Behavior Research Methods
2007, 39 (3), 482-493

A group-administered lag task as a measure of working memory

Jill T. Shelton
Louisiana State University, Baton Rouge, Louisiana

Richard L. Metzger
University of Tennessee, Chattanooga, Tennessee

and

Emily M. Elliott
Louisiana State University, Baton Rouge, Louisiana

The purpose of the present studies was to evaluate the utility of a group-administered version of the *n*-back, or lag task. Experiments 1 and 2 describe the construction of the task and reveal that the modified lag task (MLT) produces the same performance trends as have been observed in individually administered versions of the lag task; performance decreased significantly as lag conditions increased in difficulty. Experiments 3 and 4 established convergent validity by comparing the MLT to another common working memory task, the operation–word span task, as well as the updated version of this task, the automatic operation span task. The results showed that MLT performance was significantly correlated to scores on both measures. These experiments provide important details about the MLT as a measure of working memory, in a group- or individual-administration setting.

The study of working memory (WM) has become a focal point of memory research over the past 20 years. WM has been defined as the storage and manipulation of processed information (Baddeley, 1999). The manipulation aspect of this definition is what distinguishes WM from traditional views of short-term memory (STM). The typical WM span task is made up of an interleaving set of to-be-remembered (TBR) information along with a secondary processing task, such as making decisions about math operations or reading a series of sentences (Daneman & Carpenter, 1980; Turner & Engle, 1989). The active maintenance of TBR information, through focal attention, is a common theme among WM tasks. Another common feature of traditional WM tasks is that they typically employ individual-administration techniques.

The present studies report the development and utility of a group-administered version of the *n*-back or lag task as a measure of WM. The ability to assess cognitive functioning in groups could be useful in laboratory or applied settings. Group assessment is a valuable prescreening tool in the laboratory setting because it allows researchers to easily identify the high- and low-functioning groups for further examination. Group assessment of cognitive functioning would also be useful in applied settings, such as in the armed forces, police academies, and other organizations. Mass screening would provide a more cost-efficient and less time-consuming method of evaluation. The potential implications of group assessment of WM function using traditional tasks have been discussed but not directly examined.

Conway et al. (2005) discussed the issues associated with group-administered versions of traditional span tasks such as reading span (Daneman & Carpenter, 1980), counting span (Case, Kurland, & Goldberg, 1982), and *operation–word span* (OSPAN; Turner & Engle, 1989). In group-administered versions of these tasks, participants who are faster at completing the secondary processing component (e.g., completing the math problem, reading sentences) may have additional time to spend rehearsing the TBR items. This could lead to greater error in the assessment of WM capacity and could alternatively be measuring the rehearsal strategies more highly associated with STM capacity. The lag task, however, should not be susceptible to the same kind of group-administration issues. In this task, participants view lists of items (typically digits or symbols), and at the end of the list they are prompted to identify an item that appeared in a particular position in the list (Dobbs & Rule, 1989). In the group-administered version of the lag task, stimulus items are presented at the same rate for everyone; therefore, none of the participants are at an advantage in terms of rehearsal time. Furthermore, the processing component is to remember the item that appeared in a particular location of the list, which should not be different for individual versus group versions of the task, considering that the processing requirement emerges during the retrieval stage.

J. T. Shelton, jshelt9@lsu.edu

It is important to include a processing requirement in any measure of WM; however, it is not essential for this requirement to be embedded in a dual-task format. Recently, several researchers have proposed that separate storage and processing requirements are not a necessary component of a valid WM task (Cowan et al., 2005; Haarmann, Davelaar, & Usher, 2003). In these studies, alternative WM tasks were found to be highly correlated to traditional dual storage-and-processing tasks, and were good predictors of higher order cognitive functioning. One alternative to the traditional dual-task span measures is the lag task. Demand is determined, at least in part, by the requirement that information be held active in memory until a cue appears requesting an item presented at some previous point in the information flow.

The lag task has been referred to as the "gold standard" of WM measures in the field of cognitive neuroscience (Kane & Engle, 2002); however, there is still some uncertainty about the properties of the task. Some researchers have requested more systematic evaluation of this task to support its usefulness as a measure of WM (Conway et al., 2005). The lag task was first developed by Kirchner (1958) to study memory retrieval processes. Both recall and recognition versions of the task have been developed. McElree (2001) used a recognition version of the lag task to examine the interaction between focal attention and retrieval speed. In this version of the task, participants viewed lists of letters containing 6 to 13 items, followed by a probe to match items at a position of Lag 1, 2, or 3. Friedman et al. (2006) used a spatial recognition version of the lag task to measure the ability to update and monitor information in WM. In this task, participants viewed a screen with darkened boxes and they had to report whether these boxes matched the darkened box that had been presented two trials back.

The recall procedure developed by Dobbs and Rule (1989) is closest to the one used in the present studies. Dobbs and Rule were interested in developing a procedure that measured WM, was useful with a variety of stimulus materials, and could be administered successfully to an aging population. In this study, the lag task consisted of the researcher reading lists of randomly ordered numbers to individual participants. At the end of each list of numbers, the participants were asked one of three questions: (1) What was the digit at Lag 0?; (2) What was the digit at Lag 1?; or (3) What was the digit at Lag 2? Dobbs and Rule found that performance declined as the lag conditions increased in difficulty, and this decline was more pronounced for the older participants. They suggested that performance on this task reflected people's abilities to manipulate information through WM. The lag task contains the characteristics typically present in a WM task. Controlled attention must be applied to the items being presented, in order to actively maintain the target information (the location of the item in the list) for later recall. In addition, it is necessary to inhibit irrelevant items presented in the list when attempting to recall the target item.

The Present Research

The primary goal of the present research was to examine the utility of a modified version of the lag task that allows for group administration, and we used Experiments 1 and 2 to accomplish this goal. The second objective was to test the reliability and validity of the *modified lag task* (MLT) as a measure of WM. Experiments 3 and 4 sought to accomplish this goal by testing convergent validity with a traditional measure of WM performance. We also conducted analyses in Experiment 4 to assess the reliability of the MLT.

EXPERIMENT 1

In the first experiment, we examined the materials and manipulations of the MLT. We modified the lag task; therefore, it was necessary to test the effectiveness of the new task. For example, we used words instead of digits or symbols, which could have altered the difficulty level of the task. However, the main factor we needed to consider was the group-administration technique, because the lag task has traditionally been administered to individual participants (Dobbs & Rule, 1989; McElree, 2001). We examined the performance trends to determine whether the previously observed trends (i.e., decrease in performance across lag conditions) would occur using group administration.

Method

Participants. A total of 62 introductory psychology students from the University of Tennessee at Chattanooga completed the MLT in groups of 20–25 people. All of the participants received extra credit in psychology courses for their participation. We did not record demographic information for Experiments 1 and 2, but we have included this information in Experiments 3 and 4.

