Journal of Experimental Psychology:
Learning, Memory, and Cognition
2010, Vol. 36, No. 2, 471–483


0278-7393/10/$12.00 DOI: 10.1037/a0018405

# Similarities and Differences Between Working Memory and Long-Term Memory: Evidence From the Levels-of-Processing Span Task

Nathan S. Rose, Joel Myerson, Henry L. Roediger III, and Sandra Hale
Washington University in St. Louis

Two experiments compared the effects of depth of processing on working memory (WM) and long-term memory (LTM) using a levels-of-processing (LOP) span task, a newly developed WM span procedure that involves processing to-be-remembered words based on their visual, phonological, or semantic characteristics. Depth of processing had minimal effect on WM tests, yet subsequent memory for the same items on delayed tests showed the typical benefits of semantic processing. Although the difference in LOP effects demonstrates a dissociation between WM and LTM, we also found that the retrieval practice provided by recalling words on the WM task benefited long-term retention, especially for words initially recalled from supraspan lists. The latter result is consistent with the hypothesis that WM span tasks involve retrieval from secondary memory, but the LOP dissociation suggests the processes engaged by WM and LTM tests may differ. Therefore, similarities and differences between WM and LTM depend on the extent to which retrieval from secondary memory is involved and whether there is a match (or mismatch) between initial processing and subsequent retrieval, consistent with transfer-appropriate-processing theory.

*Keywords:* short-term memory, working memory, secondary memory, long-term memory, levels of processing

The construct of working memory (WM) has become central to theories that attempt to understand a wide range of cognitive functions. Individual differences in WM capacity have been found to be related to numerous areas of higher order cognition including language comprehension (Daneman & Carpenter, 1980; Gathercole & Baddeley, 1993; Just, Carpenter, & Keller, 1996), mathematics (Hitch, 1978; Logie & Baddeley, 1987), reasoning (Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Kyllonen & Christal, 1990), and complex learning (Kyllonen & Stephens, 1990; Shute, 1991). The ubiquity of associations found between WM capacity and higher order cognitive function has led some to even refer to it as "the hub of cognition" (Haberlandt, 1997, p. 212). There is no consensus in the literature, however, as to exactly what the construct of WM represents, and how it should be distinguished from other memory constructs, that is, short-term memory (STM) and long-term memory (LTM).

## On the Distinction Between Memory Systems and Some Messy Terminology

Early models of memory made clear distinctions between short-term and long-term stores. In 1890, based purely on introspection, William James distinguished between *primary* and *secondary* memory. Primary memory reflects the current contents of consciousness, whereas secondary memory consists of memory of the distant past that must be brought back into consciousness by a retrieval process. This distinction was maintained in influential memory models developed by experimental psychologists (e.g., Atkinson & Shiffrin, 1968; Waugh & Norman, 1965) and is supported by a substantial body of evidence, including observations of neuropsychological cases (Milner, 1966; Shallice & Warrington, 1970), and patterns of serial position effects (e.g., Glanzer, 1972; Murdock, 1962).

The construct of WM evolved to capture a more dynamic STM system than that denoted by the construct of primary memory (Baddeley & Hitch, 1974). As Baddeley (1986) pointed out, "The term working memory implies a system for the temporary holding and manipulation of information during the performance of a range of cognitive tasks such as comprehension, learning, and reasoning" (pp. 33–34).

How theories distinguish between memory systems is complicated by the lack of clarity and consistency in the terminology that researchers have used over the years. Craik and Lockhart (1972) recommended that the terms used when referring to theoretical constructs be clearly distinguished from the procedures used for measuring those constructs (see also Tulving, 1983a, 1983b, 2000). They suggested, for example, that the terms *STM* and *LTM* be used to refer to tasks and procedures (e.g., immediate and delayed tests) that emphasize the involvement of the primary and secondary memory systems, respectively.

Where then does the term WM fit in? Many researchers have tried to incorporate it within previously conceived memory systems simply by combining terms—that is, short-term working memory (cf. Neath, Brown, Poirier, & Fortin, 2005) and long-term working memory (Ericsson & Kintsch, 1995)—although whether

Nathan S. Rose, Joel Myerson, Henry L. Roediger III, and Sandra Hale, Department of Psychology, Washington University in St. Louis.

We thank Linden Weiswerda for assistance in participant testing and data scoring, and Gus Craik for helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Nathan S. Rose, Washington University, Department of Psychology, Campus Box 1125, St. Louis, MO 63130. E-mail: nrose@wustl.edu

such compound terms refer to procedures, constructs, or functions is often unclear. Despite being originally developed out of the concept of a system for STM, the concept of WM, as instantiated in several recent models, is intimately related to LTM (Cowan, 1999; Ericsson & Kinstch, 1995; Oberauer, 2002; Unsworth & Engle, 2007a). Indeed, Baddeley (2000) recently suggested that WM provides an interface between STM and LTM, and has modified his original model by adding a new component, the episodic buffer, to accommodate the way in which WM and LTM interact.

Some researchers (e.g., Cowan, 1999) have conceptualized the relation between WM and LTM as one in which WM is actually a subset (i.e., the currently activated portion) of LTM. According to Cowan's (1999) embedded-process model of WM, the capacity of the focus of attention (a construct similar to William James's, 1890, description of primary memory) is limited to four chunks of information, and all other items in WM reside within, and must be retrieved from, the activated portion of LTM. Similar to Cowan (1999), Oberauer (2002) has proposed a concentric model of WM. In Oberauer's model, information in memory may exist in different states of accessibility. A limited number of chunks may be within a state of direct access and other, recently activated information remains in a passive state of readiness within LTM. Importantly, because LTM is not constrained by the same capacity limits as the focus of attention or the region of direct access, reliance upon LTM may appear to expand the capacity limitations of WM (Cowan, 1999; Ericsson & Kintsch, 1995; Oberauer, 2002).

More recently, Unsworth and Engle (2006, 2007a, 2007b) have suggested that, in addition to a primary memory component, many immediate memory tasks (e.g., WM span tasks) also involve retrieval from a secondary memory component. For example, complex span tasks (e.g., operation span) require participants to perform a secondary processing task (e.g., solving math problems) interleaved between presentation of items to be immediately recalled. According to Unsworth and Engle's dual-component model, such secondary tasks require that participants temporarily switch attention away from maintaining items in primary memory. Thus, at least some of these items must be retrieved from secondary memory (Unsworth & Engle, 2007a). In contrast, simple span tasks (e.g., digit span) capture the ability to maintain a list of items in, and report them directly from, primary memory. This is the case unless the list exceeds approximately four chunks, at which point both primary and secondary memory abilities are involved (Unsworth & Engle, 2006). Taken together, these recent models (Baddeley, 2000; Cowan, 1999; Oberauer, 2002; Unsworth & Engle, 2007a) reflect a growing consensus that WM tasks are not solely dependent on either system, thus placing WM at the intersection of STM and LTM, or the primary and secondary memory systems (see also Mogle, Lovett, Stawski, & Sliwinski, 2008; Unsworth, 2009).

The present study addresses the relation between WM and LTM by comparing how they are affected by a manipulation known to affect LTM: levels of processing (LOP; Craik & Lockhart, 1972). That is, one characteristic of LTM is that it is highly sensitive to the qualitative depth with which memory items are processed when they are initially encoded. For example, it is well established that conceptual (semantic) processing at encoding leads to superior long-term retention on most episodic memory tests, relative to processing that focuses on more structural aspects of the memory items, such as phonological or visual features (Craik & Lockhart, 1972; Craik & Tulving, 1975; Hyde & Jenkins, 1973; Roediger, Gallo, & Geraci, 2002). Thus, if the performance of a WM span task depends in part on retrieval from secondary memory, it would seem to follow that the type of processing at encoding should affect performance on a WM span task in the same way that it affects delayed memory tests. More specifically, if one designs a WM span task in which the secondary task involves varying LOP, then one might expect deeper (semantic) processing to result in better immediate recall (i.e., increased WM span) than if the secondary task focuses attention on more structural aspects of the memory items (e.g., phonological or visual features).

