Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale
2013, Vol. 67, No. 4, 260–270

1196-1961/13/$12.00 DOI: 10.1037/a0034351

Individual Differences in Working Memory, Secondary Memory, and Fluid Intelligence: Evidence From the Levels-of-Processing Span Task

Nathan S. Rose
The Rotman Research Institute of Baycrest, Toronto, ON, Canada

Individual differences in working memory (WM) are related to performance on secondary memory (SM), and fluid intelligence (gF) tests. However, the source of the relation remains unclear, in part because few studies have controlled for the nature of encoding; therefore, it is unclear whether individual variation is due to encoding, maintenance, or retrieval processes. In the current study, participants performed a WM task (the levels-of-processing span task; Rose, Myerson, Roediger III, & Hale, 2010) and a SM test that tested for both targets and the distracting processing words from the initial WM task. Deeper levels of processing at encoding did not benefit WM, but did benefit subsequent SM, although the amount of benefit was smaller for those with lower WM spans. This result suggests that, despite encoding cues that facilitate retrieval from SM, low spans may have engaged in shallower, maintenance-focused processing to maintain the words in WM. Low spans also recalled fewer targets, more distractors, and more extralist intrusions than high spans, although this was partially due to low spans' poorer recall of targets, which resulted in a greater number of opportunities to commit recall errors. Delayed recall of intrusions and commission of source errors (labeling targets as processing words and vice versa) were significant negative predictors of gF. These results suggest that the ability to use source information to recall relevant information and withhold recall of irrelevant information is a critical source of both individual variation in WM and the relation between WM, SM, and gF.

Keywords: working memory, secondary memory, fluid intelligence, levels of processing

Individual differences in working memory (WM) are related to a wide variety of higher-order cognitive abilities, such as fluid intelligence (gF) (see Unsworth & Engle, 2007, for a review). For this reason, a great deal of research has been conducted in an attempt to understand performance on WM tasks. However, debate surrounds the specific sources of individual variation in WM and the relation between WM and gF (e.g., Mogle, Lovett, Stawski, & Sliwinski, 2008; Shelton, Elliott, Matthews, Hill, & Gouvier, 2010; Unsworth, Brewer, & Spillers, 2009).

According to Unsworth and Engle's (2007) dual-component framework of WM, individual differences in *secondary memory* (SM) retrieval underlie individual differences in performance on WM tasks and are responsible, in part, for the relation between WM and gF. Unsworth and Engle (2007) proposed that WM tasks (e.g., complex span tasks such as operation span) predict gF because they require retrieving items from SM when they have been displaced from *primary memory* (PM). According to their framework, controlled retrieval is the key ability that is common to measures of WM, SM, and gF.

Working Memory and Secondary Memory

Several lines of research support the notion that individual differences in performance on WM tasks are driven by differences in the ability to retrieve information from SM. For example, individuals with higher scores on WM span tasks ("high spans") outperform individuals with lower scores on WM span tasks ("low spans") and on SM tasks as well (e.g., Unsworth, 2007). This is particularly true for SM tasks that emphasize recollection over familiarity, such as local versus global recognition (Oberauer, 2005) or source recognition versus item recognition (Unsworth & Brewer, 2009).

Most studies examining variation in both WM and SM tasks have focused on differences in retrieval processes as the key source of individual differences. However, it is important to also consider the way in which encoding processes impact the nature of subsequent retrieval processes. To overcome this limitation, the current study examined the impact of an encoding manipulation on individual differences in WM and SM.

If low spans use poor retrieval cues to retrieve items from SM, would their recall benefit if they encoded better retrieval cues? Some studies have shown that explicitly instructing or requiring participants to use semantic encoding strategies benefits WM over phonological or maintenance (repetition) encoding strategies (Bailey, Dunlosky, & Hertzog, 2009; Bailey, Dunlosky, & Kane, 2008; Loaiza, McCabe, Youngblood, Rose, & Myerson, 2011; Rose & Craik, 2012, Experiment 2), although some have not (e.g., Rose & Craik, 2012, Experiment 1; Rose, Myerson, Roediger, & Hale, 2010; Turley-Ames and Whitfield, 2003). It remains to be seen whether low spans' WM capacity might benefit from deeper levels of processing (LOP) at encoding. Addressing this issue was a primary goal of the current study.

I thank Joel Myerson, Roddy Roediger, Sandy Hale, Jill Shelton, Gus Craik, and Klaus Oberauer for their comments on an earlier version of this article and Linden Weiswerda and Molly Ruben for their assistance with data collection and scoring.