MLT procedure. We constructed a variant of the lag task used by Dobbs and Rule (1989) for the present study. One difference was that we administered the task to groups of participants, and automated the task using Microsoft PowerPoint. A second difference was our use of words instead of digits or symbols. The words were generated from the MRC Psycholinguistic Database (www.psych.rl.ac.uk/MRC_Psych_Db.html) using specifications of length, familiarity, meaning, and frequency of usage. Finally, we presented a cue at the end of each list, prompting the recall of a word from a particular position in the list.

All of the words consisted of five to seven letters, and an equal number of words of each length were used for the task. Another specification was that the words had to have a high level of familiarity, meaning, and frequency of usage in the English language. The database generated approximately 1,500 words, and we randomly chose 189 words from this list to use in the first experiment. The words used for the MLT are presented in Appendix A. We then randomly assembled the words and placed them onto slides. In the actual lag exercise, a computer running Microsoft PowerPoint was used to display the information onto a white screen, via an overhead projector. The words were printed in a black, 44-point font, on a white background. At the end of each trial, a slide appeared with one of three questions: (1) What word was at Lag 0?; (2) What word was at Lag 1?; or (3) What word was at Lag 2?

The participants were shown 27 lists of randomly ordered words. We used three different list lengths for each of the three lag conditions: six, seven, and eight items. The participants completed three trials at each of the nine factorial combinations of lag $\times$ length. The main reason for using different list lengths was to prevent the participants from being able to predict the end of the list.

The participants were informed about the nature of the lag task, and the experimenter explained each lag condition. Then three practice trials were presented, consisting of one example of each lag condition. The participants were told that there would be varying numbers of words in each list. The practice trials were presented in the same form as the actual test trials. After the participants completed the practice trials, they were given the opportunity to ask questions and repeat the practice trials if necessary. Once the participants appeared to understand the task, they completed the test trials.

After the practice trials, we presented the lag task in an automated fashion. The slides indicating that the next trial was about to begin had a 5-sec duration, each word had a 1-sec presentation rate, and the recall cue remained on screen for 10 sec. The participants were given no indication of the lag position being requested for each trial prior to the recall cue slide. The participants were expected to write the appropriate one-word answer on a designated answer sheet for each trial, and the experimenter remained in the room to monitor this behavior. The scores were the sum of the correctly answered trials for each condition. Therefore, each participant had nine scores for the MLT, each with a maximum of three points. For example, each participant had a score for List Length 6, Lag 2, and List Length 8, Lag 1, and so on. This task took approximately 15 min to complete.

Results and Discussion

The results of the first experiment confirmed that group administration of the MLT produced outcomes similar to those previously reported with individual administration (Dobbs & Rule, 1989; McElree, 2001). We used a 3 (length) $\times$ 3 (lag) repeated measures ANOVA to analyze the lag scores; Table 1 presents the mean performance at each lag $\times$ length condition. All significant effects were based on an alpha level of .01 so we would have a more conservative test of this newly developed task. There was a main effect of lag condition [$F(2,122) = 62.43$, $MS_e = 0.703$, $\eta_p^2 = .506$]. As anticipated, the participants performed worse as the lag conditions increased in difficulty (92%, 81%, and 60%, respectively). The average percentage correct for each lag condition will be presented in this way throughout the article to allow for direct comparisons across experiments. There was also a main effect of length [$F(2,122) = 5.62$, $MS_e = 0.417$, $\eta_p^2 = .084$], suggesting that memory performance was less accurate when longer lists of words were used.

The length $\times$ lag interaction was significant as well [$F(4,244) = 3.78$, $MS_e = 0.447$, $\eta_p^2 = .058$]. This interaction was driven by the greater effect of length in the most difficult lag condition. However, as stated above, we manipulated the variable of length to prevent the prediction of the end of the list; therefore, the interaction of list length with lag condition is not of major theoretical interest.

Overall, these findings suggest that the MLT can be administered to groups and yield similar results to those reported by Dobbs and Rule; a significant decrease in performance was found as lag conditions increased in difficulty. Comparisons between the present study and previous research provided insight into the utility of the MLT, but we felt that it would also be informative to use the same stimulus materials to make comparisons for the MLT in a group versus individual setting. Experiment 2 was designed to make this comparison.

Table 1
Lag Task Means and Standard Deviations for Experiment 1

	Length 6		Length 7		Length 8	
	M	*SD*	*M*	*SD*	*M*	*SD*
Lag 0	2.76	0.67	2.79	0.60	2.76	0.62
Lag 1	2.60	0.78	2.21	0.98	2.50	0.84
Lag 2	2.05	1.05	1.79	1.12	1.60	1.11

Note—Maximum score per cell = 3.

EXPERIMENT 2

We had three primary objectives in the second experiment. First, we wanted to ensure that the group administration did not influence performance on the MLT in comparison with performance when using individual administration. Second, we wanted to determine whether the procedures used for group administration could be further adapted for individual administration. That is, would the same results be obtained when each participant was seated at his or her own computer, but all other procedures remained the same? Third, we created versions of the OSPAN and digit span tasks that could be administered in a group setting, and we wanted to see whether performance on these tasks would differ in group versus individual settings. We used the OSPAN and digit span tasks to assess the validity of the MLT in Experiments 3 and 4, but the purpose for including them in this experiment was to evaluate the utility of the group-administered versions of these tasks.

The mass group-administration technique used in Experiment 1 is appropriate for some situations—for example in classrooms or in settings with limited equipment; however, individual administration may be more appropriate in other settings, such as when individual administration is necessary for additional tasks used in the research session. Therefore, it was important to be sure the procedure would work well in either administration mode.

Method

Participants. A total of 48 undergraduate volunteers from the University of Tennessee at Chattanooga were randomly assigned to complete the tasks in a group ($n = 27$) or individual ($n = 21$) setting. Fifteen participants completed the tasks in both settings.

MLT. The task materials used to construct the MLT in Experiment 2 were identical to those used in Experiment 1. Again, Lags 0, 1, and 2 were used; the one difference was that in Experiment 2, only List Lengths 6 and 8 were used. Using only two list lengths allowed us to increase the number of trials for each condition. In this experiment, participants completed 5 trials at each lag/length condition, for a total of 30 MLT trials.

Digit span task. A forward digit span task, similar to that used in previous research, was used to assess STM capacity (Saito & Baddeley, 2004). The materials for this task consisted of five lists of numbers that were placed onto Microsoft PowerPoint slides. The digits were presented individually, one per second, on a white projector screen. At the end of each list of digits, a slide with a question mark was presented, cuing participants to recall all of the digits from the list, in the order in which they were presented. The digits 1–9 were used for each list; they were randomly ordered among lists. The number of digits in each list ranged from five to nine (Saito & Baddeley). A participant's digit span was the number of digits in the longest list he or she could correctly recall. For instance, if he or she correctly recalled the digit lists with five and six numbers, but incorrectly recalled the list with seven numbers, then his or her digit span would be six. If the participant could not accurately recall the shortest digit list offered (five digits), he or she was assigned a digit span of four. The task took approximately 5 min to complete.

