Working Memory Span: The Effect of Prior Learning
Author(s): Cindy Lustig and Lynn Hasher
Source: *The American Journal of Psychology*, Vol. 115, No. 1 (Spring, 2002), pp. 89-101
Published by: University of Illinois Press
Stable URL: http://www.jstor.org/stable/1423675
Accessed: 25-10-2017 18:49 UTC

# Working memory span: The effect of prior learning

CINDY LUSTIG AND LYNN HASHER
Duke University

Recent work suggests that working memory span (WMS) tasks are not simple measures of the capacity to simultaneously store and process new information. Instead, these measures may be influenced by numerous factors, including proactive interference (PI). The current study examined whether WMS, like other memory tasks, is also influenced by PI from prior memory experiments. Experimentally experienced and naive participants completed a speaking span task. Span scores were lower for experienced than for naive participants, but other cognitive scores were not. In combination with other work, these results suggest that WMS estimates are not pure measures of capacity and may be partially determined by PI.

Working memory capacity is widely thought of as the mental workspace available for the simultaneous processing and storage of information (Baddeley, 1986; Daneman & Carpenter, 1980; Just & Carpenter, 1992). It is this capacity (in contrast to the capacity for simple, passive storage as measured by traditional span measures) that various working memory span (WMS) measures were devised to assess (Daneman & Merikle, 1996). Both individuals and groups are thought to differ in this fundamental cognitive capacity. Consistent with this view, WMS tasks predict performance on a range of skills including reading comprehension, text recall, and reasoning (see Daneman & Merikle, 1996, for a review). WMS tasks also predict performance across the life span, from childhood to at least middle age (Siegel, 1994). However, WMS measures are complex, and as a result it has been difficult to determine the cognitive components that influence the size of span scores and are responsible for their predictive power (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985; Tirre & Pena, 1992; Waters & Caplan, 1996).

The current study investigates the impact of participation in previous memory experiments on WMS. Prior experience of this sort has large, detrimental effects on other memory tasks. For example, for a series of lists learned over consecutive days, retention of the current day's list decreases as a function of the number of previous lists learned for both 5- and 48-hour retention intervals (Greenberg & Underwood, 1950).

AMERICAN JOURNAL OF PSYCHOLOGY
Spring 2002, Vol. 115, No. 1, pp. 89–101

Likewise, 24-hour delayed recall of a single, laboratory-learned list is dramatically affected by the number of lists learned in prior experiments. Recall of the target list declines from 75% to 25% as the number of previously learned lists increases, even when the target list uses words different from those in the previous lists (Underwood, 1957; see also Keppel, Postman, & Zavortink, 1968; Zechmeister & Nyberg, 1982). This decline in performance with increasing lists usually is attributed to proactive interference (PI), the generally disruptive effect of prior learning on the ability to retrieve more recently learned information.

Several lines of work now suggest that PI also plays a major role in determining WMS scores. First, individuals and groups particularly vulnerable to the effects of PI (e.g., poor readers, young children, older adults) also perform poorly on span tasks compared with less PI-vulnerable young adults (Chiappe, Hasher, & Siegel, 2000; Dempster, 1992). Likewise, young adults with high WMS scores perform better on PI tasks than do those with low WMS scores (Dempster & Cooney, 1982; Rosen & Engle, 1998).

Also, a task analysis reveals that many WMS tasks unintentionally encourage the buildup of PI within the task because they consist of a series of study–recall trials that start with the shortest lists and end with the longest (Dempster, 1992; May, Hasher, & Kane, 1999; Whitney, Arnett, Driver, & Budd, 2001). Because PI can build up quickly across trials within a task (Keppel & Underwood, 1962), items studied and recalled on early trials of WMS tasks can disrupt recall of items from later trials. In most WMS tasks the long trials essential to obtaining a high span score are also the later trials and thus are those most likely to be disrupted by PI. The potential for PI to disrupt performance on these longest, latest, and most important trials means that WMS measures may be heavily influenced by participants' susceptibility to PI.

