggests that the difference between the span groups is largest for omission errors, followed by transpositions and then intrusions. For instance, the ratio of low span to high span errors for omissions was 3.99:1, for transpositions 3.51:1, and for intrusions 2.55:1.[4] Thus, the majority of variability seems to lie within differences in omission errors.

Input errors

Scoring the results by input suggests a highly similar pattern of results. Specifically, because the number of transpositions does not change (only the position of errors does) and because the input omissions is simply output omissions plus output intrusions, the results

---

[2] Note that the same pattern of results were obtained when examining each complex span task individually as examining the composites (see the Appendix A). The only major differences occurred when examining output omission and intrusion errors for the two span tasks. Intrusions were twice more likely in reading span than in operation span (2.5:1). This difference was localized to the first three serial positions for list-length five. This is to be expected given that the stimuli for the Rspan were from a fixed pool and for the Ospan were from an unlimited pool of items.

[3] We analyzed the data using ANOVAs instead of linear regressions because we were primarily interested in the pattern of errors as a function of serial position, which are clearly nonlinear. Although trichotomizing the data, as we have done, may not be the optimal statistical strategy, we chose to examine the data this way because we were principally concerned with the pattern of errors across serial positions as well as individual differences in those patterns.

[4] Intrusion errors are a combination of both intra- and extra-experimental intrusions. That is, intrusions could either be items from previous sets or items that were not from previous sets. The results suggest intra-experimental errors were three times more likely than extra-experimental errors (3.01:1).

should show a similar pattern with large differences in input omissions followed by input transpositions. This is indeed the case. There were main effects of both error type, $F(1, 147) = 275.62$, $p < .01$, partial $\eta^2 = .65$, and complex span, $F(2, 147) = 310.24$, $p < .01$, partial $\eta^2 = .81$. These two factors interacted, $F(2,147) = 29.34$, $p < .01$, partial $\eta^2 = .29$, suggesting that the differences between the span groups was largest for input omissions followed by input transpositions.

Correct responses by serial position for list-length five

Before presenting the error serial position curves, we examined the correct response serial position curve. As shown in Fig. 2, the resulting curve shows both primacy and large recency effects. Examining this by complex span, suggests that, as expected, low spans have poorer recall at list-length five than the other span groups, but that the span effect is equivalent at all serial positions. These impressions were confirmed by a 3 (complex span) × 5 (serial position) mixed ANOVA with serial position as the within subjects effect. There were main effects of both serial position, $F(4, 588) = 53.84$, $p < .01$, partial $\eta^2 = .27$, and complex span, $F(2,147) = 159.9$, $p < .01$, partial $\eta^2 = .69$. These two factors, however, did not interact, $F(8, 588) = 1.42$, $p > .18$, partial $\eta^2 = .02$.

Output errors by serial position for list-length five

Next, we examined the serial position curves for each error type at list-length five by complex span to determine whether the differences between the span groups occurred across all serial positions or whether it was localized to specific serial positions. Furthermore, examining each error type by serial position should provide us with a fairly detailed picture of the processes that occur during recall. As with the overall error analyses, errors were examined by both output and input position.

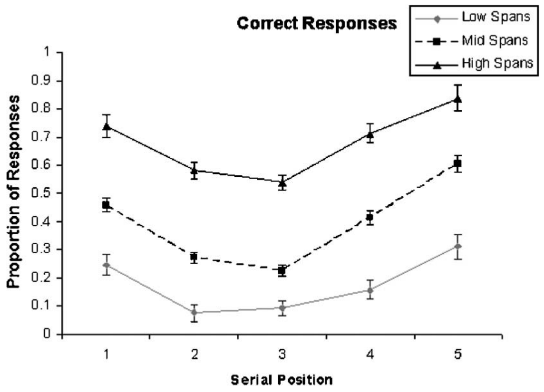

Fig. 2. Serial position curve for correct responses as a proportion of total responses and a function of complex span.

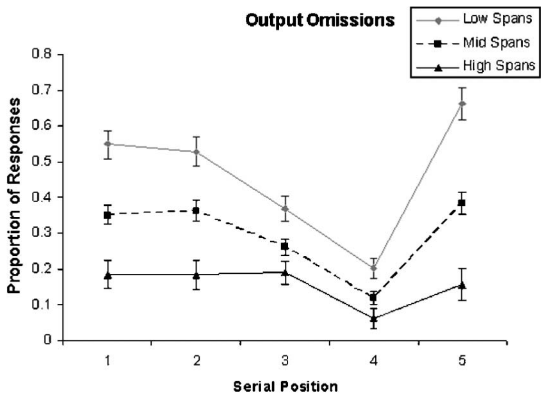

Fig. 3. Serial position curve for output omissions as a proportion of total responses and a function of complex span.