OSPAN task. We reconstructed the Turner and Engle (1989) OSPAN task using materials from Cantor and Engle (1993). We placed the test materials onto PowerPoint slides. The operation portion of this task consisted of simple two-step mathematical operations. Answers to each mathematical operation were presented at the end of the operation; the answer was correct half of the time. The participants responded to the correctness of the answer by circling "yes" or "no" on an answer sheet.

Each math operation was presented with a TBR word (e.g., [8/2] − 1 = 3 TOOL). The words were high-frequency one-syllable words utilized by La Pointe and Engle (1990). The math operations and words were randomly paired with one another. The participants saw 12 trials of math operation–word pairs presented individually on slides. Each trial consisted of four to seven slides; there were 3 trials with each list length (four, five, six, and seven words). Therefore, each participant viewed 66 operations and TBR words. At the end of each trial, the participants were cued by a question mark to recall all of the words from that set, in the order in which they were presented. After a practice trial was presented, the slides were presented in a timed, automated fashion. The slides indicating that the next trial was about to begin appeared for 5 sec. The slides presenting the math operations and words had an 8-sec duration. Finally, the slides cuing the recall of the words remained on the screen for 15 sec. We determined the participants' span scores by totaling all of the words for trials they answered completely correctly. Scores on the OSPAN task could range from 0 to 66. In line with previous research, we dropped participants from the analyses if they did not respond accurately to at least 80% of the math problems (Elliott, Barrilleaux, & Cowan, 2006).

Procedure. First, the participants were randomly assigned to complete the tasks using either group- or individual-administration techniques. Approximately one week after completing the tasks in one administration condition, 15 participants returned to complete the tasks in the other administration condition. Nine participants completed the group-administered version first, and 6 completed the individual-administration condition first. It is important to make comparisons between the participants who completed each condition first, because simply making within-group comparisons of group- versus individual-administration performance would fail to take any potential practice effects into account. In the group condition, the administration technique was identical to that used in Experiment 1. A group of participants (n = 27) sat in a classroom and viewed the stimulus materials for all three tasks on a wall screen, presented using an overhead projector. They recorded their answers for all three tasks on predesignated answer sheets. In the individual condition, the participants were all seated at their own computers, and 5–10 participants completed the tasks at one time. The same test materials were used for both conditions, and the instructions were read aloud to participants in both conditions. The difference between the group- and individual-administration conditions was that in the individual condition, each person was seated at his or her own computer and viewed the stimuli on the computer screen as opposed to seeing them on a projector screen, as presented by the overhead projector in the group condition. In both conditions, participants filled out consent forms and then completed all three memory tasks: the forward digit span task, the OSPAN task, and the MLT. The entire session lasted approximately 45–50 min.

Results and Discussion

We used several approaches to analyze the data. First, we analyzed lag performance using a mixed-model ANOVA, with administration as a between-subjects variable and length and lag as within-subjects variables. Second, we ran an independent-samples t test with OSPAN and digit span scores as dependent measures, and administration as the grouping factor, to determine whether there were performance differences for those people who completed the tasks in group versus individual settings. Third, we ran a 2 (administration) × 2 (length) × 3 (lag) repeated measures ANOVA on the data from the 15 participants who completed both sessions, to determine whether lag performance differed when they completed the task in a group versus individual setting. We also conducted a paired-samples t test to examine potential within-subjects differences for OSPAN and digit span scores. Table 2 presents average performance in the group- and individual-administration conditions for each lag condition (collapsed across lengths), the OSPAN task, and the digit span task.

The first analysis produced a main effect of lag, indicating that performance decreased as lag conditions increased in difficulty [$F(2,92) = 67.78$, $MS_e = 1.361$, $\eta_p^2 = .596$]. There was no main effect of length [$F(1,46) = 1.15$, $MS_e = 0.547$, $\eta_p^2 = .024$] or lag × length interaction [$F(2,92) = 2.32$, $MS_e = 0.609$, $\eta_p^2 = .048$]. This finding suggests that length was not an important variable in this analysis. Most importantly, there was no main effect of administration [$F(1,46) = .929$, $MS_e = 7.905$, $\eta_p^2 = .02$], indicating that there were no significant differences in lag performance in the group (Lags 0, 1, and 2 = 89%, 79%, 49%) versus the individual (93%, 85%, 57%, respectively) administration conditions. Furthermore, the administration variable did not interact with length or lag in any way.

The results of the second analysis revealed that there were no significant differences in OSPAN scores for those who first completed the tasks in group versus individual settings [$t(46) = 0.171$]. The same was true for digit span scores [$t(46) = 0.990$]. Finally, the third set of analyses concerning the data for those participants who completed the tasks in both individual and group conditions demonstrated that the administration variable was not a significant factor for any of the tasks being examined. Participants performed similarly on the MLT [$F(1,14) = 1.46$, $MS_e = 2.577$, $\eta_p^2 = .097$], OSPAN task [$t(14) = 0.668$], and digit span task [$t(14) = 1.047$], regardless of whether they were in a group or an individual setting.

The results of Experiment 2 demonstrated that the same performance trends were observed whether the MLT was administered in a group or an individual setting. In addition, the modifications applied to the OSPAN and digit span tasks did not affect their utility in a group setting. One caveat to this finding is that the individual-administration approach used in this study is still different from the typical method used for OSPAN administration. Traditionally, in the OSPAN task, a participant responds aloud to the

Table 2
Lag Task Means and Standard Deviations for Experiment 2

	Group Administration (n = 27)		Individual Administration (n = 21)	
	M	SD	M	SD
Lag 0	8.85	2.43	9.33	2.20
Lag 1	7.85	2.57	8.48	2.79
Lag 2	4.89	3.00	5.71	2.88
OSPAN	30.74	17.89	29.81	19.63
Digit span	6.52	1.74	6.71	1.31

correctness of the math problems and the experimenter immediately moves on to the next trial. This minimizes the time spent rehearsing the TBR word. The implications of this difference are discussed further in the Discussion section of Experiment 3. The most important outcome of Experiment 2 was that the MLT proved to be an appropriate task for group administration. Experiment 3 was used to make comparisons among the group-administered versions of all three tasks used here.