Direct evidence of the buildup of PI within WMS tasks was reported by May et al. (1999), who showed that reversing the usual order of administration (so that the longest lists came first rather than last) raised the span scores of older adults (who are differentially vulnerable to the effects of PI) to the level of college students' scores (see also Lustig, May, & Hasher, 2001). In addition, reducing the amount of PI in WMS tasks reduces the ability of WMS to predict performance on other measures (Lustig et al., 2001; see Dempster & Corkill, 1999). Finally, WMS scores are influenced by factors affecting the buildup of PI, such as the distinctiveness or similarity between the to-be-remembered items or between the to-be-processed and to-be-stored items (Conrad & Hull, 1964; May et al., 1999; Postman & Underwood, 1973; Shah & Miyake, 1996; Young & Supa, 1941).

Thus, WMS tasks appear to be vulnerable to PI as it builds up within

a task. This raises the question of whether WMS scores are also vulnerable to PI from previous learning obtained in prior memory experiments, as is known to be the case for other memory tests (Underwood, 1957; see Keppel et al., 1968; Zechmeister & Nyberg, 1982). To address this question, young adults initially naive to psychology experiments were given a WMS task either as their very first experimental task or after having participated in two prior experiments that were held on different days. Based on research showing within-task PI in WMS and the effects of between-experiment PI on other memory tests, we expected that participation in prior experiments would lead to lower WMS scores.

## EXPERIMENT

### METHOD

#### Participants

Seventy-two (mean age = 20.1 years, *SD* = 1.6) Duke University undergraduates who had not been in any psychology experiments for at least 2 years served as participants in this study. All participants volunteered to be in multiple studies over a 3-day period and were paid $5 for each session. Participants were assigned to one of two groups: experimentally naive ($n$ = 36) and experimentally experienced ($n$ = 36). The naive participants did the span task on Day 1. The experienced participants took part in two unrelated memory studies on Days 1 and 2 and then returned for the span task on Day 3.

#### Materials and procedure

The WMS task we used was the speaking span (see Daneman & Green, 1986).[1] Each participant silently read a series of words presented on a computer screen for 1 s per word. At the end of a series (which in this instance ranged in size from two to four words), an asterisk appeared to cue the participant to produce one sentence for each word in the series. If the participant did not begin within 2 s after the asterisk, a tone sounded to remind the participant to do so. Words could be used in any order, with the exception that the last word in a series could not be used to generate the first sentence. The trial ended when the participant correctly produced sentences for each word in the series or indicated that he or she could not remember any more words. The next trial commenced when the participant indicated a readiness to begin. Participants were warned when the size of the series was about to change (e.g., from two to three words per trial).

Set sizes of two, three, and four words were used in this task, with three trials given at each set size. Half the participants completed the trials in an ascending order, starting with sets of two items and ending with sets of four; the other half completed trials in descending order, starting with sets of four items and ending with sets of two.[2] Twenty-seven medium frequency, two-syllable, unrelated, concrete nouns served as stimuli for the span task. There were six

different arrangements of words across set sizes, and each was used equally often in the two conditions.

Stimuli were presented on the monitor of an IBM-compatible personal computer using Micro Experimental Laboratory (MEL) software. All participants also completed Version 3 of the Extended Range Vocabulary Test (ERVT; Educational Testing Service, 1976) and the Digit Symbol subtest of the Wechsler Adult Intelligence Scale–Revised (WAIS-R, Wechsler, 1981). The ERVT and Digit Symbol were included to test for any differences in motivation or ability (vocabulary and speed) between the two groups. The ERVT is a very difficult vocabulary test, and it was used here as a potentially sensitive way to determine whether our two experimental groups differed as a result of our manipulation of experience. Because participants were randomly assigned to the two conditions, our expectation was that the naive and experienced groups would not differ on these two additional (cognitive, but nonmemory) measures.