Correspondence concerning this article should be addressed to Nathan S. Rose, The Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, ON, M6A 2E1. E-mail: nrose@research.baycrest.org

Individual Differences in Recall Errors

Unsworth and Engle (2007) also hypothesized that low WM span individuals perform worse than high WM spans because they are less effective at using source or contextual cues when retrieving information from SM. Unsworth and Engle suggested that low spans (relative to high spans) use noisier source or context cues for retrieval, which produces a memory search set that contains less relevant and more irrelevant information. Some findings are consistent with these hypotheses. For example, low span individuals recall more off-target, irrelevant items (intrusions) and have more source errors (transpositions) on WM span tasks (Unsworth & Engle, 2006a), and older adults, who have lower WM capacity than younger adults on average, recall more intrusions on the reading span task, particularly when the trials of the task were ordered in a way to induce proactive interference (Lustig, May, & Hasher, 2001). De Beni and colleagues found that, relative to young-old adults (55–65 years of age), old-old adults (over 75 years of age) had worse recall and more intrusions on a memory updating task (De Beni & Palladino, 2004) and on the listening span task (Palladino & De Beni, 1999).

However, these studies did not consider the relation between recall of intrusions and veridical recall. Lower scoring individuals may recall more intrusions than higher scoring individuals in part because they have more opportunities to commit recall errors due to their poorer level of veridical recall. For example, consider a task with 10 lists of six to-be-remembered words per list (60 total words). If one subject only recalls 2 of 6 words per list correctly, then there are many opportunities to recall an intrusion (i.e., 40 total) because the correct words were not recalled in these instances. If another subject recalls 5 of the 6 words per list correctly, then there are only 10 opportunities to recall an intrusion. Suppose the first subject recalls four intrusions and the second subject recalls three intrusions. Even though the first subject recalled one more intrusion than the second subject, the first subject recalled an intrusion 10% of the times they could have recalled an intrusion (4/40 = .10), whereas the second subject did so 30% of the time (3/10 = .30). Of course, this assumes that subjects know how many items are to be recalled on each trial, which was the case in the current study, and subjects would not recall more than the total number of to-be-recalled items, which may not be true in all cases.

Although investigation of recall errors can be informative, it is important to consider the way in which accurate recall and recall errors can interact. Lower scoring individuals produce more errors (including omissions) than higher scoring individuals by definition, and so they have a greater chance of producing intrusions simply because they recall fewer targets. Thus, it is unclear if the reason low spans recall more intrusions is because they have more opportunities to recall an error. Addressing this issue was a second goal of the current study.

Relation Between gF and WM, SM, Intrusion Errors, and Source Memory

Unsworth and Engle (2007) also suggested that the ability to use source or context cues to constrain retrieval and monitor retrieval output may be the ability that underlies the relation among measures of WM, SM, and gF. Previous research has shown a relation between gF and measures of intrusions and source monitoring derived from different WM and SM tasks (e.g., Unsworth, 2009). However, it is unclear whether the key source of the relation among WM, SM, and gF is the ability to use source cues to constrain retrieval or monitor the products of retrieval—that is, to withhold report of irrelevant items. If the ability to withhold recall of irrelevant information via source-constrained retrieval processes is a critical component of individual variation in WM, SM, and gF, do individual differences in intrusions and source errors on WM and SM tests predict individual differences in gF even after controlling for differences in the opportunity to produce errors? Addressing this issue was a third goal of the current study.

The Current Study

The present study used a novel design to examine individual differences in WM, SM, and gF by manipulating LOP at encoding and assessing participants' abilities to access both target items and items that were previously to be withheld from recall on WM and SM tests and to provide source judgments for retrieved items. This was accomplished by examining immediate and delayed recall of relevant and irrelevant words from the LOP span task (Rose et al., 2010). In the LOP span task, participants are required to recall a series of words (in uppercase) but, in between the presentation of each to-be-remembered "target" word, two "processing" words (in lowercase) are given, and the participant must choose which processing word matches the target, when "match" is specified as either visual features (color), phonology (rhyme), or meaning (semantic). An example of a four-item list is: BRIDE (dried, groom), LEG (arm, beg), STONE (grown, rock), CHEF (cook, deaf). To accurately recall the list, the participant must recall all of the target words ("bride, leg, stone, chef") and none of the processing words. In the present study, participants performed the phonological and semantic processing conditions of the LOP span task as well as a word span task for the purpose of comparing individual differences on the complex LOP span task to a simple span task (Unsworth & Engle, 2006b). Then, after performing the immediate recall tests for all lists, participants took a final free recall test after a 10-min delay. For this final free recall test, participants were instructed to recall both the target words from the span tasks, as well as the related processing words from the LOP span task that were initially to be withheld from recall. In addition, participants were instructed to provide source judgments as to whether each recalled item had initially served as a target or a processing word.

Thus, in this experiment, participants had to process distractor words that were highly related to the to-be-remembered target words, to withhold recall of these items on the initial test, and then recall them on the delayed test and indicate the source or context in which the items were initially experienced. Therefore, the LOP span task and the subsequent delayed recall test are well suited for addressing issues concerning individual differences in controlled retrieval processes because the tasks required a high degree of controlled search and retrieval monitoring (Unsworth & Engle, 2007). In addition to examining individual differences in various measures of WM and SM, how these measures predict performance on a gF test was also examined.