EXPERIMENT 3

Our main objective in Experiment 3 was to explore the validity of the MLT using a larger sample of participants. We included the two additional memory tasks in the procedure in Experiment 2 to provide converging evidence of the validity of the MLT. As before, we used a digit span task to assess STM capacity, and we used the OSPAN task to assess WM capacity. WM tasks have been shown to be more highly correlated with one another than with measures of STM (Engle, Tuholski, Laughlin, & Conway, 1999).

The investigation of the MLT as a measure of WM in group settings has led to the formulation of three hypotheses. First, the group administration of the MLT should produce a significant decrease in performance across lag conditions, as seen in studies using individual participants. Second, performance on the lag task should be significantly correlated with performance on the OSPAN task, thus demonstrating that the MLT shares variance with other WM tasks. Third, MLT performance should be more correlated with the OSPAN task than with performance on a measure of STM, in this case, the digit span task.

To evaluate the second hypothesis, we calculated an alternative index of overall lag performance, which we refer to as lag score. This index is more comparable to the OSPAN index, with more credit earned for responding correctly to more difficult trials. Lag score is a weighted summary of MLT performance calculated by multiplying the number correct at each lag by a weighting factor represented by the lag:

$$\text{Lag score} = (\text{Lag 1 \# correct} * 1) + (\text{Lag 2 \# correct} * 2) + (\text{Lag 3 \# correct} * 3).$$

Method

Participants. The participants were 87 students from undergraduate psychology courses at the University of Tennessee at Chattanooga who volunteered to receive extra course credit. The data from 3 participants were not used because their performance on the math portion of the OSPAN task was below 80%. Of the remaining 84 participants, 18 were males and 66 were females, with a mean age of 24.88 years. No gender differences were found among any of the tasks.

MLT. The MLT procedure used in the third experiment was identical to that used in Experiment 2, except that Lags 1, 2, and 3 were used in this experiment. We made this change in order to avoid the ceiling effect observed in the Lag 0 condition. Participants were shown 30 lists of randomly ordered words. There were 15 lists consisting of six words, and 15 lists consisting of eight words. Each lag condition was employed on an equal number of trials; the lag conditions were used randomly across the 30 trials. In other words, there were five trials of each of the three lag conditions using both list lengths (e.g., List Length 6/Lag 3, List Length 8/Lag 1).

Procedure. In this experiment, the participants completed the MLT, the OSPAN task, and the digit span task in groups of 20–25 people. The OSPAN and digit span tasks were identical to those used in Experiment 2. We used partial counterbalancing to vary the order of task presentation; all three tasks appeared in each ordinal position of task administration.

Results and Discussion

We approached the analyses with two strategies. First, we investigated how the pattern of results produced by the MLT compared with the pattern of results described by previous research. Second, we investigated the relationship between MLT performance and the validation measures embedded in the design.

The first hypothesis was that there would be a significant decrease in performance across lag conditions, suggesting that the group-administered MLT was producing the same outcomes as have been seen in individually administered recall versions of the lag task (e.g., Dobbs & Rule, 1989). In addition, we were attempting to replicate the MLT performance trends observed in Experiments 1 and 2. We used a 2 (length) $\times$ 3 (lag) repeated measures ANOVA to assess MLT performance. There was a main effect of lag, confirming this hypothesis [$F(2,166) = 157.54$, $MS_e = 1.34$, $\eta_p^2 = .66$]. Performance decreased as the lag conditions increased in difficulty (75%, 47%, 30%, respectively). We conducted a Bonferroni test to determine whether average performance in each lag condition differed, and the results revealed a significant difference between mean performance in each lag condition. Table 3 lists the average performance for each of the six lag/length conditions; there were five trials per condition. Table 4 presents the descriptive statistics for each lag condition (collapsed across length), as well as the lag score index, digit span scores, and OSPAN scores.

There was no significant main effect of length [$F(1,83) = 3.91$, $MS_e = 0.938$, $\eta_p^2 = .05$] and no lag $\times$ length interaction [$F(2,166) = 2.7$, $MS_e = 0.98$, $\eta_p^2 = .03$]. There was, however, some indication that list length had a greater impact in the more difficult lag conditions. Due to the inconsistent length effects, we are hesitant to make inferences about the role of list length in the MLT. We still encourage the use of different list lengths, because blocking the ability to predict the end of the list forces participants to maintain higher levels of focal attention on list items.

The second hypothesis predicted that a significant positive relationship would exist between performance on the OSPAN task and performance on the MLT, providing a

Table 3
Lag Task Means and Standard Deviations for Experiment 3

	Length 6		Length 8	
	M	*SD*	*M*	*SD*
Lag 1	3.67	1.39	3.79	1.55
Lag 2	2.52	1.57	2.21	1.50
Lag 3	1.67	1.51	1.35	1.25

Note—$N = 84$.

Table 4
Descriptive Statistics for Memory Tasks in Experiment 3

Memory Task	Score Minimum	Maximum	Mean	Standard
Lag 1	0	10	7.45	2.69
Lag 2	0	10	4.74	2.72
Lag 3	0	10	3.01	2.31
Lag score	0	54	25.96	13.03
Digit span	4	8	4.88	2.66
OSPAN	0	66	30.4	19.14

Note—$N = 84$.

measure of convergent validity. We conducted analyses to determine whether each individual lag condition was significantly correlated with OSPAN performance. The weighted score for each lag condition, which was used to calculate lag score, was used for the correlational analysis since it was more comparable to the OSPAN index. Table 5 illustrates all of the correlations between individual lag conditions (based on the individual lag weighted scores) and the other tasks. The results demonstrated that performance at Lags 1, 2, and 3 shared a positive relationship with performance on the OSPAN task (rs = .39, .44, .43, respectively).

The third and final hypothesis dealt with the relationship between MLT and digit span performance. Performance on the MLT was predicted to have a stronger relationship with OSPAN than it would with digit span performance, thus demonstrating that the MLT was more similar to a measure of WM than a measure of STM. As indicated previously, there was a significant relationship between OSPAN and lag performance. Conversely, digit span performance was not significantly correlated with the lag score index ($r = .28$). Furthermore, digit span was not significantly correlated to performance in the Lag 1 and 3 conditions (rs = .07, .23, respectively). Digit span and Lag 2 performance did share a significant positive relationship ($r = .33$). The most important observation here was that overall lag task performance was significantly correlated to a measure of WM, the OSPAN task, and not significantly correlated to a measure of STM, the digit span task.