Omissions

As shown in Fig. 3, plotting omission errors by serial position as a proportion of total responses, we see that output omissions occur most frequently at the first and last serial positions. Output omissions are less likely at the middle serial positions (3 and 4). Additionally, examining this by complex span suggests that there are large individual differences for the first and last serial positions but that these differences are drastically reduced for the middle serial positions. These impressions were confirmed by a 3 (complex span) × 5 (serial position) mixed ANOVA with serial position as the within subjects variable. The analysis yielded main effects of serial position, $F(4, 588) = 31.05$, $p < .01$, partial $\eta^2 = .17$, and complex span, $F(2, 147) = 78.57$, $p < .01$, partial $\eta^2 = .52$. There was also a reliable two-way interaction between serial position and complex span, $F(8, 588) = 4.01$, $p < .01$, partial $\eta^2 = .05$. These effects suggest that the difference between the span groups was moderated by serial position.

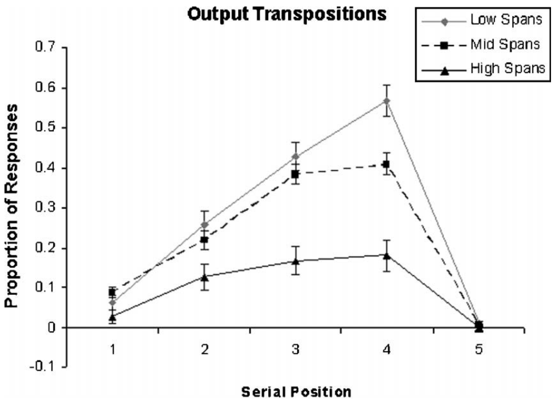

Fig. 4. Serial position curve for output transpositions as a proportion of total responses and a function of complex span.

Transpositions

Fig. 4 shows output transpositions by serial position as a proportion of total responses. In contrast to omissions, the curve shows that there are very few transpositions at the first and last serial positions, but that the majority of transpositions occur for the middle serial positions. Examining this by complex span, shows that individual differences are also localized to the middle three serial positions with virtually no individual differences occurring at the first and last serial positions. This was supported by a 3 (complex span) × 5 (serial position) mixed ANOVA with serial position as the within subjects variable. Main effects of serial position, $F(4, 588) = 127.78$, $p < .01$, partial $\eta^2 = .47$ and complex span, $F(2, 147) = 31.26$, $p < .01$, partial $\eta^2 = .30$. The two-way interaction between these factors was also reliable, $F(8, 588) = 8.86$, $p < .01$, partial $\eta^2 = .11$.

Intrusions

For our final analysis of output errors by serial position, we examined output intrusions. As shown in Fig. 5, plotting intrusions by serial position suggests that the occurrence of an intrusion is more likely at the first serial positions, and then decreases to near zero for the last serial position. Thus, participants are more likely to intrude an item from a previous list, or an extra-experimental item, at the first few serial positions than at the last serial position. Examining this by complex span suggests that the different span groups do not differ in the likelihood of an intrusion at any of the serial positions except for the very first position. Thus, it is not the case that low spans make more errors at all serial positions, but rather the differences can be localized to a given position. These impressions were confirmed by a 3 (complex span) × 5 (serial position) mixed ANOVA with serial position as the within subjects variable. The only reliable effect was a main effect of serial position, $F(4, 588) = 26.26$, $p < .01$, partial $\eta^2 = .15$. Neither the main effect of span, $F(2, 147) = 2.24$, $p > .10$, partial $\eta^2 = .03$, nor the two-way interaction was reliable, $F(8, 588) = 1.40$, $p > .18$, partial $\eta^2 = .02$. However, follow up analyses revealed a reliable span difference only at the first serial position, $F(2, 147) = 5.48$, $p < .01$, partial $\eta^2 = .07$, all other ps > .33 and partial η^2s < .02.

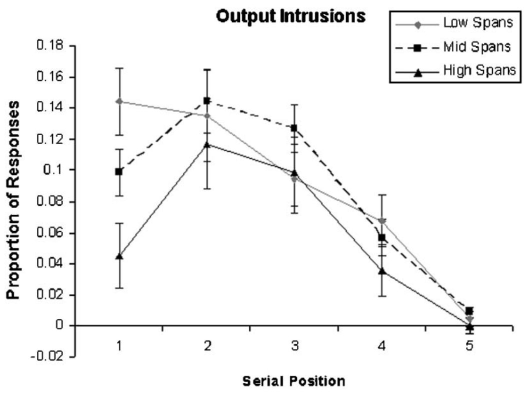

Fig. 5. Serial position curve for output intrusions as a proportion of total responses and a function of complex span.