A variety of effects might be expected for the current task conditions. One possibility is that low spans will generally benefit from deeper LOP at encoding more than high spans. An analogous pattern may be seen in age differences in LOP effects. For example, older adults often obtain larger benefits of deeper processing at encoding, likely because younger adults spontaneously engage in more elaborative encoding than older adults (see Craik & Rose, 2012, for a review). Alternatively, high spans may benefit more from deeper LOP than low spans because, even when low spans are encouraged to engage in deeper processing at encoding, they may resort to shallower, maintenance-focused processing for immediate recall. Some findings suggest the latter possibility may be the case. For example, Bailey et al. (2008) found that low spans reported using less effective strategies than high spans.

With regards to recall errors, it is likely that low spans will have more recall errors (e.g., intrusions) than high spans, but it is unclear whether this difference will go away after statistically controlling for differences in veridical recall. If it does, this would suggest that the primary reason that low spans commit more recall errors is because they do not recall as many target items, and so intrusions are emitted instead. If the difference remains, however, it would suggest that low spans are indeed deficient in controlling the retrieval of targets and withholding recall of irrelevant items. Such a deficit is hypothesized to be due to differences in the ability to use source information to control retrieval output.

With regards to the hypothesis that source memory is a key cause of the WM–gF relationship, if this is true, then individual differences in source memory should predict variation in gF and may mediate the relation between WM, recall of intrusions, and gF. Accordingly, the nature of encoding was experimentally manipulated in the current study and individual differences in WM, SM, and gF were examined using multiple regression/correlational analyses.

Method

Participants

Fifty Washington University undergraduate students participated in the experiment in exchange for course credit. All participants were native English speakers. Data for one participant were removed because the person did not follow the instructions for the WM tasks (i.e., the person wrote to-be-recalled items down as they were presented). Data were unavailable for the matrix reasoning test for three participants due to computer error, and one participant was dropped as an outlier (data greater than 3 SDs below the mean).

Design and Procedure

The design was a repeated-measures design with span task (Word, Phonological LOP, Semantic LOP), list length (four items, eight items), and time of test (Immediate, Delayed) as within-subject factors. Each experimental session began with the participants performing the WM tasks, after which they solved mental arithmetic problems. Then, they were given a surprise free recall test on words from the WM tasks. Finally, the participants took a gF test. Details concerning the procedures for all of the tests are provided below. Following practice trials to familiarize participants with the various conditions of the WM tasks, participants performed a simple word span task as well as the phonological and semantic processing conditions of the LOP span task (Rose et al., 2010). The three types of test trials were mixed (rather than blocked, as in Rose et al., 2010), and the type of task to be performed on each trial was indicated on the screen immediately before the trial began (i.e., word span task: target words only; phonological LOP span task: target words plus rhyme decisions; or semantic LOP span task: target words plus meaning decisions). In addition, each type of trial could involve presentation of either a four-item series or an eight-item series; the order of these series lengths was pseudorandom, as was the order of the three types of trials, so that all of the participants experienced the various trial types and series lengths in the same order. There were three trials of each type.

On each trial of the word span task, participants viewed a series of to-be-remembered target words on a computer screen. Each target word was presented in uppercase letters in the center of the screen for 1750 ms, followed by a blank screen for 250 ms. At the end of each series, participants saw a recall signal (a green box). They were instructed to try and recall the words in the order presented and to report what they recalled by writing on an answer sheet provided by the experimenter.

The LOP span task trials were similar to the word span trials except that, following presentation of each to-be-remembered target word, two nontarget "processing" words were presented side by side in lowercase letters until the participant made a selection. One of the processing words rhymed with the target word and the other was a semantic associate of the target word. Half of the trials on the LOP span task were from the phonological condition, and on these trials participants were instructed to indicate which processing word (either the word on the left or the word on the right) rhymed with the target word by pressing the corresponding key (labeled left or right) on the keyboard. The other half of the trials on the LOP span task were from the semantic condition, and on these trials participants were instructed to indicate which processing word was related in meaning to the target word. The stimuli were the same as in Rose et al. (2010), Experiment 1. Target words were monosyllabic, high-frequency nouns, and imageability was matched across conditions.

On each trial of the LOP span task, presentation of target words and processing words continued until the complete series of either four or eight target words and their corresponding four or eight pairs of processing words were presented. Participants were instructed that when they saw the recall signal, they should write down the target words in the order in which the words had been presented. Participants were specifically told not to write down any of the related processing words.