Interestingly, OSPAN was significantly correlated to digit span ($r = .38$). This finding could reflect the overestimation of OSPAN, due to the techniques used for group administration. The descriptive analysis for the OSPAN task indicated that average performance (M = 30.4, SD = 19.14) on the group-administered version of this task was considerably higher and more variable than is typically observed when this task is used with individual participants (M = 14.6, SD = 7.5; Engle et al., 1999). One of the reasons this task is more appropriate when performed individually is that people's spans may be over- or underestimated in a group setting because of the length of time it takes to answer the math problems. When individuals are tested with the OSPAN task, they respond to the correctness of the math problem aloud and the researcher immediately moves on to the next math problem–word pair. Therefore, one could assume that the participant has no extra time to rehearse the TBR word before moving on to the next problem. In the present research, the math problems were presented for 8 sec each, and the groups of participants recorded their answers. Those people who finished the math problem quickly had more time to rehearse the TBR word, whereas those who did not answer quickly had little to no time to rehearse the word. The fact that average performance on the OSPAN task in Experiment 3 exceeded average performance typically observed using individual administration suggests that participants' OSPAN was overestimated in the present study (Engle et al.). A different version of the OSPAN task, the automatic operation span task, was used in Experiment 4.

We conducted a final analysis to determine whether the order of task presentation (counterbalancing of task order) affected the outcome of the study. We conducted a one-way ANOVA with OSPAN, digit span, and lag score as dependent measures, and procedure order as the independent measure. The results demonstrated that the order of task presentation did not affect performance.

Experiment 3 provides evidence that the MLT is a valid measure of WM. MLT performance produced the predicted decline as a function of lag. The pattern of results found when using group administration mirrors the results found using individual-administration techniques. The pattern of correlations between the OSPAN, lag score, and digit span was consistent with predictions based on the observations of Engle et al. (1999). Engle et al. reported the correlation between OSPAN and digit span as .31, whereas the present data showed a correlation of .38. OSPAN was correlated with lag score, and both had lower correlations with digit span. As previously mentioned, there were some problems with the version of OSPAN used in this study. The digit span task we used was also problematic, because only one list was presented at each list length, and participants who could not accurately recall the shortest digit list of five items automatically received a digit span of four. Experiment 4 sought to remedy this problem by using a more reliable version of these tasks, as well as an updated version of the MLT.

Table 5
Correlations Between Performance on Memory Tasks in Experiment 3

Memory Task	1	2	3	4	5	6
1. Lag 1	–					
2. Lag 2	.69	–				
3. Lag 3	.47	.64	–			
4. OSPAN	.39	.44	.43	–		
5. Digit span	.07	.33	.23	.38	–	
6. Lag score	.74	.90	.89	.49	.28	–

Note—Correlations greater than .3 are significant at $p < .01$. $N = 84$.

EXPERIMENT 4

The purpose of this experiment was to test the reliability and validity of the MLT using individual-administration techniques, which are more commonly used in laboratory studies. We also used different versions of all three tasks, to reflect typical methods currently being utilized in the laboratory setting.

Method

Participants. The data from 78 undergraduates from Louisiana State University were used for this experiment. All of the participants received extra credit in psychology courses for their participation. Two of the original 80 participants were excluded from the analyses: one because English was not his/her first language and the other because he/she scored below 80% on the math accuracy portion of the operation span task. Of the remaining 78 participants, 16 were males and 62 were females, with a mean age of 20.34 years. There were no gender differences among any of the tasks.

MLT. This task was designed very similarly to the versions used in the previous experiments, but there were some differences. First, the task was programmed using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002). This allowed for easier scoring of the task and gave us the opportunity to randomly generate the word lists for each participant. The materials for the task were words generated from the MRC Psycholinguistic Database. The word bank was generated according to certain parameters: All words were one-syllable nouns with four to five letters, and they had a high level of frequency and familiarity in the English language. The generated list contained nearly 500 words, and we randomly selected 280 words from this list to be used for the task. The words used for the MLT in Experiment 4 are presented in Appendix B.

The MLT consisted of 40 lists of words: Half of the lists contained six words and the other half contained four words. Shorter list lengths were used in this experiment because in the previous experiments, participants reported ignoring the first few words of the list (i.e., predicting that they would never be asked to recall these words). Lag Conditions 0, 1, 2, and 3 were all used in this study. There were five trials at each lag/list length condition. The program was designed to randomly generate the word lists from the predesignated word bank for each participant. The lag/list length conditions were initially assigned randomly throughout the trials, but they remained constant across participants. Each word was presented at a rate of one per second. At the end of each list, a recall screen appeared and prompted the participant to recall either the last word in the list, or the word that was 1-back, 2-back, or 3-back. Five practice trials were presented prior to the actual task in order to acquaint the participants with the lag conditions. Participants used the keyboard to type their responses and then pressed ENTER to continue to the next list. There was no penalty for spelling errors. Full credit was given for misspelled responses as long as it was clear that the participants were attempting to report the correct answer.

We used a variety of scoring techniques to assess performance on the MLT. First, we gave participants scores for the total number of trials recalled correctly in each lag/list length condition. We also gave scores for the total number of trials recalled correctly in each lag condition, collapsed across list length. Third, we calculated the weighted score used in the previous experiment, referred to as lag score. We calculated this cumulative index using the following formula:

$$\begin{aligned}\text{Lag score} &= (\text{Lag 0 \# correct} * 1) + (\text{Lag 1 \# correct} * 2)\\ &\quad + (\text{Lag 2 \# correct} * 3) + (\text{Lag 3 \# correct} * 4).\end{aligned}$$

Automatic operation span task. An alternative version of the OSPAN task, the *automatic operation span* (AOSPAN) task, was constructed recently (Unsworth, Heitz, Schrock, & Engle, 2005) and was used for the present study. The AOSPAN task was also administered using E-Prime software. Recently, the AOSPAN task has been shown to be a reliable and valid measure of WM capacity. This version of the task is similar to the OSPAN task used in the previous experiment in that it still consists of a processing and storage component. Participants have to respond to the correctness of simple math problems while simultaneously retaining lists of items. One difference is that these items are letters instead of words. Also, instead of having to generate the list of items at the end of the trial, participants are provided a list of letters and have to choose the appropriate letters in the correct order. Thus, serial reconstruction of the list, as opposed to pure recall of the list, is the memory requirement in this version of the task. A detailed account of this task is provided by Unsworth et al.

We generated multiple performance indices for this task. First, OSPAN score is a weighted score formed by taking the sum of letters from trials with 100% accuracy in the letter-identification portion. Second, OSPAN total is the total number of letters correctly identified from the task, independent of whether a particular list was answered correctly. OSPAN score is the most comparable to lag score, considering that additional credit is earned in the more difficult test trials; therefore, OSPAN score will be used in the analyses.

Digit span task. We used a more reliable form of the digit span task in this experiment. This task was also administered using E-Prime software. In this task, participants viewed lists of digits and were asked to recall the digits in the order they were presented. The digits were presented at a rate of one per second, and the participants responded by typing in the digits on the number keypad. They viewed four digit lists at each list length, ranging from 3–9 digits per list. If they accurately recalled at least one list at a particular list length, then they moved on to the next consecutive list length. If they could not accurately recall the digits from at least one list at a particular list length, then the task was terminated.