Summary of output errors by serial position

Taken together, these results paint an interesting picture of performance on the complex span tasks. The results suggest that the occurrence of an output omission is higher at the first and last serial positions compared to the middle serial positions. In contrast, the output transposition data suggests that the likelihood of a transposition is highest for the middle serial positions and lowest for the endpoints. Together, this suggests that first and last serial positions are most likely to be left blank and that perhaps the items that were supposed to be in those positions were moved inward to the middle serial positions resulting in a transposition error. Additionally, although intrusion errors were infrequent compared to the other error types, the results suggested that when an intrusion was made, it most likely occurred at the first few serial positions and rarely, if ever, occurred at the last serial position. The complex span effects suggest that although low spans tend to make more omissions and transpositions, these effects are generally localized to specific serial positions. For the output omissions, the span differences are largest at the first and last serial positions and much smaller at position four. In contrast, for the output transpositions, all the span differences occur at the middle three serial positions with virtually no differences occurring at the first and last serial positions. Thus, it is not simply the case that low spans make more of each type of error at all serial positions, but rather that they are more likely to make a given error at specific serial positions. Indeed, examining the output intrusions suggests that the span groups do not differ in the amount of intrusions made for a list-length of five. However, significant span differences did appear at the very first serial position, suggesting that low spans are more likely to intrude an item at the first serial position than are the other two span groups.

Input errors by serial position for list-length five

Next we examined errors by input position for the different span groups. Recall that input errors represent errors plotted by their input positions. Thus, not recalling an item presented at position one would be considered an input omission for position one. An input transposition is the same as an output transposition (an item recalled correctly but in the incorrect serial position) except that the position of the error has changed. For

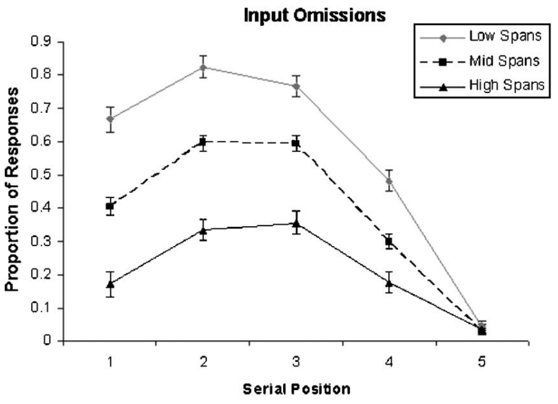

Fig. 6. Serial position curve for input omissions as a proportion of total responses and a function of complex span.

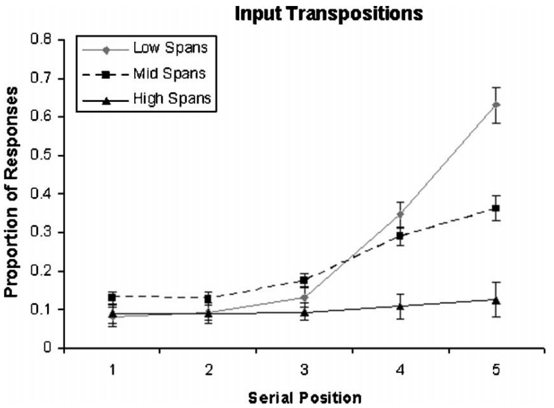

Fig. 7. Serial position curve for input transpositions as a proportion of total responses and a function of complex span.

instance, recalling an item presented at position five but placing it in position four would be considered an input transposition at position five. Thus, input errors contain the same basic information as output errors but the serial positions of the errors has changed. Note that output intrusions are subsumed under input omissions.

Omissions

Similar to the output analyses reported above, we examined the serial position functions for input omissions and input transpositions. Input omissions reflect those items that were presented to the participant but not recalled (see description). As shown in Fig. 6, plotting input omissions by serial position demonstrates that the last item presented (input position five) is nearly always recalled. However, the farther back in time an item was presented, the greater the likelihood of an omission. The first item presented (input position one), however, seems to have some special accessibility as demonstrated by the lower probability of an omission error compared to positions two and three. Examining these effects by complex span suggests that all participants, regardless of complex span score, recall the last item presented. However, the further back in time an item was presented, the larger the individual differences become. That is, individual differences begin at input position four and get larger moving back towards position one. These results were supported by a 3 (complex span) × 5 (serial position) mixed ANOVA with serial position as the within subjects variable. There were reliable main effects of both serial position, $F(4, 588) = 199.74$, $p < .01$, partial $\eta^2 = .58$, and complex span, $F(2,147) = 137.17$, $p < .01$, partial $\eta^2 = .65$. These two factors interacted, $F(8, 588) = 11.49$, $p < .01$, partial $\eta^2 = .14$.