After completing all trials for the span tasks, participants performed mental arithmetic problems for 10 min. Following the arithmetic filled delay, participants performed a final free recall test. For this delayed recall test, participants were told to recall and write down on the sheet of paper as many words as they could from the experiment—both target and processing words. Then, after recalling the words, participants were required to make a source judgment for each recalled word. They were to indicate whether each word was a target or processing word by writing either a T or a P next to each word. Participants completed the recall test and the source judgments within 5 min. After completing the delayed recall test, participants completed the gF test (the matrix reasoning subtest of the WAIS-III, Wechsler, 1997). For both the immediate and delayed recall tests, the number of targets and processing words recalled, and the number of intrusions and source errors were recorded. Intrusion errors consisted of words that were not to be recalled on the current test. For the immediate recall tests, these errors consisted of prior list intrusions (recalling an item on the current list that was to-be-recalled on a previous list) and extra list intrusions (recalling items that were not presented in the experiment). For the delayed recall test, extra list intrusions and source errors were considered. Source errors consisted of both targets labeled as processing words and processing words labeled as targets.

Data Analysis

Correlation and regression analyses were conducted on the full sample. To characterize differences between high- and low-scoring individuals in measures of veridical recall, false recall (intrusions), and source errors, the 20 participants who had the highest overall average span task performance (averaged across the three span tasks) were classified as high-span individuals and the lowest scoring 20 participants were classified as low spans. The split group analyses simply illustrate the direction of associations revealed by the correlation/regression analyses. However, because of the dangers and inefficiencies of extreme group designs, statistical inferences were only drawn from the correlation/regression analyses.

The delayed recall data for targets were conditionalized on initial level of recall in order to control for baseline differences in initial level of recall between the simple and complex span tasks and between the low- and high-span groups. Therefore, differences observed in delayed recall of targets were not simply the result of differences in initial level of recall.

Results

In keeping with the three goals of the current study (to examine individual differences in WM and SM, individual differences in recall errors, and relations among WM, SM, and gF), the results are organized in three sections. Of primary interest was whether, as has been hypothesized, individuals with lower and higher WM scores differed in the amount of benefit from encoding deeper cues on recall on WM and SM tests. To test this hypothesis the relation between WM and the LOP effect—the amount of benefit of deep versus shallow cues—was examined with regression analyses.

Individual Differences in LOP Effects on WM and SM Tests

First, it should be noted that performance on the processing component of the LOP span task (rhyme and semantic judgments) was high in both the phonological (93%) and semantic (94%) conditions, and neither processing accuracy nor reaction time (RT) (phonological = 1483 ms, semantic = 1499 ms) differed between the two LOP conditions or between the high- and low-span groups, *t*s <1. Thus, the nature of initial encoding was similar between high and low spans; however, as can be seen below, recall was substantially different between the groups.

Table 1 reports the mean proportion of words recalled on the three span tasks and the mean proportion of those words that were subsequently recalled on the delayed test for the full sample and for the high- and low-performing individuals. Correlations among all measures for the full sample may be found in the Appendix.

Figures 1 and 2 illustrate the differences in LOP effects on the WM and SM tests between low and high spans. For immediate recall, high spans recalled more words than low spans overall, with the exception of words recalled on the word span task for short lists (see the top half of Table 1). Additionally, the proportion of

Table 1

Proportion of Words From the Span Tasks Recalled on the Immediate Tests and the Delayed Test (Conditionalized on Initial Recall) for the Whole Sample and the Low- and High-Span Participants

Time of test by list length and group	Word span	Phonological LOP	Semantic LOP	Mean
Immediate recall				
Four-item lists				
Whole sample	.97 (.01)	.86 (.02)	.88 (.02)	.91 (.01)
Low spans	.95 (.02)	.75 (.03)	.76 (.03)	.82 (.02)
High spans	.98 (.01)	.93 (.02)	.96 (.01)	.96 (.01)
Eight-item lists				
Whole sample	.73 (.02)	.58 (.03)	.55 (.02)	.62 (.02)
Low spans	.65 (.03)	.44 (.03)	.41 (.02)	.50 (.02)
High spans	.82 (.03)	.75 (.03)	.70 (.03)	.76 (.02)
Mean				
Whole sample	.85 (.01)	.72 (.02)	.72 (.02)	
Delayed recall				
Four-item lists				
Whole sample	.08 (.02)	.08 (.02)	.19 (.02)	.12 (.01)
Low spans	.05 (.02)	.12 (.03)	.15 (.03)	.10 (.02)
High spans	.12 (.04)	.07 (.02)	.23 (.04)	.14 (.02)
Eight-item lists				
Whole sample	.19 (.02)	.27 (.03)	.32 (.03)	.26 (.02)
Low spans	.14 (.03)	.26 (.04)	.28 (.05)	.23 (.03)
High spans	.26 (.04)	.29 (.05)	.38 (.04)	.30 (.03)
Mean				
Whole sample	.14 (.02)	.18 (.02)	.25 (.02)	

Note. Values are means with standard error of the mean in parentheses. Slight differences between the marginal means and means computed from the cell means are due to rounding. LOP = levels of processing.

words recalled from the semantic condition was not different from the phonological condition for either high spans or low spans. That is, deeper LOP did not benefit immediate recall for either group.