Participants assigned to the experienced group completed two other experiments in the days before the span experiment. These two experiments were conducted in the same laboratory as the critical span experiment but by different experimenters and in different rooms within the laboratory. On the first day, participants listened to true and false trivia statements spoken by a male or female speaker, then performed a recognition test in which they made judgments as to either the truth or the source (male or female speaker) of the sentence. On the second day, they studied a single list of words made up of three blocked sets of related words (taken from the "cold," "needle," and "sleep" lists from the DRM lists; Deese, 1959; Roediger & McDermott, 1995) and then attempted to recall the studied words. Neither of the two prior experiments (the truth or source judgment task and the list recall task) completed by the experienced participants shared items with the critical span experiment, but both prior experiments used verbal materials and tested for memory, as did the critical span experiment.

## RESULTS

To determine whether the naive and experienced groups differed in general ability or motivation, performance of the two groups on the Digit Symbol and vocabulary (ERVT) tasks was compared. The experienced group had a marginally higher vocabulary score, $M = 28.1$, $SD = 6.7$, than the naive group, $M = 25.3$, $SD = 6.9$, $t = 1.76$, $p = .08$, but not discriminably different Digit Symbol scores, $M = 76.6$, $SD = 11.5$ for experienced and $M = 76.1$, $SD = 8.6$ for naive participants, $t = .24$, $p = .81$. There was no indication that the experienced participants were less knowledgeable about word meanings (and so possibly of word usage) or that they were less motivated than the naive participants. If anything, the results on these general performance tests are slightly biased *against* the prediction of lower span scores for the experienced group.

The span data were scored using three different methods. The first

was similar to that developed by Daneman and Carpenter (1980). This scoring procedure (called "standard") determined a participant's span as the highest set size at which he or she correctly recalled all words for the majority (two out of three) of trials. Partial credit (of .5) was received if all the items for one trial (out of three) at that set size were correctly recalled. Scoring began with set size 2. Criteria (at least two out of three trials correct) for set size 2 had to be obtained before continuing to set size 3, and criteria for set size 3 had to be met before continuing to set size 4. However, this scoring method is arbitrary, and the requirement that criteria for one set size must be met before the next set size can be scored results in its ignoring a great deal of data about participants' performance at the larger set sizes, limiting the range of scores and making it an insensitive measure (Daneman & Green, 1986). Therefore, we also used two other, more sensitive scoring methods. The second scoring method was a lenient version of the "standard" procedure, in which criteria for the previous set size did not have to be met before scoring for the next largest set size could begin. The final method used was an items span measure (Engle, Nations, & Cantor, 1990) that allows a wider range of scores. The items span measure calculates span as the total number of words correctly recalled in fully correct trials and has become increasingly popular (e.g., Chiappe et al., 2000; Engle et al., 1990; May et al., 1999).

The central question was whether span scores for the experienced group differed from those for the naive group. Means for all three scoring methods suggest that the experienced group had smaller spans than the naive group (Table 1). Although the difference did not reach statistical significance for the standard scoring method, $t = 1.18$, $p = .24$, it did for the more sensitive lenient, $t = 3.46$, $p = .0009$, and items methods, $t = 2.86$, $p = .006$. Experimentally naive participants had higher span scores than did experimentally experienced participants. It is worth noting that effect size increased with the sensitivity of the measure, $d = 0.66$ ("medium"), for the lenient method, $d = 1.47$ ("large") for the items method (Cohen, 1988). In addition, there was a trend for naive participants to produce fewer intrusions (sentences that did not contain any words from the current trial) than experienced participants, .17 versus

Table 1. Mean span scores and standard deviations (in parentheses) for experienced and naive participants

| | Standard method | Lenient method | Items method |
|---|---|---|---|
| Experienced | 2.54 (0.61) | 2.58 (0.69) | 8.78 (4.01) |
| Naive | 2.72 (0.68) | 3.13 (0.64) | 11.97 (5.38) |

.44, $t = 1.56$, $p = .12$, $d = 0.33$ (a small effect size by the Cohen, 1988, guidelines), although intrusion rates in general were very low.