Secondary processing tasks on WM span tasks typically reduce spans below levels observed on simple storage tasks (e.g., Engle et al., 1999; Hale, Myerson, Rhee, Weiss, & Abrams, 1996; Unsworth & Engle, 2007b). Having to perform a secondary processing activity may disrupt the ability to actively maintain a list of to-be-remembered items by interrupting rehearsal (Baddeley, 1986) or by displacing the items from the focus of attention (Cowan, 2005). It should be noted, however, that the secondary tasks used with most WM span procedures (e.g., operation span) do not manipulate the way in which the to-be-remembered information is processed. In fact, we know of only one study (Mazuryk & Lockhart, 1974) that has had participants perform an immediate memory task similar to present-day simple and complex WM tasks with conditions that manipulated the depth of processing of the to-be-remembered items.

Mazuryk and Lockhart (1974) presented participants with five words for immediate free recall. Participants were instructed that, following presentation of each to-be-remembered word, they were to process that word in one of four different ways, depending on the condition: either rehearse the word silently, rehearse the word overtly, generate a rhyme (shallow processing), or generate a semantic associate (deep processing). The two rehearsal conditions both produced near-perfect immediate recall, which was considerably better than performance for the two conditions with a secondary processing demand (rhyme or semantic generation). Interestingly, the two conditions that most closely resembled a complex WM span task with deep versus shallow processing requirements failed to show an LOP effect. That is, generating a semantic associate (semantic processing) did not produce significantly better immediate recall than generating a rhyme (phonological processing). After performing several trials of immediate recall, participants were given a delayed free recall or recognition test on all of the studied words. Semantic processing, despite producing immediate recall performance that was equivalent to phonological processing and worse than either covert and overt rehearsal, resulted in performance superior to all other conditions on both delayed recall and delayed recognition tests.

If it is true that performance of WM tasks involves retrieval from secondary (long-term) memory, then one might expect processing tasks that manipulate the depth to which memory items are processed to affect performance on WM and LTM tasks in the same way. On the other hand, if the nature of retrieval differs for WM and LTM tasks, then an LOP manipulation may affect performance on the two tasks differently. The dissociation between LOP effects on immediate recall and LTM shown by Mazuryk and Lockhart (1974) is clearly consistent with the latter interpretation.

Although recent research suggests retrieval from secondary memory is involved in performance of WM tasks, task dissociations between WM and LTM tasks of the sort shown by Mazuryk and Lockhart (1974) may be accommodated within the transfer-appropriate-processing theory of memory (Morris, Bransford, & Franks, 1977). Even if both WM and LTM tasks involve retrieval from the same secondary memory system, the demands of WM tasks may bias the use of processes that would be less appropriate for LTM tasks. Whereas rehearsal and the use of more transient cues (e.g., acoustic, temporal) tend to be sufficient for WM tests (Mazuryk & Lockhart, 1974), LTM tests typically involve sets of to-be-remembered material that are too large, and delays that are too long, for the same type of retrieval processes to be effective. Rather, more durable semantic cues tend to produce optimal LTM retrieval (Craik & Lockhart, 1972; Craik & Tulving, 1975). Put simply, WM and LTM tasks may involve retrieval from the same secondary memory system, yet their retrieval processes may differ. As a result, differences in the effects of many variables (e.g., LOP) may be expected.

Because recent theories assume that WM involves retrieval from secondary (long-term) memory, we set out to examine similarities and differences between retrieval in a new WM task and in a standard LTM task (delayed recognition) as a function of an LOP manipulation. Two experiments addressed this issue. Experiment 1 assessed the effect of LOP on WM and LTM (subsequent recognition for the same items). The second experiment was designed to replicate and extend the results obtained in Experiment 1. We addressed the hypothesis that retrieval from secondary memory is involved in performance of WM tasks by examining how the retrieval practice provided by the initial WM task affected long-term retention (relative to a condition without immediate testing). We also examined how retrieval practice effects differed as a function of list length.

## Experiment 1

### Method

**Participants.** Twenty-four Washington University undergraduate students participated in exchange for course credit. All participants were native English speakers, except for one who reported speaking English since the age of 4. Participants were screened for normal or corrected-to-normal visual acuity, as well as for color vision deficiencies. Their mean age was 18.9 ($SD$ = 0.9), and their mean score on the vocabulary subtest of the Wechsler Adult Intelligence Scale—Third Edition was 56.6 ($SD$ = 6.8; Wechsler, 1997).

**LOP span task.** We developed a new complex span procedure, the LOP span task, to assess whether depth of processing affects WM and LTM in similar or different ways. In this task, participants were presented with lists of to-be-remembered target words, with each target word followed by two processing words presented side by side. Depending on the condition, the participant was to determine which of the processing words was the same color as the target word, which one rhymed with the target word, or which one was semantically related to the target word. We hypothesized that the secondary task (picking a match based on color, rhyme, or meaning) would function like an orienting task in the standard LOP experimental paradigm. Following presentation of the list of target words and their associated processing words, participants attempted to recall the target words in serial order.

As depicted in Figure 1, a to-be-remembered target word (e.g., *BRIDE*, presented in red), would be followed by two processing words (e.g., *dried*, in blue; and *groom*, in red). Depending on the processing condition, the participant would indicate which processing word matched in color (*groom*), rhyme (*dried*), or semantic relatedness (*groom*) by pressing the left or right labeled key. The color-matching processing word was counterbalanced to appear as the semantic and phonological associate equally often, and blue and red colored words alternated between the left and right positions randomly. Following this response, another target word would be presented (e.g., *LEG*), followed by two more processing words (e.g., *arm* and *beg*). After several target words and pairs of processing words were presented, participants were prompted to recall the target words in order (e.g., "*bride, leg*"). At issue was whether "deeper" (i.e., semantic) processing would provide a benefit to WM relative to "shallower" (i.e., visual, phonological) processing.

For each condition, 54 monosyllabic target words were selected from the English Lexicon database (Balota et al., 2007). The mean lengths of the sets of target words for the color, rhyme, and semantic conditions were 4.3, 4.0, and 4.1 letters, and their mean log-HAL frequencies were 10.0, 9.9, and 10.0, respectively; neither difference was significant, $Fs(2, 161) = 2.43$ and 0.83, $ps > .05$. Each target word was paired with a rhyming word obtained from the Washington University (2009) Speech & Hearing Lab Neighborhood Database and with a semantically associated word obtained from the University of South Florida Free Association Norms Database (Nelson, McEvoy, & Schreiber, 1998). The mean forward-associative strengths for the visual, phonological, and semantic processing sets of target-associate pairs were .49, .45, and .46, respectively; this difference was not significant, $F$(2,

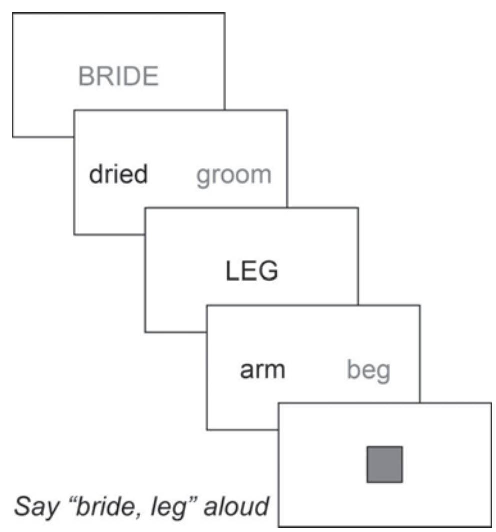


*Figure 1.* Example of the levels-of-processing span task for a two item list. Words in uppercase are the to-be-remembered target words. Depending on the condition, the participant was to determine which of the two intervening processing words was the same color as, rhymed with, or was semantically related to the target word. At the end of the trial, the participant was to recall the target words aloud in the order presented.