We assessed performance on this task using two indices. The integer index represents the longest list length at which participants accurately recalled at least one list. We calculated the cumulative index by first finding the lowest list length at which all four lists were answered correctly, and establishing that as the base score. Then we gave each participant .25 point for each accurately recalled list above the base score. The cumulative score has been shown to be a more sensitive index of digit span performance, and will therefore be used in the present study (Elliott, 2002).

Procedure. After filling out informed consent forms, the participants completed the MLT, AOSPAN, and digit span tasks during one session. We used partial counterbalancing to minimize potential order effects. There were three procedure orders, so that all of the tasks appeared in each ordinal position at least once. The test session lasted approximately 45 min.

Results and Discussion

The purpose of Experiment 4 was to demonstrate that the MLT would produce a similar pattern of results using individual-administration techniques. In addition, we used alternative versions of all of the memory tasks that are more representative of the kinds of tasks currently being used in laboratory research. The results of Experiment 4 replicated some aspects of the previous experiments, but some interesting differences did arise.

The data were analyzed similarly to those in the previous experiments. First, we conducted a 2×4 repeated measures ANOVA, with lag and length as within-subjects factors, to examine the performance on the MLT. Second, we ran correlational analyses to determine the relationship between MLT, AOSPAN, and digit span performance. Finally, we conducted reliability analyses on the MLT.

The results of the repeated measures analysis further replicated the finding that the MLT produced a pattern of results similar to those typically observed in recall versions of the lag task (Dobbs & Rule, 1989). There was a main effect of lag [$F(3,231) = 269.60$, $MS_e = 1.583$, $\eta_p^2 = .78$], demonstrating that as the lag conditions increased in difficulty, performance significantly decreased (94%, 74%, 32%, 26%, respectively). There was also a main effect of

Table 6
Descriptive Statistics for Memory Tasks in Experiment 4

Memory Task	Score			
	Minimum	Maximum	Mean	*SD*
Lag 0	7	10	9.45	.83
Lag 1	0	10	7.36	2.58
Lag 2	0	9	3.23	2.42
Lag 3	0	7	2.56	1.97
Lag score	14	79	44.22	15.41
Cumulative span	4	9	6.47	1.10
AOSPAN	0	75	38.46	18.07

Note—$N = 78$.

length [$F(1,77) = 19.31$, $MS_e = 0.968$, $\eta_p^2 = .20$], but no length × lag interaction [$F(3,231) = 2.63$, $MS_e = 0.886$, $\eta_p^2 = .033$]. This suggests that the longer list length also resulted in performance decrements on the MLT. Table 6 presents the descriptive statistics for each lag condition (collapsed across length), OSPAN scores, digit span scores, and the lag score index.

We conducted a correlational analysis to examine the relationships between the various MLT conditions with AOSPAN and digit span performance. Table 7 lists the correlations between the individual lag conditions (based on the weighted scores for each condition), OSPAN scores, digit span scores, and the lag score index. In support of the findings from Experiment 3, there was a significant correlation between lag score and AOSPAN score ($r = .51$). To examine the relationships between performance on the individual lag conditions with OSPAN score, we used the weighted versions of each individual lag condition in order to make them more comparable to OSPAN score. The results revealed a significant correlation between OSPAN score and Lag 1 and Lag 2 conditions (rs = .38, .54, respectively). There was not, however, a significant correlation between OSPAN score and the Lag 0 and Lag 3 conditions (rs = .05, .27, respectively). There was a lack of variability in these two conditions, resulting in a ceiling effect at Lag 0 and a positive skew at Lag 3. Thus, we calculated a lag score index excluding these two conditions. The new index, which only contained performance at Lag 1 and Lag 2, had a marginally higher correlation with OSPAN score ($r = .57$).

Some results of the correlational analyses were somewhat different from those observed in Experiment 4. There was a significant correlation between lag score and digit span performance ($r = .53$). Digit span performance was significantly correlated to the Lag 1, 2, and 3 conditions (rs = .38, .48, .35, respectively). There was not, however, a significant relationship between digit span and Lag 0 performance ($r = .09$). An additional analysis was done to determine whether using the lag score index that only included Lag 1 and 2 scores would alter its relationship with digit span, and no major changes were observed ($r = .52$). Digit span performance was also significantly correlated to OSPAN score ($r = .69$).

Table 7
Correlations Between Performance on Memory Tasks in Experiment 4

Memory Task	1	2	3	4	5	6
1. Lag 0	–					
2. Lag 1	.03	–				
3. Lag 2	−.04	.37	–			
4. Lag 3	−.22	.20	.52	–		
5. Lag score	−.06	.60	.84	.80	–	
6. Cumulative span	.09	.38	.48	.35	.53	–
7. AOSPAN	.05	.38	.54	.27	.51	.69

Note—Correlations greater than .3 are significant at $p < .01$. $N = 78$.

We assessed internal consistency of the MLT in order to establish reliability. We calculated Cronbach's alpha for each of the four lag conditions. The Lag 1 and 2 conditions both revealed high levels of reliability ($\alpha = .78$ and .70, respectively). The Lag 3 condition produced a moderate level of reliability ($\alpha = .58$), and the Lag 0 condition a low level of reliability ($\alpha = .24$). The lack of variability found in the Lag 0 and Lag 3 conditions was most likely responsible for the lower levels of reliability observed. This lack of variability could suppress the correlations used to calculate the alpha coefficient, as well as other statistical analyses that utilize scores from these conditions. The Lag 0 and Lag 3 conditions could still be useful for descriptive purposes; however, the Lag 1 and 2 conditions offer the most reliable assessment of MLT performance in this population of undergraduate students. Use of this task with children, older adults, or other special populations may require use of different lag conditions.

GENERAL DISCUSSION

The goal of the present studies was to determine whether the modifications applied to the lag task would alter its measurement capabilities, primarily its ability to accurately measure WM function in a group setting. The first modification to the lag task was to place the materials onto slides so that the task could be administered in groups, as opposed to the traditional individual administration. The second modification was to use words rather than digits or symbols. Scores demonstrated that participants could still perform the lag task, in spite of the word modification. Also, mean scores demonstrated that the MLT produced the same outcome as has been seen in individual administration of the lag task—a decrease in performance across lag conditions. Finally, another goal was to determine whether, after the modifications were applied to the MLT, it would still relate to other WM tasks, in this case the OSPAN and AOSPAN tasks. The significant relationships among the performance results on these tasks demonstrated that it would.