For delayed recall, high spans recalled more words than low spans for the word span task and the semantic condition of the LOP span task, but low and high spans did not differ in recall of words from the phonological condition of the LOP span task. It may also be noted that, although high spans' delayed recall demonstrated an LOP effect—a benefit of deep (semantic) processing over shallow (phonological) processing–low spans' recall did not.

These observations were confirmed with linear regression analyses predicting the LOP effect from WM scores (see Table 2). The correlation between WM and the LOP effect on immediate recall was zero, but there was a significant positive correlation on delayed recall ($r = +.37$, $p < .01$, see Figure 2). Thus, even though the processing judgments of the LOP span task required that both low and high spans encoded the words in terms of phonological or semantic characteristics for the LOP span task and even though analysis was restricted to those words that were initially recalled in order to control for baseline differences in immediate recall, high spans' delayed recall was better than low spans, specifically for words from the word span and semantic condition of the LOP span. Low spans did not differ from high spans in delayed recall of words from the phonological condition of the LOP span task. This pattern suggests that, despite having encoded deep retrieval cues in the semantic condition of the LOP span task, low-span participants may have focused on shallower phonological/articulatory cues to maintain the words for immediate recall. This possibility is considered further in the Discussion section.

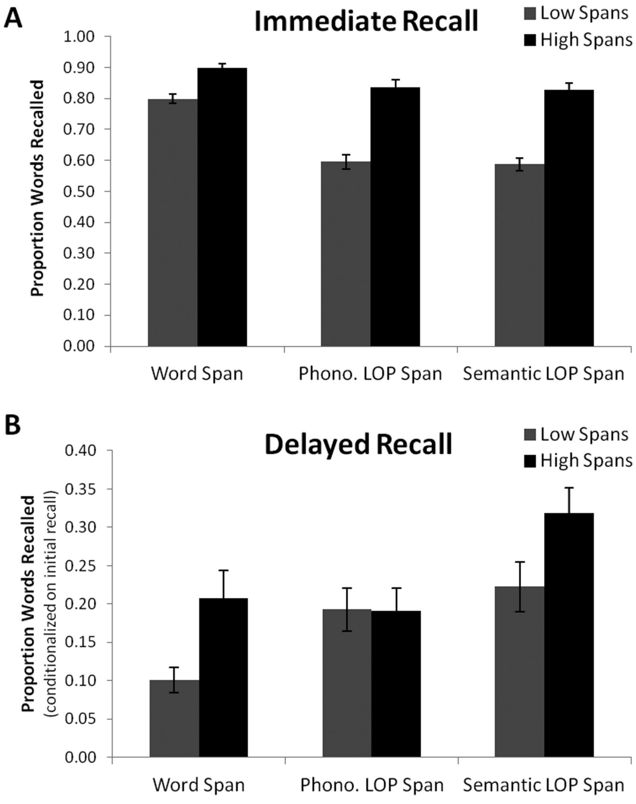

Figure 1. (A) Mean proportion of words recalled on the immediate recall tests of the word span task and the phonological and semantic conditions of the levels-of-processing span task, and (B) the mean proportion of those words that were subsequently recalled on the delayed recall test for low and high spans. Error bars represent the standard error of the mean.

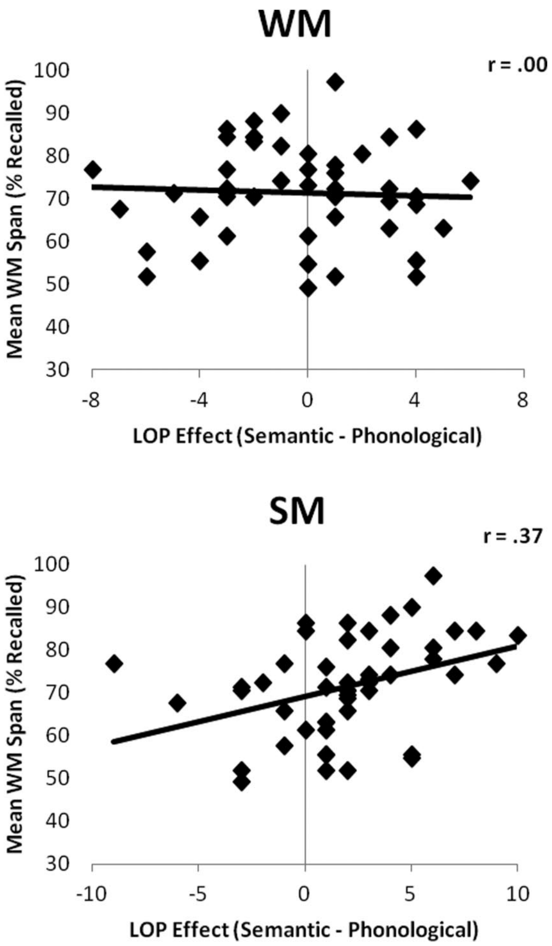

Figure 2. Scatterplots showing the LOP effect (number of words recalled from the semantic minus the phonological condition) as a function of mean WM span (percent words recalled) for the whole sample ($N = 50$) on the immediate (WM) and delayed (SM) recall tests.