161) = 0.65, $p > .05$. The sets of words were also matched on imageability according to the mean rating from the combined norms from the MRC database (Coltheart, 1981) and the Bristol norms (Stadthagen-Gonzalez & Davis, 2006). The mean imageability rating for each set was 549, 534, 546, respectively; this difference was not significant ($F < 1$).

**Procedure.** Participants were tested individually. Stimuli were presented on a computer monitor. On each trial a fixation cross appeared where each target word would be presented. The participant began each trial by pressing the space bar when ready, after which a to-be-remembered target word was displayed in either blue or red for 1,750 ms. The participant was to say the word aloud and remember the word for recall at the end of the trial. After a 250-ms blank screen, the target word was replaced by two processing words (i.e., the semantic associate and rhyme, one of which was presented in blue and the other in red).

In all conditions, the participant was to select the appropriate processing word by pressing either the left or the right key to indicate whether the matching word was on the left or right. Prior to testing, the participant was instructed to make each decision as quickly as possible without sacrificing accuracy. After the processing decision was made, the screen was blank for 750 ms before the next target word appeared. At the end of the trial, a green box and a tone cued the participant to recall the target words aloud in the order presented. Participants were told that if they were unable to recall all of the target words, they were to recall as many as possible in the order presented. Before starting the test trials, participants performed four practice trials of two, three, four, and five sets of target and processing words in order to familiarize them with the procedure. Recall responses were recorded by electronic voice recorders for later scoring.

The color, rhyme, and semantic processing trials were blocked by condition, and the order of blocks was counterbalanced across participants. Within each block, there was one trial each of two, three, four, five, six, and seven target words. List length was varied in a pseudorandom order in order to prevent participants from predicting how many words were to be remembered on each trial (i.e., list length did not increase or decrease in a predictable pattern). Between blocks, participants performed nonverbal reaction time tasks involving shape and distance judgments. These tasks were intended to allow a rest from processing verbal stimuli and minimize proactive interference across conditions.

Following the third and final block, there was a filled delay during which participants performed a mental arithmetic task. This task consisted of 15 problems that each involved solving for a term in an equation using either addition or subtraction (e.g., $x + 76.31 = 164.89$; $x =$ ___?). Upon completion of all of the mental arithmetic problems, which took participants approximately 4 min ($M = 3.9$, $SD = 1.4$), they were given a surprise recognition memory test on the target items from the LOP span task.

In the recognition test, participants saw 162 individually presented words: the 81 target words from the three conditions of the LOP span task plus 81 lure words that had not been previously presented. Lures were matched to the target words based on length and word frequency. None of the processing words from the LOP span task were included in the recognition task, and participants were informed of this. For each word, participants were instructed to indicate whether it had been one of the target words from any of the previous conditions. Participants reported old–new decisions by pressing the left mouse button to indicate old and the right button to indicate new (i.e., not presented in any previous part of the experiment). Following each recognition decision, participants provided a confidence rating. They were instructed that pressing the *1*, *2*, *3*, or *4* key on the keyboard indicated *definitely old*, *probably old*, *probably new*, or *definitely new*, respectively.

**Scoring.** Performance on the LOP span task was scored in two different ways: memory span (i.e., the maximum number of target words that could be recalled in correct serial order) and the overall proportion of target words recalled, irrespective of their serial position. The latter measure is more typical of traditional LTM experiments. For all experiments the $p$ value was set to .05.

## Results

We first verified that participants performed the processing operations required by the LOP span task. The proportion of correct processing decisions was high in all conditions: visual = .94, ($SD = .07$), phonological = .98 ($SD = .03$), and semantic = .96 ($SD = .05$), $F(2, 23) = 2.21$, *ns*. Perhaps surprisingly, the LOP manipulation did not significantly influence WM performance (see the immediate test data in the left panel of Figure 2). The results of an analysis of variance on the memory span measure failed to show an effect of processing condition, $F(2, 23) = 1.3$, *ns*, and similar results were obtained for the overall proportion of words recalled ($F < 1$).[1] When processing times for each condition were used as a covariate, the effect of processing was again not significant, $F(2, 19) = 1.84$, *ns.*

On the other hand, as can be seen in the delayed test data (see the right panel of Figure 2), recognition for the *same* items revealed a different pattern when it was assessed after a brief delay. Deeper LOP benefited delayed recognition of the same words that were previously processed in the LOP span task, $F(2, 23) = 13.48$, $p < .001$. Semantic processing produced a significant advantage over both phonological and visual processing, $t$s(23) = 3.65 and 4.60, respectively ($p$s < .01). Phonological processing produced an advantage over visual processing, although this difference did not reach significance, $t(23) = 2.05$, $p = .05$. Analysis of the confidence judgments for correctly recognized items showed that the semantic processing condition was associated with greater reported confidence than both the visual and phonological conditions, both $t$s(23) > 3.3, $p$s < .01, whereas the visual and phonological conditions did not differ from one another, $t(23) < 1$.

## Discussion

The results of Experiment 1 revealed an interesting dissociation between WM and LTM that conceptually replicates the findings of Mazuryk and Lockhart (1974). If depth of processing had an effect on WM similar to that typically observed with delayed episodic

[1] The LOP span task results were replicated in an independent sample of 24 Washington University undergraduate students, with two trials at each list length in order to estimate performance in terms of a more traditional memory span measure (e.g., Hale et al., 1996). These spans were 4.4 ($SD$ = 0.8) for the visual condition, 4.0 ($SD = 0.8$) for the phonological condition, and 4.5 ($SD = 0.9$) for the semantic condition. Span did not differ across conditions, $F(2, 23) = 2.3$, $p = .10$, thus replicating the absence of an LOP effect found in Experiment 1.

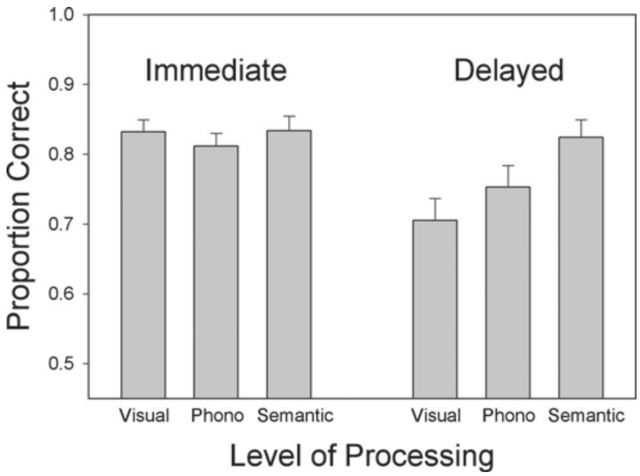


*Figure 2.* Experiment 1: Immediate data are the proportion of target words recalled on the levels-of-processing span task for the visual (color), phonological (rhyme), and semantic conditions. Delayed data are the proportion of target words from the levels-of-processing span task that were called "old" (i.e., hits) on the delayed recognition test. The mean false alarm rate was .30 ($SD$ = .16). Error bars are the standard error of the mean.

memory tests, then one might have expected semantic processing to have resulted in better memory performance than phonological and visual processing. However, the present results did not demonstrate this pattern. The semantic condition of the LOP span task did not result in significantly higher WM scores than the conditions that focused processing on more shallow, perceptual features. However, when LTM for those same words was assessed, the classic LOP effect was obtained.

In our experiment, WM was assessed with immediate recall, whereas LTM was assessed with delayed recognition, which may raise the concern that the difference in LOP effects could be due to differences between recall and recognition procedures. Although delayed tests of recall and recognition have both been traditionally used as measures of episodic memory, they are known to differ in many ways (e.g., Haist, Shimamura, & Squire, 1992; Tulving, 1976). However, using an experimental design that was very similar to that of the present experiment, Mazuryk and Lockhart (1974) found that delayed recall and recognition of items from prior immediate recall tests showed the same pattern of results. Therefore, it is unlikely that the reason our immediate test failed to show an LOP effect but the delayed test did was due to the difference between recall and recognition procedures.