To more closely examine the WMS performance of the naive and experienced groups, we used a 2 (group) × 3 (set size) ANOVA to analyze the number of trials correctly completed at each set size, with a maximum of three (Table 2). Experimentally naive participants correctly recalled more sets (had more completely correct trials) overall than did experienced participants, $F(1, 70) = 5.78$, $p = .02$. The group × set size interaction was reliable, $F(2, 140) = 3.09$, $p = .049$; naive participants correctly completed more trials at the largest set size than did experienced participants, $F(1, 70) = 12.07$, $p = .0009$. Thus, not only did naive participants perform better on the span task overall, but the benefit of naiveté was greatest on the longest, most difficult trials.

To ensure that our results reflected memory differences between the two groups and not differences in strategy, we recorded and transcribed the sentences of a subset of 10 participants from each group. Two independent judges, blind to condition, then rated each sentence on a 1–7 scale for how interesting the sentence was (1 = *not interesting,* 7 = *extremely interesting*) and for how unusual or creative the sentence was (1 = *not unusual or creative,* 7 = *extremely unusual or creative*). There was no difference between the naive and experienced groups in terms of how interesting (1.9 vs. 1.7) or unusual or creative (1.4 vs. 1.4) their sentences were, $p > .20$ in both cases. Also, the sentences produced by the two groups did not differ in the number of words per sentence (5.9 vs. 6.1), $t < 1$. Thus, the naive and experienced groups did not differ in the quality of the sentences they produced but only in the number of words they were able to recall and use to generate sentences.

## DISCUSSION

Prior laboratory experience lowered the working memory scores of experimentally experienced participants relative to similar participants who lacked those same experiences. How are these results to be explained? We first consider the possibility that despite random assignment to conditions, something about the two groups of participants

Table 2. Mean number of correct trials and standard deviations (in parentheses) at each set size for experienced and naive participants

| | Set size 2 | Set size 3 | Set size 4 | Total sets |
|---|---|---|---|---|
| Experienced | 2.47 (0.8) | 1.05 (0.8) | 0.17 (0.4) | 3.69 (1.5) |
| Naive | 2.50 (0.6) | 1.36 (0.9) | 0.72 (0.9) | 4.58 (1.7) |

influenced the outcome. Two possibilities seem to be likely candidates: differences in overall ability or differences in motivation. Either of these possibilities could lead to a general decline in performance rather than specific impairment of memory caused by proactive interference.

Several aspects of the data argue against this general ability or motivation explanation of the memory span differences between the groups. First, there was no reliable difference on the Digit Symbol subtest of the WAIS-R (Wechsler, 1981), a widely used test of intellectual ability. Second, scores on the vocabulary test actually *favored* the experienced group, and it is this group that had the reduced span scores. The vocabulary test used here, the ERVT, is part of an experimental intelligence battery, the Educational Testing Service's Kit of Factor-Referenced Tests (1976), deliberately created to be sensitive to differences between people in the upper ranges of ability. If anything, then, the suggestion is that for these two measures of general ability, the advantage, however small, is in favor of the experimentally experienced participants or those with the lower span scores.

The tendency for the experienced group to have a small advantage in ability is particularly noteworthy given the recent suggestion that general ability and WMS are correlated (Dempster & Corkill, 1999; Engle, Tuholski, Laughlin, & Conway, 1999). This suggestion predicts an outcome of higher span scores for the higher-ability group, which in this instance is the experimentally experienced group. This prediction is in direct contrast with what was found. Likewise, a lack of motivation on the part of the experienced group should also have led to lower scores not just on the span measure but on all measures, including the Digit Symbol and ERVT, and these data show the opposite pattern. Ability and motivation differences do not seem to be reasonable explanations of the WMS differences found to be associated with experience.