Individual Differences in Recall of Processing Words, Intrusions and Source Errors

Next, individual differences in the number of different types of errors (intrusions and source errors[1]) and the number of processing words recalled on the immediate tests (when they were to be withheld from recall) and on the delayed test (when they were

---

[1] The number of source errors made by one high-span participant (14) was greater than 4.5 SDs from the mean of the whole sample. Based on a binomial distribution function, the probability of committing 14 source errors out of 29 words recalled, when the probability of a source error is .25 (the highest proportion of source errors committed by any other high-span participant), is $p < .005$. Therefore, this outlier was removed from analysis. No other outliers were observed.

Table 2
Summary of Results From Linear Regression Analyses Predicting the LOP Effect on Immediate and Delayed Recall From Mean Percent WM Span for the Whole Sample

Dependent variable	R^2	β	*t*	*p*
LOP effect (sem. − phono.)				
Immediate recall	.00	−.05	−0.38	.76
Delayed recall	.14	.37	2.78	<.01

Note. LOP effect = number of words recalled from the semantic minus the phonological encoding condition; WM = working memory; sem. = semantic; phono. = phonological.

to-be-recalled items) were considered. For the immediate tests, there were three different types of intrusions: the distracting processing words that were to-be-withheld from recall (processing words), words that were to be remembered on a previous list (prior-list intrusions), and nonpresented words (extralist intrusions). On the delayed recall test, recall of processing words and recall errors were of interest. There were two types of recall errors (besides omissions): recall of nonpresented words (extralist intrusions) and source errors (i.e., either labeling target words that were recalled as processing words or labeling processing words as target words). The data showing the differences between low and high spans are presented in Table 3.

On the immediate recall test, low spans produced more of all three types of error than high spans. However, as discussed in the Introduction, because recall errors might interact with the level of accurate recall, it is important to consider the number of recall errors as a proportion of the number of opportunities to commit a recall error (i.e., the number of times a target was not recalled). When these measures were calculated, a different pattern was found: The number of processing words, prior list intrusions, or extra list intrusions recalled on the LOP span task as a proportion of the total number of recall errors was not greater for low spans than high spans (see Table 3).

On the delayed recall test, low and high spans did not differ in the number or proportion of processing to targets recalled. However, low spans committed a greater number of both intrusions and source errors than high spans. The same pattern was found when the number of intrusions was calculated as a proportion of the number of recall errors, and the number of source errors was calculated as a proportion of the number of items correctly recalled (see Table 3).

These observations were confirmed with linear regressions predicting each variable from WM span for the whole sample. The proportion of variance accounted for (R^2), the standardized beta weights (β), and *t* and *p* values for two-tailed *t* tests of whether the beta weights significantly differed from zero are presented in Table 3.

Taken together, these findings indicate that low spans recalled fewer targets and more intrusions than high spans, but, on the immediate recall (WM) tests, when differences in the number of opportunities to commit a recall error were taken into account, the proportions were similar for low and high spans. On the delayed recall (SM) test, low spans made proportionally more recall errors (intrusions and source errors) than high spans. This suggests that the ability to constrain retrieval from SM, perhaps by relying on source information to recall targets and withhold recall of intrusions, may be an important component of differences between low

Table 3
Number of Processing Words and Intrusions Recalled on the Immediate and Delayed Recall Tests, and Source Errors on the Delayed Test for Low and High Spans, and Results From Regression Analyses Predicting Each Variable From Mean WM Span for the Whole Sample

	Low span		High span		Whole sample			
Variable	Mean	*SEM*	Mean	*SEM*	R^2	β	*t*	*p*
Immediate recall								
Processing words (nontargets)								
Number	4.7	0.5	1.9	0.3	.24	−.49	−5.21	<.001
Proportion	.11	.01	.11	.02	.00	.05	0.33	.72
Prior list intrusions								
Number	1.0	0.3	0.5	0.2	.11	−.33	−2.40	<.05
Proportion	.02	.01	.03	.02	.02	.15	1.02	.31
Extra list intrusions								
Number	4.1	0.7	2.5	0.4	.14	−.37	−2.75	<.01
Proportion	.09	.06	.15	.15	.09	.31	2.21	<.05
Delayed recall								
Processing words (targets)								
Number	6.4	0.9	7.7	1.3	.01	.08	0.51	.61
Proportion	.31	.04	.25	.03	.05	−.23	−1.58	.12
Intrusions								
Number	6.9	1.1	3.9	0.6	.10	−.32	−2.31	<.05
Proportion	.023	.004	.013	.002	.17	−.41	−3.10	<.01
Source errors								
Number	2.9	0.5	1.7	0.4	.10	−.32	−2.32	<.05
Proportion	.204	.046	.079	.019	.16	−.41	−3.00	<.01

Note. Values are means with standard error of the mean. On the immediate recall tests, processing words were nontargets (i.e., items that were to-be-withheld from recall), but on the delayed test, processing words were targets (i.e., to-be-recalled items).

and high spans. The next section suggests that this ability may also be a critical source of the relation among measures of WM, SM, and gF.