The findings of Experiment 1 suggest that the LOP effect typically observed on explicit tests of LTM does not occur with tests of WM. We consider the difference between LOP effects an interesting dissociation between WM and LTM. One implication of these findings is that WM appears to obey different principles, at least with regard to the effect of LOP. One possible reason is that WM processes may not simply represent a subset of those involved in LTM. An alternate interpretation is that WM and LTM tests rely on different types of retrieval processes. This latter interpretation need not suppose that different "systems" were involved in the different tests, but rather that the demands of the two tests bias the use of different processes.

Before discussing the implications of our research, we sought to provide additional tests of the hypothesis that retrieval from secondary memory is involved in the performance of WM span tasks (Unsworth & Engle, 2007a). To this end, we conducted a second experiment, which also served to address some methodological concerns that might cloud interpretation of the findings of Experiment 1. These issues will be discussed in turn.

## Experiment 2

If WM span tasks do involve retrieval from secondary memory (Unsworth & Engle, 2007a), why did an LOP manipulation fail to show the classic effect when memory was tested immediately? After all, the LOP effect is ubiquitous in explicit memory tests of long-term retention. The failure to observe an effect of LOP in a WM task may seem especially puzzling given that such an effect was observed when recognition for the same items was assessed after a delay, when secondary memory was certainly involved. Does this pattern indicate that secondary memory was not involved in performing the LOP span task and that the task relied entirely on primary memory? Experiment 2 addressed this question in two different ways.

The first approach was based on the finding that recalling items from secondary memory (i.e., retrieval practice) can have important consequences for the long-term retention of those items (Craik, 1970; Karpicke & Roediger, 2008; Roediger & Karpicke, 2006). Practice retrieving items from secondary memory results in substantial benefits to long-term retention on later memory tests, even when compared to control conditions in which the items are restudied rather than tested (Karpicke & Roediger, 2008). Thus, if it is the case that performing a span task does involve retrieving items from secondary memory, then recalling items for the LOP span task should benefit participants' long-term retention compared to a condition in which immediate recall of the words was not required. Importantly, repeated retrieval from primary memory often has little or no effect on a long-term test (e.g., Karpicke & Roediger, 2007; see below).

To assess whether the LOP span task does provide practice retrieving items from secondary memory, we had half of the participants in Experiment 2 perform the LOP span task as in Experiment 1, whereas we had the other half make the same processing decisions on the same words but we did not have them engage in immediate recall. At issue was whether the group that performed the LOP span task with immediate testing would show less forgetting of the items on a surprise, delayed test than the group without immediate testing. If the immediate testing group were to show less forgetting, this would suggest that performance of the LOP span task provides retrieval practice from secondary memory.

The second way in which Experiment 2 addressed the issue of whether the LOP span task involves retrieval from secondary memory was based on comparing the long-term retention of items from supraspan lists that exceed WM capacity with retention of items from shorter lists that are within capacity limitations. Retrieval from secondary memory should play a larger role in the LOP span task when participants try to maintain and recall items from supraspan lists, compared to when the items are from shorter lists because items from shorter lists are more likely to be reported directly from primary memory (Unsworth & Engle, 2006). As previously noted, practice retrieving items from secondary memory results in substantial benefits to long-term retention (e.g., Roediger & Karpicke, 2006). In contrast, reporting items directly from primary memory does not provide practice that benefits delayed tests, as demonstrated by the negative recency effect and the relative ineffectiveness of rote rehearsal as a mnemonic technique (Craik, 1970; Craik, Gardiner, & Watkins, 1970; Craik & Watkins, 1973; Jacoby & Bartz, 1972; Madigan & McCabe, 1971; Mazuryk & Lockhart, 1974; L. McCabe & Madigan, 1971; Rundus, Loftus, & Atkinson, 1970; Smith, Barresi, & Gross, 1971).

In the present experiment, it was expected that initial recall would be better for shorter lists. However, because of the increasing involvement of secondary memory as list length increases, we hypothesized that items recalled from longer lists should benefit more from retrieval practice and would subsequently show better long-term retention than would items from shorter lists. In Experiment 1, the mean memory span on the LOP span task was approximately 4.3 items. Therefore, participants in Experiment 2 were presented with four-item lists and eight-item lists. On average, we reasoned that four-item lists should be within WM capacity whereas eight-item lists should be well above span. Thus, maintaining and recalling items from eight-item lists would be more likely to involve retrieval from secondary memory. If this were the case, then this retrieval practice would render items from the longer lists less likely to be forgotten than would those from the shorter lists.

Experiment 2 also addressed a possible methodological concern. In Experiment 1, differences existed in the amount of time it took to process words in the various conditions: Phonological (rhyme-matching) and visual (color-matching) processing decisions both took significantly less time than did semantic processing decisions, and the difference was especially great with visual processing. Table 1 presents the mean reaction times for the different processing conditions for Experiment 1 (as well as for Experiment 2, discussed below). Note that the amount of time that the to-be-remembered target words were displayed was the same in all conditions; nonetheless, the differences in the amount of time spent on the processing operations could have affected the results. Although extensive research on the LOP effect suggests that the type of processing is more important than the amount of processing time for later retention (Craik, 2002; Craik & Tulving, 1975), one might question whether this finding generalizes to the LOP span task.

Therefore, in addition to the semantic and phonological processing tasks used previously, Experiment 2 included a more time-consuming, shallow processing task: vowel counting. Participants had to count the vowels in the target and processing words and then decide which processing word provided the closest match to the target word in this regard. Pilot data indicated that vowel counting would be more time consuming than semantic processing and thus permit a rigorous test of the hypothesis that the better LTM for semantically processed items from the LOP span task truly reflected the type and not simply the amount of processing that these words received.

## Method

**Participants.** Forty-eight Washington University undergraduate students participated in exchange for course credit. Half were assigned to the condition with immediate testing (i.e., LOP span task), whereas the other half were assigned to the condition without immediate testing (i.e., processing decisions only). All participants were native English speakers, except for two who reported speaking English since the ages of 1 and 3. Participants were screened for normal or corrected-to-normal vision, as well as for color vision deficiencies. The mean ages for the groups with and without immediate tests were 19.8 ($SD = 1.4$) and 19.8 ($SD = 1.0$), and their mean scores on the Mill Hill Vocabulary test (Raven, 1958) were 14.6 ($SD = 1.6$) and 14.8 ($SD = 1.8$), respectively.

**Stimuli.** Three sets of 36 new monosyllabic target words were selected from the English-Lexicon Project Database (Balota et al., 2007). The mean word length for each set was 4.2, 4.2, and 4.4 letters and the mean log-HAL frequency for each set was 9.61, 9.65, 9.73; neither difference was significant, $Fs(2, 105) = 1.20$ and 0.12, respectively ($ps > .05$). Each of the target words was paired with both a semantically associated word obtained from the University of South Florida Free Association Norms Database (Nelson et al., 1998) and a rhyming word, obtained from either www.rhymer.com or the Washington University (2009) Speech & Hearing Lab Neighborhood Database. The mean forward associative strength between each target word and its semantic associate (i.e., the semantic processing word) for each set was .366, .369, .378; this difference was not significant ($F < 1$). The sets of words were also matched on imageability according to the combined norms from the MRC database (Coltheart, 1981) and the Bristol norms (Stadthagen-Gonzalez & Davis, 2006). The mean imageability rating for each set was 548, 535, 553; this difference was not significant ($F < 1$).

Table 1
*Mean Processing Decision Times (in Milliseconds) on the Levels-of-Processing Span Task for Experiments 1 and 2*

| | Visual (color or vowel) | | Phonological | | Semantic | |
|---|---|---|---|---|---|---|
| Experiment and group | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Experiment 1 | 630 | 175 | 1,022 | 199 | 1,243 | 185 |
| Experiment 2 | | | | | | |
| With immediate tests | 1,852 | 352 | 1,034 | 159 | 1,140 | 178 |
| Without immediate tests | 1,300 | 372 | 730 | 147 | 823 | 143 |
| *M* | 1,576 | 454 | 882 | 216 | 982 | 226 |

*Note*. Visual processing was color matching for Experiment 1 and vowel matching for Experiment 2. Values in the row labeled *M* are the average processing decision times for the three conditions of Experiment 2 collapsed across the groups with and without immediate tests.