Further evidence that the WMS differences between the naive and experienced groups did not result from differences in motivation or ability comes from the observation that only the memory aspect of the task suffered as a result of prior experience. If the experienced group were lacking in motivation or were fatigued by participating in three experiments in three days, we might have detected differences between the types of sentences this group produced and those produced by people tested on the first day. However, the two groups did not differ in either the quality or the length of the sentences they produced; the two groups differed only in the number of span task words they correctly recalled and used to generate sentences and the number of extraneous words they incorrectly intruded during recall.

In summary, then, evidence both external and internal to the span task argues against the suggestion that the naive participants' WMS score

advantage was caused by any advantage in ability or motivation. External measures of general ability and performance (Digit Symbol, ERVT) either showed no differences between the groups or tended to favor the experienced group. Internal to the span measure, aspects of performance unrelated to memory were the same for the naive and experienced groups; only the aspects of span performance related to the correct retrieval of the studied words changed as a result of participation in prior experiments.

Other, older work shows that prior sessions in a laboratory can disrupt retrieval of the most recently learned information, even when there is a substantial delay (5, 24, or 48 hours) between the prior session and new learning (Greenberg & Underwood, 1950; Underwood, 1957; see Keppel et al., 1968; Zechmeister & Nyberg, 1982). We acknowledge that our methods are somewhat different from that used in this older work, where the prior sessions typically consisted of tasks (e.g., paired-associate or serial learning) identical to that used in the critical session. In our investigation the previous tasks (truth or source judgment and list recall) were not identical to the critical task (production span). However, we note that all three tasks were memory tests and used verbal materials. In particular, both the source memory task and the critical span task involved sentences, and both the list recall task and the critical span task involved recalling lists of words. A number of experiments have shown that retroactive interference (detrimental effects of *subsequent* learning) can occur even if the interfering materials are presented in a task different from that used to present the critical materials (Bird & Weaver, 1975; Gibson & Gibson, 1934; Lehr, Frank, & Mattison, 1972; Posner & Konick, 1966). To our knowledge, the current study is the first demonstration of proactive interference (detrimental effects of *previous* learning) across different tasks.

It will be important to replicate this finding of proactive interference from nonidentical previous tasks in future experiments. Nonetheless, the reduction in WMS seen here as a result of prior experience is consistent with other recent studies suggesting that WMS tasks are vulnerable to interference. For example, variables that influence PI, including similarity between competitors and number of potential competitors, are now known to affect working memory span tasks much as they do other long- and short-term memory tasks (see Conrad & Hull, 1964; May et al., 1999; Postman & Underwood, 1973; Shah & Miyake, 1996; Young & Supa, 1941). Furthermore, manipulations intended to reduce the impact of PI have been shown to boost span scores substantially (Lustig et al., 2001; May et al., 1999). Thus, this new finding of differences between naive and experienced participants is consistent with recent evidence on the vulnerability of WMS scores to interference.

The demonstrated impact of interference on WMS raises the question of what WMS tasks measure and thus the basis on which WMS predicts performance on measures of text comprehension, reasoning, and memory (see Daneman & Merikle, 1996). For some time now, working memory capacity as measured by WMS tasks has been considered a cognitive "primitive" of fundamental importance in many areas of cognition (e.g., Just & Carpenter, 1980; Kyllonen & Christal, 1990). A common interpretation is that WMS measures the general capacity to simultaneously store and operate on information (Baddeley, 1986; Daneman & Carpenter, 1980). Others consider WMS a measure of the capacity for "activation" (Anderson, Reder, & Lebiere, 1996; Just & Carpenter, 1992) or of the amount of attention available to engage controlled rather than automatic attentional processes (Rosen & Engle, 1997).[3] An alternative viewpoint is that the cognitive primitive underlying performance on both WMS and many other tasks is the inhibitory-based ability to delete irrelevant information from working memory, thus reducing competition (Hasher, Zacks, & May, 1999; Lustig et al., 2001; May et al., 1999; Zacks & Hasher, 1994). Competition may stem from previous occasions in a laboratory, as the present study suggests, from a previous list of items (see Postman & Underwood, 1973) or even, as in WMS tasks and other short-term memory tasks, from previous sets within a single series (Keppel & Underwood, 1962; May et al., 1999). In any event, it is increasingly clear that WMS scores do not measure a single construct such as capacity but instead are determined by multiple factors, at least one of which is proactive interference.