Relations among Measures of WM, SM, and gF

First, the utility of veridical recall and the number of intrusions and source errors in predicting performance on a measure of gF (matrix reasoning) was examined. The results of a simultaneous regression analysis on the full sample are presented in Table 4. As may be seen in Table 4, recall of targets and processing words did not predict matrix reasoning. Only the number of intrusions and source errors on the delayed recall test were significant predictors of matrix reasoning, indicating that participants that recalled more intrusions or had more source errors had lower scores on the gF test.

According to Unsworth and Engle (2007), the ability to withhold recall of intrusions may be achieved by relying on source cues to control retrieval, and this ability is thought to be the source of the relation among measures of WM, SM, and gF. To test this hypothesis, a stepwise regression analysis was conducted to see if source memory (as measured by the number of source errors committed on the delayed recall test) mediated the association between the number of delayed recall intrusions and gF. The results are presented in the top half of Table 5 (Model 1). The number of intrusions recalled remained a significant predictor of matrix reasoning after controlling for the number of source errors committed, accounting for an additional 12% of unique variance in matrix reasoning.

However, as noted above, it is important to control for differences in the number of opportunities to recall intrusions and make source errors. A second stepwise regression model was computed to see if the number of intrusions recalled as a proportion of veridical recall was associated with matrix reasoning performance and whether the number of source errors committed as a proportion of opportunities (number of items not veridically recalled) mediated the association. The results are presented in the bottom half of Table 5 (Model 2). Controlling for the number of opportunities to commit a source error reduced the association with matrix reasoning ($\beta = -.27$, $p = .07$). Nevertheless, the number of delayed recall intrusions as a proportion of the number of opportunities to recall an intrusion remained a significant predictor of matrix reasoning, accounting for 30% of the variance. Taken together these findings suggest that the ability to withhold recall of irrelevant words (intrusions) on the SM test captured an ability associated with performance on the gF test over and above the ability to recall source information about the words.

Table 4
Results of the Simultaneous Regression Analysis Predicting Matrix Reasoning From the Immediate and Delayed Recall Variables

Variable	r	β	t	p
Immed. word span	.19	−.05	−.30	.77
Immed. phono. LOP span	.14	.10	.43	.67
Immed. seman. LOP span	.25	−.23	−.99	.33
Immed. processing words	.03	.17	1.15	.26
Immed. PLIs	−.14	.02	.14	.89
Immed. ELIs	−.32	−.24	−1.39	.17
Delay word span	.19	.15	.88	.38
Delay phono. LOP span	−.09	−.27	−1.58	.12
Delay seman. LOP span	−.01	−.08	−.34	.74
Delay processing words	−.12	.13	.72	.48
Delay intrusions	**−.51**	**−.37**	**−2.40**	**.02**
Delay source errors	**−.45**	**−.34**	**−2.04**	**.05**

Note. $R^2 = .465$, $F(12,32) = 2.36$, $p = .03$. Immed. = immediate; phono. = phonological; seman. = semantic; LOP = levels of processing; PLIs = prior list intrusions; ELIs = extra list intrusion. Boldface indicates significant predictors of matrix reasoning.

Table 5
Results of the Stepwise Regression Analysis Predicting Matrix Reasoning From the Numbers of Source Errors and Intrusions (Model 1) or Errors as Proportions of the Number of Opportunities (Model 2) From the Delayed Recall Test

Model and step	β	R^2	F	p
Model 1				
1. Number of source errors	**−0.45**	**0.21**	**11.12**	**0.002**
2. Number of intrusions	**−0.38**	**0.33**	**7.41**	**0.009**
Model 2				
1. Proportion of source errors	−0.27	0.07	3.37	0.07
2. Proportion of intrusions	**−0.52**	**0.30**	**13.57**	**0.001**

Note. Boldface indicates significant predictors of matrix reasoning.

Discussion

The present study was conducted to examine sources of individual variation in WM and SM and their relation to gF. Specifically, the present study attempted to address (1) whether low spans would benefit from encoding cues that might facilitate retrieval from SM, (2) whether low spans recall more irrelevant information than high spans on WM and SM tasks even when controlling for differences in veridical recall, and (3) whether intrusions and source monitoring errors predict individual differences in gF even when controlling for the number of opportunities to produce such errors.