We selected the phonological processing words so that target–rhyme pairs would be as orthographically dissimilar as possible (e.g., *height–site*) in order to increase the duration of rhyme-matching decisions (Seidenberg & Tanenhaus, 1979) and make them more similar to semantic matching decisions in this regard. To provide a processing word for the visual (vowel) matching condition, we selected the phonological and semantic associates for each target so that one was equal or closer to the target word in number of vowels; the matching word was the phonological associate in half of the instances and the semantic associate in the other half. Participants were instructed not to consider *y* as a vowel in the experiment. The target words and their processing words were counterbalanced across the three conditions such that each target word was a to-be-remembered word for each of the three conditions of the LOP span task. As in the previous experiment, the order of processing conditions was counterbalanced across participants. The word triads used in Experiment 2 may be found in the Appendix.

**Procedure.** The procedure was identical to that of Experiment 1 except that participants performed three trials of four- and eight-item lists of target words rather than one trial each of two-, three-, four-, five-, six-, and seven- item lists. Additionally, whether or not recall was required after performing the processing decisions for the lists of four or eight items was varied between participants. That is, half of the participants performed the LOP span task and delayed recognition test just as in Experiment 1, whereas the other half only performed the processing decisions and then the delayed recognition test. The group of participants that did not engage in immediate serial recall tests made the same processing decisions as in the LOP span task, but under the guise of a reaction time experiment. Participants were instructed to make each processing decision as fast and as accurately as possible. Following a set of four or eight of these decisions, a green box appeared. For the group performing the LOP span task with immediate tests, the green box served as the cue to begin recalling the target words. For the processing-only condition, participants were instructed to pause momentarily and continue on to the next trial when ready.

Again, the processing conditions (vowel, rhyme, and semantic) were blocked. Within each block, we varied pseudorandomly list length to try to prevent participants performing the LOP span task from predicting the number of target words to be remembered on a given trial. The order of list lengths was four, eight, four, eight, eight, four. Following the third and final block, all of the participants performed mental arithmetic for a fixed amount of time (10 min), rather than a fixed number of problems as in the previous experiment. Following the 10-min filled retention interval, all participants were given a surprise recognition test.

For the recognition test, the 102 target words that were presented in the LOP span task and 102 new lure words that had never appeared in the experiment were presented individually. Lures were matched to the target words based on length and word frequency. For each word, participants were instructed to indicate whether that word was one of the to-be-remembered target words from any of the three processing conditions. Old–new decisions were made by pressing the *1*, *2*, *3*, or *4* key on the keyboard to indicate *definitely old*, *probably old*, *probably new*, or *definitely new*, respectively. Accuracy of old and new responses was scored by combining 1 and 2 responses, and 3 and 4 responses, respectively.

## Results and Discussion

This section is organized into three parts. The first part concerns the effects of the processing tasks on performance of the LOP span task and on delayed memory for the same items; these results bear on whether the findings of Experiment 1 were influenced by the fact that the semantic processing decisions took longer than the shallower processing decisions. The second and third parts focus on the issue of whether the LOP span task involves retrieval from secondary memory. We contrast LTM for the group with immediate testing (i.e., the LOP span task) with the group that engaged in the same initial processing of items but without immediate testing. Finally we consider how remembering different list lengths for the LOP span task affected subsequent long-term retention of memory items. At issue in both cases is whether the key manipulations (immediate testing and supraspan list lengths) provided the benefits to long-term retention that would be expected if the LOP span task provided practice retrieving items from secondary memory.

**Effects of processing on immediate and delayed memory for LOP span task items.** The first set of analyses compared immediate and delayed memory performance, and therefore, focused on the group with immediate testing (i.e., the participants who performed the LOP span task). With regard to processing times, the vowel-counting decisions were the slowest, as expected. These data are presented in Table 1 along with the mean reaction times for the processing decisions for Experiment 1 for comparison. Importantly, the vowel-counting processing decisions took significantly longer than did the semantic processing decisions, $t(23) = 9.7$, $p < .001$. The vowel-counting decisions were less accurate than were the phonological and semantic decisions: vowel = .87 ($SD$ = .09), phonological = .98 ($SD$ = .02), semantic = ($SD$ = .03), both $ts(23) > 5.8$, $ps < .001$. Processing decision accuracies for the phonological and semantic conditions did not differ ($t < 1.0$).

Despite the inclusion of a shallow processing task (i.e., vowel-counting) that involved over 700 ms more processing time per item than semantic processing, on average, the pattern of results was similar to those in the previous experiment: Semantic processing did not benefit immediate recall but did benefit delayed recognition of the same items.

The immediate recall data are shown in Table 2. As in Experiment 1, there was no effect of processing condition, $F(2, 23) = 1.89$, $p = .17$. As expected, the proportion of words recalled from four-item lists was greater than from eight-item lists, $F(1, 23) = 410.57$, $p < .001$, but this effect did not interact with LOP ($F < 1$). The LOP effect also did not interact with the proportion of words recalled as a function of serial position for either the four- or eight-item lists (both $Fs < 1$).

In contrast, delayed recognition for the same items did show an LOP effect, as can be seen in the left panel of Figure 3. There was a main effect of processing condition, reflecting the fact that semantic processing benefited delayed recognition, $F(2, 23) = 3.42$, $p = .05$. There was no effect of list length ($F = 1.51$, $p = .23$), and no interaction ($F < 1$). The findings with respect to LOP replicate the results of Experiment 1 and extend them by showing

Table 2
*Immediate Serial Recall Results on the Levels-of-Processing Span Task for Experiment 2*

| Processing by list length | Proportion of items recalled | |
|---|---|---|
| | *M* | *SD* |
| Visual | | |
| 4 items | .85 | .15 |
| 8 items | .47 | .10 |
| Phonological | | |
| 4 items | .88 | .13 |
| 8 items | .47 | .12 |
| Semantic | | |
| 4 items | .89 | .15 |
| 8 items | .50 | .12 |

that semantic processing benefited delayed recognition despite involving less processing time than vowel counting.

**Effects of maintaining items for immediate memory testing.** The second set of analyses compared the delayed recognition of participants who performed the LOP span task to those who saw the same words and made the same processing decisions but who did not have to recall the words. We first verified that the group without immediate testing also successfully performed the processing decisions and that vowel processing took significantly longer than semantic processing as well. The proportion of correct processing decisions was high in all conditions: visual = .94, (*SD* = .04), phonological = .95 (*SD* = .09), semantic = .96 (*SD* = .04), $F(2, 23) = 1.36$, *ns*. The mean reaction time was significantly longer for vowel-counting processing decisions than for the semantic processing decisions, $t(23) = 7.2$, $p < .001$.

As predicted, immediate testing resulted in better delayed recognition performance (compare the left and right panels of Figure 3). Participants who had to recall items during the LOP span task subsequently showed better delayed recognition of these same items than did participants who had only made processing decisions, $F(1, 46) = 8.63$, $p < .01$. That is, trying to maintain and then immediately recall words on the LOP span task benefited delayed recognition, consistent with the hypothesis that this task provided retrieval practice. In addition, both groups demonstrated an LOP effect, $F(2, 92) = 22.94$, $p < .001$, reflecting the fact that semantic processing benefited delayed recognition more than did shallower processing. The interaction between processing condition and immediate testing failed to reach significance, $F(2, 92) = 2.4$, $p = .10$. There was, however, an interaction between list length and immediate testing that approached significance, $F(1, 92) = 3.07$, $p < .09$, reflecting the fact that the difference between the groups with and without immediate testing tended to be bigger for items from eight-item lists than from four-item lists. This finding suggests that retrieval practice was more beneficial to the long-term retention of words from longer lists. Further evidence consistent with this possibility is presented in the next section.