Transpositions

An examination of input transposition errors suggests that the majority of input transpositions occur at the last two serial positions as shown in Fig. 7. The first three serial positions show roughly the same amount of transpositions, but the likelihood of transpositions increases drastically for the last two serial positions with the greatest occurrence at the last serial position. Thus, the last two items presented are being recalled, but they tend to be recalled in the incorrect serial position. Examining these effects by complex span suggests that there are few individual differences for the first three serial positions but there are substantial individual differences for the last two serial positions. This was supported by a 3 (complex span) × 5 (serial position) mixed ANOVA with serial position as the within subjects variable. The analysis yielded reliable main effects of serial position, $F(4, 588) = 57.26$, $p < .01$, partial $\eta^2 = .28$, and complex span, $F(2,147) = 25.89$, $p < .01$, partial $\eta^2 = .26$. These two factors, also demonstrated a reliable interaction, $F(8, 588) = 14.33$, $p < .01$, partial $\eta^2 = .16$.

Summary of input errors by serial position

Combining the analyses for the input serial position curves suggests that items presented at the beginning of the list are the most likely to be forgotten, while items presented at the end of the list are usually associated with a high probability of recall. However, the items presented at the end of the list tend to be recalled in the incorrect serial position. Thus, despite the low probability of an input omission for the last serial positions, there is a high probability of an input transposition associated with the last serial positions. This suggests that the items presented at the end of the list are recalled, but they are recalled at earlier serial positions. The complex span effects suggest that the difference between the span groups seems to be localized to the first few serial positions for the input omissions, but localized to the last serial positions for the input transpositions. Thus, low complex span individuals tend to forget items from the beginning of the list more than higher scoring individuals. Additionally, when low complex span individuals are successful in recalling an item, they are more likely to recall that item in the incorrect serial position, especially for the last two presented items. High complex span individuals, however, tend to forget fewer items at all serial positions and are less likely to transpose items. In fact, as shown in Fig. 7, the probability of a transposition does not change across serial positions for high spans, $F(4, 144) = .42$, $p > .78$, partial $\eta^2 = .01$. Rather the probability holds constant around.10. The mid span individuals' performance tends to fall in between the high and low spans' performance.

Transposition gradients

Transposition errors were analyzed in more detail to better examine the hypothesis that temporal–contextual retrieval is important as discussed previously. Specifically, for each transposition error we calculated the separation between the item's presented position and its recalled position. For instance, if presented with ABCDE and a person recalls ABCED, there would be two transposition errors each with an item separation of one. This was done for each participant in each task. Note that five individuals who did not make any transposition errors at a list-length of five were excluded from these analyses. This included three high spans, one mid span, and one low span. Shown in Fig. 8 are the means for high, mid, and low complex spans as a function of item separation. The means are computed as a proportion of total transposition errors. Also shown in Fig. 8 is what would be expected if the transposition errors were distributed at random.

These results suggested that the majority of transposition errors are errors where the correct item is placed in an immediately adjacent position. Furthermore, as shown in Fig. 8, there are group differences in the pattern of transpositions. Specifically, high complex span individuals were more likely than low or mid complex span individuals to have item separations of only one. In contrast, this pattern reverses for item separations of two where, now low and mid complex span individuals were more likely to move items to two adjacent positions. That is, high complex span individuals tended to have steeper transposition gradients than mid and low complex span individuals. To examine this statistically, we computed the average separation where at least one transposition error was made. This analysis suggested that transpositions had an average separation of 1.22 across all participants. Furthermore, average item separation was highest for low spans ($M = 1.30$, $SE = .04$), followed by mid spans ($M = 1.24$, $SE = .03$), and then high spans ($M = 1.11$, $SE = .05$). An ANOVA with Complex span as the between subjects variable supported these observations with a significant main effect of Complex span, $F(2, 142) = 4.65$, $p < .05$, partial $\eta^2 = .06$. These results suggest that when a transposition error occurs, the item is likely to be moved from the correct position to an immediately adjacent position. Furthermore, high spans were more likely to move items a shorter distance from the target position than either mid ($p < .05$) or low spans ($p < .01$).

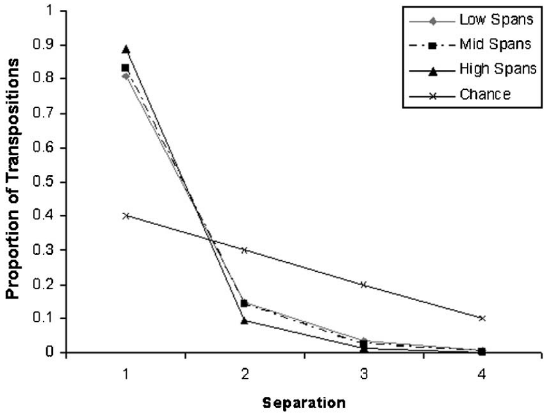

Fig. 8. Mean proportion of transposition errors as proportions of total transposition errors for list-length five as a function of separation and complex span. The gradient expected by chance alone is also shown.