Independent of the exact nature of the cognitive primitive that may underlie performance on WMS and other cognitive tasks, the present findings raise serious questions about the use of WMS as a measure of working memory capacity both within and across subject populations, at least if one considers capacity to be a stable group and individual differences variable. Two groups that differed only in their degree of experimental experience and were identical on all other measures nonetheless had disparate span scores, with lower scores being obtained by the experienced group. These data suggest that rather than being a stable characteristic of an individual or group, WMS can be influenced by participation in prior experiments in addition to other factors. We note that the influence of prior experiments may be especially problematic when attempting to measure potential capacity differences between special populations such as older and younger adults or between patient and control groups. The same limited set of volunteers from the special population may participate in many more experiments than volunteers from the reference population because the reference population typically is larger, and new volunteers from it are easier to obtain.

In summary, the current study found that experienced participants, who participated in prior verbal memory experiments, obtained lower scores on a critical WMS task than did naive participants who did not participate in prior experiments. Because the naive and experienced groups were identical on all other measures of ability and motivation, PI from prior experimental participation is the most likely explanation of the experienced participants' low WMS scores. This explanation receives further support from recent findings showing that interference has a strong effect on WMS (Chiappe et al., 2000; Dempster, 1992; May et al., 1999; Shah & Miyake, 1996) as well as an older literature demonstrating that other memory tasks are also vulnerable to PI from prior experimental sessions (Greenberg & Underwood, 1950; Underwood, 1957). These findings raise theoretical questions as to the nature of working memory capacity as measured by WMS and important practical concerns about the use of WMS as an index of capacity across groups and individuals who may differ in experimental experience.

## Notes

Lynn Hasher is now at the Department of Psychology, University of Toronto, and Rotman Research Institute of Baycrest Centre. Cindy Lustig is now at Washington University in St. Louis.

The research reported here was supported by the National Institute on Aging grants AG 2753 and 4306. Our thanks to Karen Z. H. Li, Tamara Rahhal, Cynthia P. May, Darin Wiechmann, Alisha Ruddy, Jaime Kelley, Jennie Goldstein, Carolyn Yoon, and Michelle Li, whose helpfulness at various times on this project is greatly appreciated.

Correspondence about this article should be addressed to Cindy Lustig, Department of Psychology, Box 1125, Washington University, St. Louis, MO 63130 (E-mail: clustig@artsci.wustl.edu) or Lynn Hasher, Department of Psychology of the University of Toronto, 100 St. George Street, Toronto, Ontario, M5S 3G3, Canada (E-mail: hasher@psych.utoronto.ca). Received for publication May 10, 2000; revision received September 29, 2000.

1. The speaking span task is not as commonly used as the reading span task (Daneman & Carpenter, 1980) but is highly correlated with the reading span task and is as predictive as the reading span task is of cognitive performance across various populations (Daneman, 1991; Daneman & Green, 1986; McCutchen, Covill, Hoyne, & Mildes, 1994).

2. Previous work (Lustig et al., 2001; May et al., 1999) shows that a simple directional manipulation affects the span estimates of older adults but not young adults. Also, unpublished data from two laboratories (C. May and L. Hasher) suggest that for the directional manipulation to be successful in influencing estimates of WMS, a sufficient number of lists at each set size (e.g., 5 rather than 3) must be given so that an adequate amount of PI can build up across trials. Thus, we did not expect (or find) that the direction of adminis-

tration would influence the span estimates of the young adults who served as participants in the current experiment.