Analysis of the confidence data for both groups showed that hits (i.e., *definitely old* and *probably old* responses) in the semantic processing condition were associated with greater reported confidence than were hits in either the visual or phonological conditions, $ts(47) > 5.6$, $ps < .001$, whereas the visual and phonological conditions did not differ from one another in this regard, $t(47) < 1$.

**Differences in long-term retention of four- and eight-item lists.** The third set of analyses compared forgetting of items from longer, supraspan lists with forgetting of items from shorter lists. Because these analyses focused on the forgetting of previously recalled items, they examined only data from the group that performed the LOP span task. One piece of evidence suggesting that retrieval from secondary memory benefited delayed recognition is provided by the List Length × Immediate Testing interaction just discussed. Converging evidence comes from comparing the conditional probabilities of failing to recognize previously recalled items from shorter and longer lists. As noted previously, trying to remember a list of items that exceeds the limits of WM capacity is likely to involve retrieval from secondary memory to a greater extent than trying to remember a shorter list, and practice retrieving items from secondary memory benefits long-term retention. Therefore, items from supraspan lists, such as the eight-item lists in this experiment, should be less likely to be forgotten than items from lists that are closer to capacity limitations, such as the four-item lists. Consistent with our hypothesis, items that were initially recalled from four-item lists were less likely to be recognized later than items initially recalled from eight-item lists. The mean hits were .66, for items recalled from four-item lists, and .77, for items recalled from eight-item lists, $t(23) = 2.7$, $p < .05$.

Figure 4 illustrates the difference in long-term retention for items initially from four- and eight-item lists for the group that performed the LOP span task. Plotted separately for words from four- and eight-item lists is the proportion of words recalled initially, on the immediate tests, and the proportion of words recognized as old on the delayed recognition test (which includes both words that were not initially recalled as well as those that were). As may be seen, for four-item lists, participants recognized fewer items after the delay than they initially recalled, whereas the

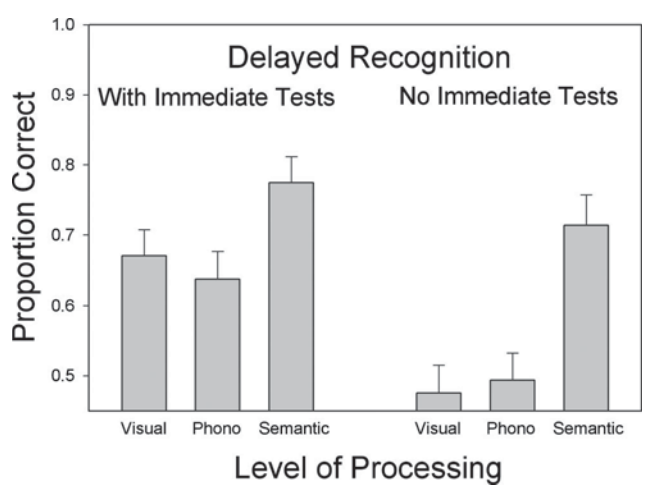


*Figure 3.* Experiment 2: Delayed recognition of target words (i.e., hits) from the visual (vowel), phonological (rhyme), and semantic conditions of the levels-of-processing span task for groups with or without immediate recall tests. The mean false alarm rates for the groups with and without immediate tests were .25 (*SD* = .11) and .279 (*SD* = .13), respectively. Error bars are the standard error of the mean.

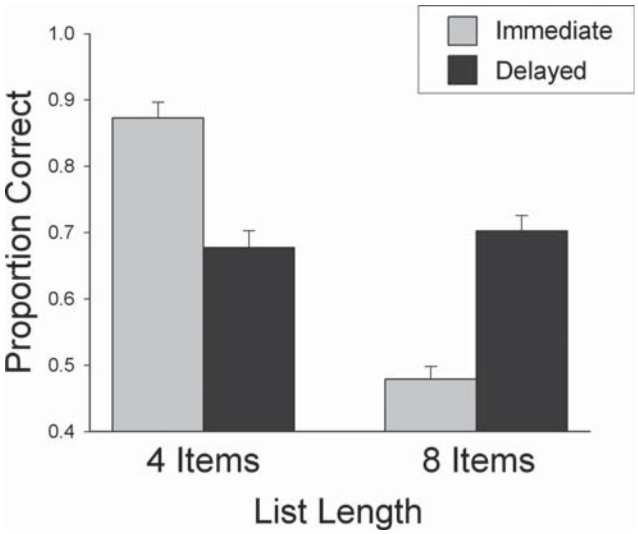


*Figure 4.* Experiment 2: Immediate data are the proportion of target words recalled from four- and eight-item lists on the levels-of-processing span task, collapsed across condition. Delayed data are the proportion of target words from four- and eight-item lists on the levels-of-processing span task that were called "old" (i.e., hits) on the delayed recognition test, collapsed across condition. The mean false alarm rate was .25 ($SD = .11$). Error bars are the standard error of the mean.

reverse was true for the eight-item lists: Participants recognized more items than they initially recalled. Although the level of performance is not directly comparable across immediate recall and delayed recognition tests, the observed pattern reflects the greater forgetting of items from shorter lists, which may have occurred because items from such short lists received less secondary memory retrieval practice than did items from supraspan lists.

This result, taken together with the finding that trying to maintain and immediately recall items resulted in better delayed recognition than merely processing the items, provides converging evidence that the LOP span task involves retrieval from secondary memory. This evidence makes the failure to observe an LOP effect in immediate testing seem all the more puzzling, especially given that such an effect emerged in subsequent delayed testing. Although the present results are consistent with the hypothesis that both WM and LTM tests involve retrieval from secondary memory, they also represent a clear task dissociation between WM and LTM.

## General Discussion

Two experiments examined whether LOP effects occurred in WM tasks. If WM tasks involve retrieving items from secondary memory, as suggested recently (Mogle et al., 2008; Unsworth, 2009; Unsworth & Engle, 2006, 2007a, 2007b), then it would seem that the type of processing that items receive when they are initially encoded should have similar effects on WM and LTM performance. However, in neither of the present experiments did deeper (semantic) processing benefit WM performance. Nevertheless, when recognition memory for the same items was assessed after a 5- to 10-min delay, the typical LOP effect was consistently observed. Although this finding points to differences between WM and LTM, other findings point to their similarities. For example, performing the immediate recall tests for the LOP span task resulted in better delayed recognition, relative to performing the same processing operations but without immediate testing (i.e., a testing effect). Finally, words recalled from supraspan lists on immediate tests were less likely to be forgotten later than were words recalled from shorter lists, a result similar to negative recency effects reported in delayed tests (e.g., Craik, 1970).

### Retrieval From Secondary Memory in WM Tasks

**Benefits of retrieval practice.** One piece of evidence in support of the hypothesis that the LOP span task involves retrieval from secondary memory comes from comparing the effects of trying to maintain and recall items for immediate tests with the effects of processing the items in the same way at encoding but without the immediate test requirement. In Experiment 2, the condition with immediate testing (i.e., the LOP span task), which provided retrieval practice, facilitated LTM for those items on a surprise delayed recognition test, relative to the condition without testing.

This result is similar to the testing effects reported by others (e.g., Carrier & Pashler, 1992; Roediger & Karpicke, 2006). Tests provide retrieval practice, and repeated retrieval practice has been shown to result in a substantial benefit to long-term retention, even when compared to repeated study (Karpicke & Roediger, 2008; Roediger & Karpicke, 2006). We believe that the benefits to long-term retention associated with the LOP span task are also the result of retrieval practice, and are consistent with the hypothesis that this task, like other complex span tasks, involves retrieval from secondary memory (D. P. McCabe, 2008; Unsworth & Engle, 2007a).

**Greater benefits from retrieving items from secondary memory.** Another piece of evidence in support of the hypothesis that the LOP span task involves retrieval from secondary memory comes from comparing the long-term retention of items from supraspan lists with retention of items from shorter lists. In Experiment 2, words recalled from eight-item lists were less likely to be forgotten following a filled delay than were words from four-item lists. If immediate recall provides retrieval practice for both supraspan lists and shorter lists, why should their long-term retention differ?