Regression analysis of error types and fluid abilities

Our final analysis concerned the unique and shared contributions of each error type in predicting fluid abilities. Note for these analyses we only analyzed the three output error types. This was done because input omissions are a combination of both output omissions and intrusions. Thus, the output errors were examined to get a detailed picture for all three error types. Our goal was to determine if the shared variability between complex span tasks and fluid abilities that has been demonstrated in the past (e.g., Engle et al., 1999) was due to the shared variability between the different errors, due to unique variability from each error, or from a combination of both unique and shared variability. To examine this, we used the total number of each output error type for all list-lengths just as was done in the previous overall error analyses. Table 1 shows the zero-order correlations between the three error types with each other and

Table 1
Zero-order correlations between the three output error types and fluid abilities

Variable	1	2	3	4
1. Omissions	—			
2. Transpositions	.41**	—		
3. Intrusions	−.08	.31**	—	
4. gF	−.27**	−.37**	−.39**	—

** $p < .01$.

Table 2
Simultaneous regression predicting fluid abilities

Variable	*B*	*t*	sr^2	R^2	*F*
Omissions	−.23	-2.89^{**}	.04		
Transpositions	−.17	-1.99^{*}	.02		
Intrusions	−.35	-4.52^{**}	.10	.26	17.03^{**}

* $p < .05$.
** $p < .01$.

with fluid abilities (gF). The zero-order correlations suggest that both omissions and intrusions are moderately related to transpositions, but that omissions and intrusions are not related to one another. Furthermore, the zero-order correlations suggest that all of the different error types are moderately related to fluid abilities.

To examine the unique and shared contribution of the different error types in predicting fluid abilities, we entered all of the error types into a simultaneous regression predicting the fluid abilities composite variable. As shown in Table 2, the regression results suggested that together the three error types accounted for 26% of the variance in gF, which is similar to what has been shown previously. In addition, the regression demonstrated that each of the error types makes a unique contribution to predicting gF. Specifically, omissions accounted for 4% unique variance, transpositions accounted for 2% unique variance, and intrusions accounted for 10% unique variance in predicting fluid abilities. This suggests that of the 26% total variance accounted for, 16% is unique to the three error types and 10% is shared by the three error types. This suggested that the predictive utility of the complex span tasks may be partially due to a combination of separate processes, each of which underlies the different error types. That is, individuals may differ in their ability to recall any item (i.e., make many omissions), they may differ in their ability to resist previous target items that are now non-target items (i.e., make many intrusions), or they may differ in their ability to recall where target items occurred within the list relative to other target items (i.e., make many transpositions). This suggests that there are a number of ways in which a participant can obtain a low score on a complex span task, each of which may add predictive utility.

Discussion

The present study investigated the patterns of errors made in complex span tasks. The results showed that the majority of errors were omissions, followed by transpositions, and then intrusions. The correct response serial position curve demonstrated large recency and primacy effects similar to those found with the continuous distractor paradigm (Bjork & Whitten, 1974). Examining each error by serial position for both input and output position suggested that the majority of output omissions occurred at the first and last serial positions, while output transpositions were localized to the middle serial positions. Furthermore, intrusions were most likely to occur at the first few serial positions and were very rare at the last serial positions. Examining each pattern of errors based on input position suggested that the last item was nearly always recalled but was likely to be recalled in an earlier serial position.

Individual differences in complex span did not occur at all serial positions for all errors, but rather were localized to a few serial positions depending on the type of error. For output omissions, the largest complex span differences were for the first few and last serial positions, for output transpositions, complex span differences only occurred for the middle three serial positions, and for intrusions, complex span differences only occurred at position one. Thus, it is not simply the case that low scoring subjects make more of each error at all serial positions, but rather the variability is localized to a few serial positions for each error. This is also the case when examining input errors by serial position. The largest complex span differences for input omissions occur at the first three serial positions and no differences occur for the very last serial position. For input transpositions, complex span differences occur at the last two serial positions. These results suggest that performance on these tasks is due to a combination of processes including the retrieval of item and order information and resistance to intrusive items from previous lists.

Although the errors in the present investigation were classified in a manner similar to previous studies that have examined errors in serial recall tasks, a number of discrepancies emerged. As noted previously, studies that have examined the patterns of errors in immediate serial recall tasks (e.g., simple span tasks with visual presentation) have generally found that output omissions and intrusions increase with serial position while transpositions show a bowed curve with the most transpositions occurring for the middle serial positions (e.g., Maylor et al., 1999). The present results, however, showed that both input omissions and intrusions declined across serial positions, output omissions and output transpositions showed a complementary pattern of errors, and input transpositions increased across serial positions. Thus, there seems to be some divergence between the pattern of errors demonstrated in verbal simple span tasks and verbal complex span tasks and point to a possible reason why simple and complex span tasks load on separate factors yet are still highly correlated (e.g., Engle et al., 1999; Unsworth & Engle, 2006).

Reconstructing serial position curves

As shown in Fig. 2, the serial position curves for correct responses show large recency and primacy effects.