The most notable finding was that this task was an appropriate measurement tool, even in a group setting. The main effect of lag confirmed that the MLT produced the same outcome in a group setting as it did using individual testing. As the lag conditions increased in difficulty, performance declined considerably. This is an intriguing finding, because researchers are often trying to assess the WM function of more than one individual. The MLT would allow for this type of testing, since it is constructed on slides and viewed using an overhead projector. One main advantage of the E-Prime version of the task is computerized scoring. (The MLT program is available from the first author upon request.)

The findings of the present studies support the notion that the MLT is a good measure of WM performance. Scores on this task were significantly correlated to performance on the OSPAN, a widely used measure of WM capacity, as well as the more recent AOSPAN task. Furthermore, in Experiment 3, MLT performance was not significantly correlated to digit span, a widely used measure of STM function. However, problems with the tasks in Experiment 3 limit the ability to draw firm conclusions. In Experiment 4, MLT did share a significant relationship with digit span. This is not surprising, given that other studies have demonstrated that whereas WM tasks should be more highly correlated with one another, they are still related to STM tasks (Engle et al., 1999). Another interesting piece of information related to these findings came from verbal feedback given by the participants after they had completed the study. When asked about the difficulty level of the MLT, participants' replies tended to relate to the focus of attention required to carry out the task. As previously discussed, focal attention is a distinguishing factor between WM and STM tasks, and the lag task has been used to study focal attention processes (McElree, 2001). Thus, the focal attention required by the lag task suggests that it is drawing from WM resources.

Another finding worth noting was the strong correlation observed between digit span and AOSPAN performance. In fact, they were more highly correlated with one another than were AOSPAN and MLT performance. The differential relationships between these three tasks could offer some insight into the processes being measured by each. Given that a fixed set of the digits 1–9 was used for the digit span task, both the AOSPAN and digit span tasks required serial reconstruction of list items, whereas the MLT did not. This serial reconstruction requirement for both the digit span and AOSPAN tasks could have greatly contributed to the strong relationship between the two. A unique aspect of the MLT is that it requires participants to continually update the contents of WM in order to recall the items at particular positions in the list. A common feature between the AOSPAN and MLT tasks, which is not present in the digit span task, is that they both require the active manipulation of information in addition to the storage requirement. Although the zero-order correlations between the tasks provide some insight into the similarities and differences between these tasks, their predictive capabilities would be more informative. Unsworth et al. (2005) addressed this point in their evaluation of AOSPAN by noting that the zero-order correlations between memory tasks are less important than their ability to predict a criterion construct, such as general fluid intelligence. WM tasks tend to be good predictors of fluid intelligence, whereas STM tasks do not.

A particularly interesting finding that deserves some discussion is that Lag 2 scores were the most highly correlated to digit span and AOSPAN performance. A possible explanation for this finding could be that this is the point where STM and WM function overlaps, as discussed by Engle et al. (1999). They thoroughly discussed the distinction between STM and WM. STM is primarily used for the retrieval of recently encoded information, whereas WM requires the active manipulation of the TBR or some secondary set of information. Engle et al. argue that the extent to which both STM and WM tasks require memory strategies, such as rehearsal, could determine their level of overlap. It is possible that the Lag 2 condition requires greater use of these strategies than the other lag conditions, putting it in a unique position to display this overlap. This interplay between STM and WM processes has made it difficult to define a clear distinction between them. Engle et al. noted that there may not be a pure measure of STM or WM due to this overlap.

The present research has several interesting implications. The most notable implications emerge from the actual modifications applied to the lag task. First, group assessment of cognitive functioning using the MLT would be a valuable resource for saving time and money, especially in applied settings. In addition, group testing is useful for prescreening individuals to evaluate their potential role in future research projects. Second, the use of words instead of numbers or symbols allows for greater flexibility in the manipulation of the stimuli. Using words provides a variety of opportunities for assessing memory function in many different domains. Furthermore, the lag task was modified to be an automated task using Microsoft PowerPoint or E-Prime, thus requiring little effort from the task administrator and minimizing the possibility for error. More importantly, this modification is useful for the standardization of test materials. The materials and timing of the practice and test trials will be the same for every administration. Instruction on how to explain the task to participants will be the only training needed to successfully administer the MLT. For all of these reasons, the MLT has the potential to be a particularly useful measurement tool in many settings.

There are several future directions for research with the MLT that should be considered. First, *structural equation modeling* (SEM) techniques using multiple memory tasks, including the MLT, would provide a more powerful approach to understanding the shared variance among the MLT and other measures of WM function. In addition, a criterion construct could be used to determine the predictive power of the MLT for other higher order cognitive skills. One study using SEM demonstrated that a recognition version of the lag task loaded on a WM updating factor that was highly related to both a latent construct defined as general fluid intelligence and general crystallized intelligence (Friedman et al., 2006). It would be informative to use a similar approach with MLT scores.

Another interesting idea for future research would be to use the MLT for the examination of individual difference factors. For example, we believe that the valence of the words used in the MLT could be manipulated to detect depression among participants. Mathews, Ridgeway, and Williamson (1996) used the attention deployment task to examine the responses of emotionally disturbed individuals to threatening stimuli. Participants were presented with pairs of words, one being neutral and the other threatening. Response latencies demonstrated that depressed individuals were biased toward threatening words. Using words as stimuli for the MLT allows for such interesting manipulations. The present studies provide a good foundation for carrying out future research using the MLT as a measure of WM function.

AUTHOR NOTE

Portions of this research served as partial fulfillment of the requirements for J.T.S.'s master's thesis at the University of Tennessee at Chattanooga. We thank Amye Warren, Jason Hicks, and Sean Lane for helpful comments on the manuscript, and Mike Biderman for his statistical consultation. Correspondence concerning this article should be addressed to J. T. Shelton, Department of Psychology, 236 Audubon Hall, Louisiana State University, Baton Rouge, LA 70803 (e-mail: jshelt9@lsu.edu).

REFERENCES

BADDELEY, A. D. (1999). *Essentials of human memory*. East Sussex, U.K.: Psychology Press.

CANTOR, J., & ENGLE, R. (1993). Working-memory capacity as long-term memory activation: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1101-1114.

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

CONWAY, A. R. A, KANE, M. J., BUNTING, M. F., HAMBRICK, D. Z., WILHELM, O., & ENGLE, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, **12**, 769-786.

COWAN, N., ELLIOTT, E. M., SAULTS, J. S., MOREY, C. C., MATTOX, S., HISMJATULLINA, A., & CONWAY, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, **51**, 42-100.

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

DOBBS, A. R., & RULE, B. G. (1989). Adult age differences in working memory. *Psychology & Aging*, **4**, 500-503.

ELLIOTT, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition*, **30**, 478-487.

ELLIOTT, E. M., BARRILLEAUX, K. M., & COWAN, N. (2006). Individual differences in the ability to avoid distracting sounds. *European Journal of Cognitive Psychology*, **18**, 90-108.

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.