Previous research has shown that retrieval practice does not uniformly benefit delayed retention. Rather, the amount of benefit observed depends upon the extent to which items are retrieved from secondary memory. For example, D. P. McCabe (2008) has shown that retrieving items for simple and complex WM span tasks has different effects on LTM for those items. He had participants perform both a simple span task (i.e., word span) and a complex span task (i.e., operation span) with immediate recall of two, three, or four words per list, followed by a final free recall test for all of the words from those span tasks. On the delayed test, participants recalled more items from operation span lists even though they were less likely to recall these items on immediate memory span tests. This double dissociation is consistent with the hypothesis that complex span tasks involve retrieval from secondary memory to a greater extent than do simple span tasks (Unsworth & Engle, 2007b).

As mentioned, the differential involvement of secondary memory is presumably because items on simple span tasks are not displaced from primary memory by other secondary task operations. Therefore, items for simple span tasks may be maintained within and reported directly from primary memory. As D. P. McCabe (2008) has shown, the differential involvement of retrieval from secondary memory benefits long-term retention. Importantly, for the present purposes, retrieval from secondary memory is involved to a greater extent for longer list lengths as well (Unsworth & Engle, 2006, 2007a).

During performance of a WM span task, secondary memory should be involved to a greater extent when one is trying to maintain and recall lists of items that exceed WM capacity (i.e., supraspan lists). Only a few items from such lists are likely to be reported directly from primary memory, while the remaining items must be retrieved from secondary memory (Unsworth & Engle, 2006, 2007a) or the activated portion of LTM (Cowan, 1999). In contrast, lists of four or fewer items are more likely to be maintained in and reported directly from primary memory. The consequences of this differential involvement of secondary memory are evident in the comparison of long-term retention of items from the eight-item, supraspan lists versus the four-item lists (see Figure 4). In agreement with D. P. McCabe's (2008) results, long-term retention of supraspan list items was greater than for shorter list items, presumably due to the greater reliance on secondary memory in the former case. The remaining question is this: If complex WM span tasks (such as the LOP span task) involve retrieval from secondary memory, especially for supraspan list lengths, why did depth of processing have negligible effects on immediate recall?

## On the Relation Between WM and LTM

Unsworth and Engle (2007a), among others (e.g., D. P. McCabe, 2008; Mogle et al., 2008), argued that complex span tasks involve retrieval from secondary memory as well as primary memory. As just discussed, several aspects of the current results support this claim. However, the fact that depth of processing had no effect on immediate recall points to a clear difference between WM and LTM. We attempt to reconcile this discrepancy in this section.

One possible reconciliation is to question whether the secondary memory system as conceptualized in the context of WM tasks (D. P. McCabe, 2008; Unsworth & Engle, 2007a) is the same system as that which has been traditionally conceptualized in the context of LTM tasks. The present study clearly demonstrates that secondary memory, as assessed by short-term and long-term procedures, has different properties (i.e., one does not show an LOP effect and the other does). This raises the possibility that retrieval from the activated portion of LTM has functionally different properties than retrieval of deactivated items.

Another possible solution to the puzzle raised by our results is to argue that different types of retrieval processes are involved in WM and LTM tasks. Performance on immediate (e.g., WM) and delayed (e.g., LTM) tests has long been thought to depend on different types of processes (more specifically, the use of different retrieval cues; e.g., Baddeley, 1966a, 1966b; Kintsch & Bushcke, 1969; Shulman, 1970, 1971). For WM tests, items typically undergo relatively superficial, nonsemantic encoding, which may be adequate for immediate recall; for subsequent (long-term) retrieval of the items, however, semantic encoding is more effective (Craik & Levy, 1970; Craik & Watkins, 1973; Jacoby & Bartz, 1972; Mazuryk & Lockhart, 1974; Tulving, 1968).

In the present context, even if semantic cues were encoded on the LOP span task (at least in the semantic processing condition), and even if both WM and LTM tests involved retrieval from the same secondary memory system, the demands of the two types of tests may have biased the use of different retrieval processes. That is, because WM tasks require immediate recall, participants try to maintain to-be-remembered items via covert retrieval (D. P. McCabe, 2008). Thus, it is likely that participants used covert retrieval in all three processing conditions of the LOP span task. Moreover, participants in all three conditions likely focused on and utilized more transient cues (e.g., acoustic, temporal) for the immediate recall tests. The latter possibility is especially likely because all three conditions of the LOP span task required that participants read items aloud and recall them in serial order. Consequently, all conditions involved acoustic and temporal encoding, which provides potent cues for immediate recall (e.g., Craik, 1969; Unsworth, 2009), and could therefore have resulted in the items being recalled equally well on the immediate tests. With time, however, these cues would be susceptible to decay or interference, revealing the beneficial effects of semantic processing on memory (for a similar argument, see Bartlett & Tulving, 1974).

The key point is that all conditions of the LOP span task combine processing that is appropriate for immediate recall with processing that is either more or less appropriate for delayed retrieval, depending on the condition (i.e., the orienting task). According to this hypothesis, the effect of depth of processing was initially masked by the effects of active maintenance and acoustic and/or temporal encoding processes on immediate recall and was only revealed subsequently by delayed retrieval.

The preceding account of the present findings is consistent with transfer-appropriate processing theory. According to this theory, the long-term retention of items is determined not only by the depth of processing at encoding, but also by how well the requirements of a subsequent memory test match the processes originally used to encode information (Morris et al., 1977). The processing that produces the best WM performance (e.g., active maintenance, acoustic, and/or temporal encoding) differs from that which produces the best LTM performance, and thus may result in a transfer of inappropriate processing, in the sense that it produces encoding that is not well suited for later retrieval. We believe that examining the role played by transfer-appropriate processing in the LOP span task provides a fruitful approach for addressing the distinction between WM and LTM. For example, future research should address whether LOP effects would appear on the LOP span task if active maintenance processes were eliminated.

From the perspective of transfer-appropriate-processing theory, the extent to which similarities and differences are observed between performance on WM and LTM tasks should depend on the extent to which the demands of the tasks converge or diverge. Similarities and differences between WM and LTM that were produced by LOP and retrieval practice in the present study may be seen as evidence for matches and mismatches between initial processing and subsequent retrieval requirements.

## References

Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. Vol. 2, pp. 89–195). New York, NY: Academic Press.

Baddeley, A. D. (1966a). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology, 18,* 302–309.

Baddeley, A. D. (1966b). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology, 18,* 362–365.

Baddeley, A. D. (1986). *Working memory.* New York, NY: Clarendon Press/Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bwer (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York, NY: Academic Press.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459.

Bartlett, J. C., & Tulving, E. (1974). Effects of temporal and semantic encoding in immediate recall upon subsequent retrieval. *Journal of Verbal Learning and Verbal Behavior, 13,* 297–309.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33*(A), 497–505.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (0), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York, NY: Cambridge University Press.

Cowan, N. (2005). *Working memory capacity.* New York, NY: Psychology Press.

Craik, F. I. M. (1969). Modality effects in short-term storage. *Journal of Verbal Learning and Verbal Behavior, 8,* 658–664.

Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior, 9,* 143–148.

Craik, F. I. M. (2002). Levels of processing: Past, present . . . and future?. *Memory, 10,* 305–318.

Craik, F. I. M., Gardiner, J., & Watkins, M. (1970). Further evidence for a negative recency effect in free recall. *Journal of Verbal Learning and Verbal Behavior, 9,* 554–560.

Craik, F. I. M., & Levy, B. A. (1970). Semantic and acoustic information in primary memory. *Journal of Experimental Psychology, 86,* 77–82.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Craik, F. I. M., & Watkins, M. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 12,* 599–607.

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

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

Ericsson, K., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

Gathercole, S., & Baddeley, A. (1993). *Working memory and language.* Hillsdale, NJ: Erlbaum.

Glanzer, M. (1972). Storage mechanisms in recall. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 5, pp. 129–193). New York, NY: Academic Press.