What patterns of errors give rise to this curve? Integrating across the different error types can provide us with a description of task performance and allow us to reconstruct the serial position curves. The results suggest that the last item presented is nearly always recalled by all participants but not always in the correct serial position. Thus, the reason that the recency effect is not larger is because the last item has been transposed inward to position three or four. This is supported by the fact that input transpositions tend to occur most frequently for input positions four and five and output transpositions tend to occur at positions three and four. This suggests that the last items presented are being recalled fairly accurately but are moved inward. The middle serial positions show less accurate recall because they are most likely to be forgotten as shown by the input omission serial position curve. Thus, participants are forgetting the middle list items most frequently and instead of leaving the spaces blank, they are moving the last items, which are highly accessible, inward.

The primacy effect occurs because the first presented item seems to also enjoy a special state of accessibility (perhaps due to increased attention at encoding) as demonstrated by the input omission serial position curve. However, when errors do occur at the first position, they tend to be either output omissions where the individual left the space blank, or intrusions where an item from a previous list was recalled. These two errors lead to the reduced primacy effect. Thus, the pattern of results suggest that the first and last presented items tend to have a higher accessibility than middle list items. When middle list items are forgotten, the last one or two items tend to be output before their correct serial position leading to an output omission at the last one or two serial positions. Furthermore, when the first presented item is forgotten, individuals tend to either make an output omission or an output intrusion. Items presented in the middle of the list tend to be forgotten more often than items at the endpoints, but at recall these positions are not left blank due to movement of the last presented items inward. Understanding what theoretical processes give rise to these different patterns of errors, especially input omissions and intrusions is important for understanding performance on these tasks and their relation to higher-order cognition.

Temporal–contextual retrieval in complex verbal memory span tasks

In the introduction, we suggested that a temporal–contextual retrieval process can account for performance on complex span tasks and individual differences in performance. Specifically, we suggested that individual items are associated with temporal–contextual elements (e.g., Brown et al., 2000; Glenberg et al., 1980) at encoding and that these elements change throughout the task. During retrieval, temporal–contextual cues are used to guide a search process of memory. Items are sampled based on the match between the encoding context and the retrieval context. The greater the similarity between encoding and retrieval contextual states, the higher the probability of recall will be. Hence, items presented close to the recall period will be easier to recall, while items presented further back from the recall period will be harder to recall. This is due, in part, to the fact that few items will share many contextual elements with the recall period and, thus the search set from which items are sampled from will be small. Items presented further back in time, however, will be less discriminable, and, thus the search set from which they are sampled will include many representations, leading to cue-overload and a low probability of recall. Additionally, context for these items may be more similar to the context from previous trials leading to previous list items being in the search set and intrusions. The data supported both of these notions. Specifically, the input omission serial position curves suggested that the further back an item was presented, the lower the probability of correctly recalling that item was. Furthermore, the output intrusion serial position curve suggested that intrusions were most likely to occur in the first and second serial positions, suggesting that when searching for items presented further back, context cues may activate the correct items as well as items from the immediately preceding list, leading to intrusions. Together, these results suggest that when searching for a set of unrelated items in an episodic memory task such as a complex span, individuals use temporal–contextual cues to specify items. Searching for items presented long ago leads to the use of noisy temporal–contextual cues which activate many items leading to a low probability of recall for correct items.

The results not only suggested that noisy temporal–contextual cues can lead to a low probability of recall, but they also suggested that noisy temporal–contextual cues can lead to the recall of target items that are placed in the wrong serial position (i.e., transpositions). That is, assuming a two-stage discrimination process in which target items are discriminated from non-target items and recalled at the first stage, during the second stage these items must be placed into the correct serial position. Any noise in the contextual cue during the second stage will lead to items being placed in the incorrect position. Because items presented close together will share many contextual elements, it is likely that noise in the temporal–contextual cues will lead to transpositions where items presented close together will swap positions. The transposition gradients support this notion by demonstrating that when transpositions occur, they are likely to move to immediately adjacent positions, leading to an item separation of one.

In terms of individual differences in complex span, we suggest that low span individuals do not use (or form) cues as effectively as high spans to search and retrieve items. In episodic memory tasks this inability to use cues results in low spans not being able to delimit their search sets to the same extent as high span individuals, possibly due to differences in attentional control abilities (Engle & Kane, 2004). Within the temporal–contextual retrieval account, we suggest that low spans' temporal–contextual cues are noisier than high spans' cues, which results in many items being activated by the cue and greater cue-overload for low spans compared to high spans. As with the temporal–contextual retrieval account in general, a nosier cue results in a low probability of recall for items presented far away from the recall period, an increased frequency of intrusions, and greater transposition errors. Each of these predictions were supported by the data.