FRIEDMAN, N. P., MIYAKE, A., CORLEY, R. P., YOUNG, S. E., DEFRIES, J. C., & HEWITT, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, **17**, 172-179.

HAARMANN, H. J., DAVELAAR, E. J., & USHER, M. (2003). Individual differences in semantic short-term memory capacity and reading comprehension. *Journal of Memory & Language*, **48**, 320-345.

KANE, M., & ENGLE, R. W. (2002). The role of prefrontal cortex in working memory, capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychological Bulletin & Review*, **9**, 637-671.

KIRCHNER, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, **55**, 352-358.

LA POINTE, L. B., & ENGLE, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 1118-1133.

MATHEWS, A., RIDGEWAY, V., & WILLIAMSON, D. A. (1996). Evidence for attention to threatening stimuli in depression. *Behaviour Research & Therapy*, **34**, 695-705.

MCELREE, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 817-835.

MRC PSYCHOLINGUISTIC DATABASE (n.d.) Retrieved July 5, 2007 from www.psych.rl.ac.uk/MRC_Psych_Db.html.

SAITO, S., & BADDELEY, A. D. (2004). Irrelevant sound disrupts speech production: Exploring the relationship between short-term memory and experimentally induced slips of the tongue. *Quarterly Journal of Experimental Psychology*, **57A**, 1309-1340.

SCHNEIDER, W., ESCHMAN, A., & ZUCCOLOTTO, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

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

UNSWORTH, N., HEITZ, R. P., SCHROCK, J. C., & ENGLE, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, **37**, 498-505.

(*Continued on next page*)

APPENDIX A
Words Used for Lag Task in Experiments 1–3

Author	Opinion	Chest	Diamond	Cereal	Plant
Sunburn	Church	Learn	Pedal	Device	Safety
Picture	Snail	Heart	Frost	Juice	Pickle
Glass	Miracle	Trouble	Apple	Cloak	Building
Cigar	Lesson	Sleigh	Hunger	Watch	Economy
Doctor	Fruit	Gallery	Blister	Robbery	Brush
Circus	Victory	Reason	Dreamer	Anxiety	Defrost
Library	Scarlet	Attend	Impulse	Officer	Globe
Lever	Light	Leopard	Recital	Madness	Rumble
Olive	Tragedy	Gravity	Purse	Level	Persist
Profile	Choir	Paper	Gender	Table	Shrimp
Valley	Anchor	Beach	Horse	Monkey	Breath
Painter	Collar	Colony	Sponge	Flower	Abdomen
Onion	Basket	Dinner	Badge	Queen	Freeze
Estate	Region	Nation	Vault	Patch	Choose
Tongue	Toast	Quail	Welfare	Goddess	Bribe
Number	Costume	Lemon	Dweller	Gravel	Supreme
Tobacco	Smoke	Creamer	Travel	Camera	Texture
Hearing	Cheese	Concert	Lumber	Disease	Graph
Ivory	Relic	Garden	Thought	Jacket	Squeak
Glove	Brick	Friend	Monarch	Magnet	Pepper
Candy	Circle	Ounce	Fruit	Camel	Money
Import	Velvet	Custom	Bread	Ability	Exhibit
Present	Boulder	Maple	Butter	Lunch	Liver
Fiddle	Journal	Forbid	Loyalty	Wizard	Balance
Belly	Fortune	Mileage	Charter	Settler	Pupil
Prairie	Hockey	Couch	Sodium	Woman	Testify
Build	Orange	Algebra	Helmet	Vision	Mixer
Justice	Artist	Dress	Slice	Cement	Berry
Uncle	Essence	Eagle	Floor	Volcano	Speaker
Violin	Sheep	Spinach	Dynasty	Shotgun	Thread
Cable	Prize	Resort	Umpire	Beard	Cottage
Loafer	Glacier	Sadness	Machine	Satin	Unique
Product	Tribute	Pencil	Gallon	Comrade	

APPENDIX B
Words Used for Lag Task in Experiment 4

Ache	Duke	Jump	Nerve	Rock	Sound	Vase
Arch	Dust	King	Nest	Roof	Spark	Vault
Aunt	Earth	Kiss	Note	Root	Spice	Vein
Back	Fair	Knee	Nurse	Rough	Spool	Verse
Badge	Fang	Knob	Ounce	Round	Staff	Vest
Band	Feast	Lake	Palm	Route	Stare	Voice
Bank	Fence	Lane	Paste	Rust	Stalk	Wage
Bass	Firm	Lark	Pearl	Sack	State	Waist
Beam	Fish	Latch	Peel	Sage	Steel	Wall
Beast	Fleet	Lawn	Pest	Sail	Stone	Wave
Birth	Flood	Lead	Pier	Saint	Stove	Weed
Blind	Flute	Leaf	Pine	Sauce	Suit	Wheat
Block	Foil	Lens	Plane	Scale	Surf	Wheel
Blush	Frame	Light	Plant	Scar	Swamp	Wife
Board	Fuse	Limb	Plug	Scene	Sweat	Wing
Bone	Gang	Lime	Point	Scout	Sword	Witch
Boot	Geese	Link	Pole	Seal	Tack	Wolf
Bowl	Germ	Lint	Pool	Seat	Tank	Youth
Brain	Glare	Load	Porch	Shape	Tape	
Bush	Goal	Lodge	Pork	Shear	Task	
Cage	Golf	Loop	Post	Shell	Taste	
Camp	Graph	Lord	Prey	Shoe	Team	
Card	Grass	Lung	Prize	Shore	Test	
Cash	Group	Lure	Prong	Show	Thief	
Chain	Guard	Mail	Pump	Shrub	Thorn	
Chest	Guide	Mane	Purse	Sign	Throw	
Child	Hail	March	Quake	Silk	Tide	
Clam	Harp	Mare	Quart	Sink	Toad	
Clasp	Heap	Mash	Queen	Skate	Toast	
Cloth	Heel	Mast	Race	Skull	Toll	
Coach	Herb	Match	Rack	Sleet	Tomb	
Coin	Hood	Maze	Raid	Slide	Tool	
Core	Hook	Meat	Rail	Slope	Town	
Craft	Hound	Mile	Ramp	Smack	Track	
Crown	Hunt	Mink	Range	Smell	Trail	
Crumb	Hymn	Mist	Rash	Smile	Train	
Dance	Inch	Mold	Reed	Smoke	Tribe	
Dawn	Isle	Moon	Reel	Snail	Troop	
Debt	Jail	Mouth	Rent	Soap	Trunk	
Disc	Jerk	Mule	Rhyme	Sock	Tube	
Dove	Judge	Name	Ridge	Soil	Tusk	
Dress	Juice	Neck	Roar	Sore	Twig	

(Manuscript received February 15, 2006;
revision accepted for publication May 8, 2006.)