Haberlandt, K. (1997). *Cognitive psychology* (2nd ed.). Boston, MA: Allyn & Bacon.

Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 691–702.

Hale, S., Myerson, J., Rhee, S. H., Weiss, C. S., & Abrams, R. A. (1996). Selective interference with the maintenance of location information in working memory. *Neuropsychology, 10,* 228–240.

Hitch, G. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology, 10,* 302–323.

Hyde, T., & Jenkins, J. (1973). Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior, 12,* 471–480.

Jacoby, L. L., & Bartz, W. H. (1972). Rehearsal and transfer to LTM. *Journal of Verbal Learning and Verbal Behavior, 11,* 561–565.

James, W. (1890). *The principles of psychology.* New York, NY: Holt, Rinehart & Winston.

Just, M., Carpenter, P., & Keller, T. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review, 103,* 773–780.

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

Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 704–719.

Karpicke, J. D., & Roediger, H. L., III. (2008, February 15). The critical importance of retrieval for learning. *Science, 319,* 966–968.

Kintsch, W., & Buschke, H. (1969). Homophones and synonyms in short-term memory. *Journal of Experimental Psychology, 80,* 403–407.

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

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

Logie, R., & Baddeley, A. (1987). Cognitive processes in counting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 310–326.

Madigan, S., & McCabe, L. (1971). Perfect recall and total forgetting: A problem for models of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 101–106.

Mazuryk, G. F., & Lockhart, R. S. (1974). Negative recency and levels of processing in free recall. *Canadian Journal of Psychology, 23,* 114–123.

McCabe, D. P. (2008). The role of covert retrieval in working memory span tasks: Evidence from delayed recall tests. *Journal of Memory and Language, 58,* 480–494.

McCabe, L., & Madigan, S. (1971). Negative effects of recency in recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 10,* 307–310.

Milner, B. (1966). Amnesia following operation on the temporal lobes. In C. W. M. Whitty & O. L. Zangwill (Eds.), *Amnesia* (pp. 109–133). London, England: Butterworth.

Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What's so special about working memory? An examination of the relationships among working memory, secondary memory, and fluid intelligence. *Psychological Science, 19,* 1071–1077.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16,* 519–533.

Murdock, B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482–488.

Neath, I., Brown, G. D. A., Poirier, M., & Fortin, C. (Eds.). (2005).

Short-term working memory: The Second Quebec Conference on Short-Term Working Memory [Special issue]. *Memory, 13*(3/4).

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms.* Retrieved from http://w3.usf.edu/FreeAssociation/

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 411–421.

Raven, J. C. (1958). *Guide to using the Mill Hill Vocabulary Scale with the Progressive Matrices Scales*. Oxford, England: H. K. Lewis.

Roediger, H. L., III, Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels-of-processing framework. *Memory, 10*(5/6), 319–332.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

Rundus, D., Loftus, G., & Atkinson, R. (1970). Immediate free recall and three-week delayed recognition. *Journal of Verbal Learning and Verbal Behavior, 9,* 684–688.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 546–554.

Shallice, T., & Warrington, E. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology, 22,* 261–273.

Shulman, H. G. (1970). Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 499–508.

Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin, 75,* 399–415.

Shute, V. (1991). Who is likely to acquire programming skills? *Journal of Educational Computing Research, 7,* 1–24.

Smith, E., Barresi, J., & Gross, A. (1971). Imaginal versus verbal coding and the primary–secondary memory distinction. *Journal of Verbal Learning and Verbal Behavior, 10,* 597–603.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability and familiarity. *Behavior Research Methods, 38,* 598–605.

Tulving, E. (1968). Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 2–36). Englewood Cliffs, NJ: Prentice-Hall.

Tulving, E. (1976). Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 37–73). Oxford, England: Wiley.

Tulving, E. (1983a). *Elements of episodic memory.* New York, NY: Oxford University Press.

Tulving, E. (1983b). How many memory systems are there? *American Psychologist, 4,* 385–398.

Tulving, E. (2000). Concepts of memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 33–43). New York, NY: Oxford University Press.

Unsworth, N. (2009). Examining variation in working memory capacity and retrieval in cued recall. *Memory, 17,* 386–396.

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

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132.

Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133,* 1038–1066.

Washington University. (2009). *Speech & Hearing Lab Neighborhood Database*. Retrieved from http://neighborhoodsearch.wustl.edu/neighborhood/Home.asp

Waugh, N., & Norman, D. (1965). Primary memory. *Psychological Review, 72,* 89–104.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: Psychological Corporation.

## Appendix

### Levels-of-Processing Span Task Word Triads From Experiment 2

| Set 1 | | | Set 2 | | | Set 3 | | |
|---|---|---|---|---|---|---|---|---|
| Target | Rhyme | Semantic | Target | Rhyme | Semantic | Target | Rhyme | Semantic |
| BELL | compel | ring | BAR | guitar | drink | BAND | gland | music |
| BOOT | compute | shoes | BEACH | screech | sand | BED | ahead | sleep |
| BRAIN | lane | think | BENCH | quench | park | BLAME | claim | accuse |
| BREAD | instead | butter | BEST | quest | worst | CAVE | brave | dark |
| CHOOSE | bruise | pick | BIKE | dislike | ride | CENT | meant | penny |
| CHURCH | search | god | BRIDE | denied | groom | CHAIR | square | table |
| COAL | enroll | black | BUG | shrug | insect | CHARM | disarm | bracelet |
| CORN | mourn | cob | CUTE | loot | pretty | CLERK | smirk | store |
| DASH | mustache | run | FATE | freight | destiny | CLOUD | crowd | rain |
| DAY | bouquet | night | FLAME | proclaim | fire | COLD | consoled | hot |
| FAIL | sale | pass | FOOD | argued | eat | DRAIN | reign | sink |
| FILL | quill | empty | GLAD | plaid | happy | FIX | picks | break |
| GAS | crass | fuel | GUEST | stressed | visitor | FLIGHT | campsite | airplane |
| GATE | create | fence | HEIGHT | site | tall | FOOT | put | shoe |
| GIRL | earl | boy | HEN | again | chicken | HARM | alarm | hurt |
| GREET | compete | hello | KIND | assigned | nice | HATE | equate | love |
| HALT | fault | stop | LEAK | critique | drip | JAM | program | jelly |
| HILL | goodwill | mountain | LOUD | vowed | noise | KEY | tea | lock |
| JAIL | exhale | prison | MOOD | crude | feeling | LIME | rhyme | lemon |
| KNEE | breezy | leg | MUD | flood | dirt | LOG | smog | wood |
| MAD | keypad | angry | NEAT | sweet | messy | PEACE | geese | war |
| PEARL | uncurl | necklace | PINE | sign | tree | PIPE | type | smoke |
| PLATE | weight | dish | POND | wand | lake | QUEEN | protein | king |
| RAKE | ache | leaves | RENT | accent | lease | RAISE | weekdays | lower |
| RULE | stool | law | RUSH | toothbrush | hurry | SAIL | female | boat |
| SHIFT | thrift | move | SACK | kayak | bag | SKIP | equip | jump |
| SPOT | swat | stain | SEA | three | ocean | SMILE | isle | frown |
| STIR | answer | mix | SON | none | daughter | SOFT | coughed | hard |
| SWIM | limb | water | SPEECH | each | talk | SPARE | affair | tire |
| TENT | comment | camp | STAIRS | cares | climb | STREET | elite | road |
| TOOL | module | hammer | THEME | beam | idea | THROAT | wrote | neck |
| TOWN | noun | city | THIN | begin | fat | TONGUE | young | lick |
| TRAP | mishap | mouse | VAGUE | plague | unclear | TRUCK | pluck | driver |
| VAST | harassed | large | WAX | tracks | candle | TRUTH | tollbooth | lie |
| WAIT | translate | patience | WIFE | wildlife | husband | WISE | eyes | smart |
| WORTH | birth | value | WIRE | choir | telephone | YEAR | tier | month |

Received December 16, 2008
Revision received November 12, 2009
Accepted November 16, 2009 ■