Specifically, low spans demonstrated more input omissions for items temporally distant from the recall period compared to high spans, suggesting that low spans had a greater difficulty using cues to select these items than did high spans. In addition, low spans demonstrated slightly more intrusions than high spans at the first serial position, suggesting that low spans had a greater difficulty discriminating target items from interfering items than high spans. Finally, even when both high and low spans were able to correctly recall a target item, low spans were more likely to place that item in the incorrect position than were high spans. That is, as demonstrated by the transposition gradients, when both high and low spans committed a transposition error, low spans were more likely to place the item further away from the target position than high spans.

Together, these results suggest that low scorers on complex span tasks have deficiencies in using temporal–contextual cues to guide the search and retrieval process of memory. This inability to effectively use cues results in many items being activated by the cue and leads to a fairly inefficient search of memory. We (Unsworth & Engle, 2006) have previously suggested that the ability to use cues to effectively delimit the search set in memory may be one important difference between high and low complex span individuals and one reason why complex span tasks correlate well with measures of higher-order cognition. Indeed, the present regression analyses suggest that breaking down complex span scores into different error components can provide important information about performance on complex spans and their relation to higher-order cognition. For instance, the present results demonstrate that all three types of output errors have both unique and shared variability with a fluid abilities composite. This suggests that the predictive power of complex span tasks is in part due to the ability to retrieve target representations amongst many potential distractors and to correctly place those target items in the correct serial position at output. Thus, the shared variability between the different error types may reflect the ability to use temporal–contextual cues to guide the search process.

In addition, the unique variance contributed by each error type may reflect specific deficits in retrieval and decision processes that occur during recall. For instance, assume that during recall temporal–contextual cues are used to delimit a search set from which items are subsequently sampled. During sampling, three possible representations may be sampled, intact target representations, intact non-target representations (intrusions), or degraded target representations (either target or non-target representations). If a degraded representation is sampled, no response is given leading to either an omission error, or a further sampling from the search set. If an intact representation is sampled, a decision has to be made determining whether that representation is a target representation or not. If an intrusion is sampled and correctly recognized as a non-target, then an omission error will occur. However, if an intrusion is sampled and there are deficiencies in the monitoring/decision processes, then the intrusion will not be recognized as a non-target and instead will be emitted. Additionally, these decision/monitoring processes will be influenced by the criterion at which individuals are willing to emit an unsure response. Those individuals whose criterion is set low will make many intrusions. Conversely, those individuals whose criterion is set high may make few intrusions and instead make many omissions. Finally, if an intact target representation is sampled, temporal–contextual information is still needed to place the item in the correct position. Thus, it is possible that the unique variance predicted by each error type may reflect deficiencies in using temporal–contextual cues to guide the search process and place items in their correct positions, as well as decision/monitoring processes that are needed to ensure that non-target items are not emitted. At the present, these suggestions are merely speculative and more work is needed to examine exactly what processes give rise to the unique predictive power of each error type.

Throughout we have suggested that the ability to use temporal–contextual cues to guide a search process of memory is crucial to performance on complex spans. In addition, we suggest that these same retrieval abilities are necessary for accurate performance on many measures of higher-order cognition where information that has been displaced from primary memory must be retrieved from secondary memory. We should note that this framework is not specific to complex span tasks and can be potentially useful in explaining performance in a number of paradigms that have been shown to be moderately correlated (e.g., working memory period tasks, Towse et al., 2005). Indeed the very fact that the framework is based on models of immediate serial recall (simple spans) and variations of free recall attests to this. Future research is needed to better examine possible working memory differences in encoding and retrieval processes and their relation to higher-order cognitive functioning.

Appendix A

Mean proportion of errors as a function of type and complex span task

Complex span task	Error type				
	OutOm	OutTr	OutIn	InOm	InTr
Ospan	.24 (.13)	.13 (.08)	.04 (.05)	.28 (.14)	.12 (.08)
Rspan	.13 (.13)	.13 (.09)	.09 (.08)	.24 (.13)	.13 (.09)

Note. Standard deviations are shown in paraphrases. OutOm, output omissions; OutTr, output transpositions; OutIn, output intrusions; InOm, input omissions; InTr, input transpositions; Ospan, operation span; and Rspan, reading span.

Mean proportion of errors as a function of type, complex span task, and serial position for list-length five

Error type	Serial position				
Complex span task	1	2	3	4	5
Output omission					
Ospan	.44 (.34)	.45 (.35)	.34 (.30)	.17 (.26)	.47 (.38)
Rspan	.28 (.34)	.27 (.33)	.20 (.27)	.08 (.18)	.32 (.35)
Output transpositions					
Ospan	.04 (.13)	.25 (.28)	.34 (.31)	.43 (.34)	.01 (.05)
Rspan	.09 (.18)	.16 (.24)	.34 (.32)	.35 (.33)	.01 (.04)
Output intrusions					
Ospan	.06 (.15)	.07 (.17)	.06 (.13)	.05 (.12)	.01 (.05)
Rspan	.13 (.20)	.20 (.25)	.17 (.24)	.06 (.15)	.01 (.04)
Input omissions					
Ospan	.42 (.33)	.68 (.33)	.58 (.30)	.38 (.30)	.05 (.14)
Rspan	.41 (.35)	.49 (.32)	.58 (.31)	.25 (.27)	.02 (.08)
Input transpositions					
Ospan	.12 (.22)	.08 (.16)	.16 (.21)	.27 (.29)	.42 (.39)
Rspan	.10 (.18)	.14 (.22)	.13 (.20)	.25 (.28)	.32 (.35)

Note. Standard deviations are shown in paraphrases. Ospan, operation span; Rspan, reading span.