3. The controlled attention and inhibitory views are in many respects mirror images of each other. In the controlled attention view, high-capacity individuals (as identified by high WMS scores) are better able to engage processes such as suppression to resist interference (Rosen & Engle, 1998). By the inhibitory view, the degree to which an individual is affected by interference affects WMS score and thus the estimate of capacity. We take the present results, in which a causal manipulation of interference had a direct impact on WMS scores, as supportive of the inhibitory view. Previous findings that interference manipulations affect WMS score (May et al., 1999; Shah & Miyake, 1996) also support this view, as do the correlational findings of Lustig et al. (2001). However, the larger issue is a complex one that warrants further research.

## References

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30,* 221–256.

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

Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language, 24,* 119–131.

Bird, C. P., & Weaver, G. E. (1975). Interpolated task characteristics and interference in short-term memory. *Bulletin of the Psychonomic Society, 6,* 93–95.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55,* 429–432.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research, 20,* 445–464.

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

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

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

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22.

Dempster, F. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review, 12,* 45–75.

Dempster, F., & Cooney, J. B. (1982). Individual differences in digit span, susceptibility to proactive interference, and aptitude/achievement test scores. *Intelligence, 6,* 399–416.

Dempster, F. N., & Corkill, A. J. (1999). Interference and inhibition in cognition and behavior: Unifying themes for educational psychology. *Educational Psychology Review, 11,* 1–88.

Educational Testing Service. (1976). Kit of factor-referenced tests. Princeton, NJ: Author.

Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is "working memory capacity" just another name for word knowledge? *Journal of Educational Psychology, 82,* 779–804.

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.

Gibson, E. J., & Gibson, J. J. (1934). Retention and the interpolated task. *American Journal of Psychology, 46,* 603–610.

Greenberg, R., & Underwood, B. J. (1950). Retention as a function of stage of practice. *Journal of Experimental Psychology, 40,* 452–457.

Hasher, L., Zacks, R., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance* (17th ed., pp. 653–675). Cambridge, MA: MIT Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329–354.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Keppel, G., Postman, L., & Zavortink, B. (1968). Studies of learning to learn: VIII. The influence of massive amounts of training upon the learning and retention of paired-associate lists. *Journal of Verbal Learning and Verbal Behavior, 7,* 790–796.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in the short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 2,* 153–161.

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

Lehr, D. J., Frank, R. C., & Mattison, D. W. (1972). Retroactive inhibition, spontaneous recovery, and type of interpolated learning. *Journal of Experimental Psychology, 92,* 232–236.

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

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition, 27,* 759–767.

McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology, 86,* 256–266.

Posner, M. I., & Konick, A. F. (1966). On the role of interference in short-term retention. *Journal of Experimental Psychology, 72,* 221–231.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1,* 19–40.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

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

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

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4–27.

Siegel, L. S. (1994). Working memory and reading: A life-span perspective. *International Journal of Behavioral Development, 17,* 109–124.

Tirre, W. C., & Pena, C. M. (1992). Investigation of functional working memory in the reading span test. *Journal of Educational Psychology, 84,* 462–472.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64,* 49–60.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103,* 761–772.

Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale*–Revised. New York: Psychological Corporation.

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

Young, C. W., & Supa, M. (1941). Mnemonic inhibition as a factor in the limitation of the memory span. *American Journal of Psychology, 54,* 546–552.

Zacks, R. T., & Hasher, L. (1994). Directed ignoring: Inhibitory regulation of working memory. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 241–264). San Diego: Academic Press.

Zechmeister, E. B., & Nyberg, S. E. (1982). *Human memory: An introduction to research and theory.* Monterey, CA: Brooks/Cole.