Mean proportion of transposition errors as a function of separation and complex span task for list-length five

Complex span task	Separation			
	1	2	3	4
Ospan	.75 (.36)	.11 (.21)	.02 (.08)	.01 (.05)
Rspan	.68 (.40)	.14 (.26)	.01 (.10)	.01 (.03)

Note. Standard deviations are shown in paraphrases. Ospan, operation span; Rspan, reading span.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General, 131*, 567–589.

Bellezza, F. S. (1982). Updating memory using mnemonic devices. *Cognitive Psychology, 14*, 301–327.

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior, 13*, 80–97.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6*, 173–189.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127–181.

Brown, G. D. A., Vousden, J. I., McCormack, T., & Hulme, C. (1999). The development of memory for serial order: A temporal–contextual distinctiveness model. *International Journal of Psychology, 34*, 389–402.

Carroll, J. B. (1988). Individual differences in cognitive functioning. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Steven's handbook of experimental psychology* (pp. 739–811). New York: Wiley.

Chiappe, P., Hasher, L., & Siegel, L. S. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition, 28*, 8–17.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75–84.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163–183.

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. (2004). Decline in working memory updating through ageing: Intrusion error analyses. *Memory, 12*, 75–89.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two- factor theory of cognitive control. In B. Ross (Ed.). *The psychology of learning and motivation* (Vol. 44, pp. 145–199). NY: Elsevier.

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*, 309–331.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, DC: V.H. Winston.

Estes, W. K. (1982). Learning, memory, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. New York: Cambridge University Press.

Elvevåg, B., Weinberger, D. R., & Goldberg, T. E. (2001). Short-term memory for serial order in Schizophrenia: A detailed examination of error types. *Neuropsychology, 15*, 128–135.

Glenberg, A. M. (1987). Temporal context and recency. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus centennial conference*. Hillsdale, NJ: Erlbaum.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., & Renzaglia, G. J. (1983). Studies of the long-term recency effect: Support for a contextually guided retrieval hypothesis. *Journal of Experimental Psychology: Human Learning and Memory, 9*, 231–255.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., Gretz, A. L., et al. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 355–369.

Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*, 3–15.

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*, 73–137.

Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review, 85*, 109–130.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*, 169–183.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*(1), 47–70.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217.

Kellogg, C. E., & Morton, N. W. (1999). *Revised Beta Examination* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence, 14*, 389–433.

Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences, 2*, 129–160.

Li, S., & Lewandowsky, S. (1993). Intralist distractors and recall direction: Constraints on models of memory for serial order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 895–908.

Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition, 30*, 294–301.

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

Maylor, E. A., Vousden, J. I., & Brown, G. D. A. (1999). Adult age differences in short-term memory for serial order: Data and a model. *Psychology and Aging, 14*, 572–594.

McCormack, T., Brown, G. D. A., Vousden, J. I., & Henson, R. N. A. (2000). Children's serial recall errors: Implications for theories of short-term memory development. *Journal of Experimental Child Psychology, 76*, 222–252.

Raven, J. C., Raven, J. E., & Court, J. H. (1998). *Progressive matrices*. Oxford, England: Oxford Psychologists Press.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126*, 211–227.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language, 39*, 418–436.

Tehan, G., Hendry, L., & Kocinski, D. (2001). Word length and phonological similarity effects in simple, complex, and delayed serial recall tasks: Implications for working memory. *Memory, 9*, 333–348.

The Psychological Corporation. 1999. *Wechsler Abbreviated Scale of Intelligence.* The Psychological Corporation, San Antonio, TX.

Towse, J. N., Hitch, G. J., Hamilton, Z., Peacock, K., & Hutton, U. M. Z. (2005). Working period: The endurance of mental representations. *Quarterly Journal of Experimental Psychology, 58A*, 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 & Language, 54*, 68–80.

Unsworth, N., & Engle, R. W. (2005a). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory & Cognition, 33*, 213–220.

Unsworth, N., & Engle, R. W. (2005b). Working memory capacity and fluid abilities: Examining the correlation between operation span and raven. *Intelligence, 33*, 67–81.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1302–1321.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 104*, 442–452.