Variation in WM and Retrieval From SM

To address the hypothesis that variation in WM depends on retrieval from SM, the present study questioned whether low spans benefit from encoding cues that facilitate retrieval from SM. Deeper (semantic) processing at encoding did not benefit immediate recall (WM) relative to shallower (phonological) processing for either low or high spans. Encoding semantic retrieval cues only benefited delayed recall (SM). Although the structure of the LOP span task is similar to complex WM span tasks which, according to Unsworth and Engle (2007), measure SM, these results suggest the LOP span task was not a pure measure of SM.

The pattern replicates the findings of Rose et al. (2010), but extends them using different methods (e.g., mixed rather than blocked design, delayed recall rather than recognition) and shows that neither low- nor high-spans' WM benefited from deeper processing at encoding on the LOP span task. The reason the LOP span task may not have been a pure measure of retrieval from SM is because, when performing WM tests, participants try to covertly retrieve or "refresh" the to-be-recalled items between the item presentation and processing phases of the task (Barrouillet, Bernardin, & Camos, 2004; McCabe, 2008). Because participants were likely able to periodically refresh to-be-recalled items in the current conditions, focusing on shallow (phonological, articulatory) cues would be sufficient for recalling them on an immediate test. Recently, we (Loaiza et al., 2011) showed that deeper LOP did benefit immediate recall on more traditional WM span tasks (reading span, operation span), perhaps because of their longer, more difficult processing operations (reading and verifying the accuracy of sentences or math equations). It was suggested that traditional WM span tasks are a better measure of SM because their difficult processing requirements impair the ability to covertly retrieve/refresh the to-be-remembered items between processing operations. Consistent with this idea, Rose and Craik (2012) showed that the extent to which LOP effects appear on WM tests largely depends on the amount of disruption to this covert-retrieval/refreshing maintenance process.

Delayed recall of words initially to be recalled on the LOP span task—a purer measure of SM—revealed an interesting difference between low and high spans. It was somewhat surprising that deeper LOP at encoding did not benefit low-spans' delayed recall relative to phonological LOP. One way to account for this pattern is to suggest that, even though the orienting question on semantic LOP trials guided participants to focus on semantic features during initial encoding, low spans may have preferentially focused on shallower (e.g., articulatory or phonological) features of the words while trying to maintain them for immediate recall. Doing so may have negated the large and reliable benefit of deeper processing that is typically seen on delayed recall (Craik & Tulving, 1975). If it was unnecessary to initially maintain the words for immediate recall, it is likely that all participants would have benefited from deeper LOP at encoding on the delayed recall test. For example, it was recently shown that both low and high WM participants demonstrated a benefit of semantic over phonological LOP on a delayed recall test (Unsworth, Brewer, & Spillers, 2011). This suggests that the act of maintaining the words for the initial WM tests may have differentially affected subsequent recall on the SM test for low and high spans, which may reflect a difference in the type of maintenance processes that were initially used for immediate recall between high and low spans.

Although this account is tentative, several lines of research have suggested that low spans spontaneously engage in more repetition-based maintenance processes than high spans on WM tasks (Bailey et al., 2008; Turley-Ames and Whitfield, 2003; Unsworth & Spillers, 2010). For example, low spans report using less elaborative maintenance strategies than high spans and strategy use partially mediates the relation between WM and gF (Bailey et al., 2008; Unsworth & Spillers, 2010). Future research should consider individual differences in the use of maintenance mechanisms as a potential source of individual differences in WM.

Individual Differences in Source Constrained Retrieval and Output Monitoring

The LOP span task and the delayed recall test of the current study provided a novel way to test hypotheses about the role of source constrained retrieval and output monitoring as an underlying source of individual differences in WM, SM, and gF. Because the LOP span task required processing both phonological and semantic representations that were highly related to the to-be-remembered words, yet were to be withheld from recall, source monitoring processes were likely needed to correctly discriminate between to-be-remembered items and other, associatively related items (e.g., processing words). Then on the surprise delayed recall test, because of the instructions to recall both target and processing words, the memory search set must have included words that were previously to be remembered as well as words that were previously irrelevant and distracting (i.e., processing words). Moreover, because participants were also required to make source judgments on the words recalled on this surprise test, the conditions placed heavy demands on source constrained retrieval and output monitoring processes.

Although previous studies have suggested that low spans recall more intrusions than high spans because low spans have more irrelevant, nontarget information in their memory search set (Unsworth & Engle, 2007), low spans also recall less target information, thereby giving them greater opportunity to produce intrusions. Nonetheless, the current study showed that even when differences in the level of veridical recall were taken into account, low spans still recalled more intrusions, although this was only true for the delayed recall test—the test that maximally implicated cue-driven search and retrieval from SM. This result represents an important replication and extension of prior research on individual differences in recall errors.

In addition to recalling more intrusions, low spans also committed a greater number of source errors on the delayed recall test. This finding is consistent with the hypothesis that an important source of variation between low and high spans may be the ability to retrieve relevant information from SM (Unsworth & Engle, 2007), but, more specifically, the present results suggest that monitoring the products of retrieval and withholding recall of irrelevant items is a critical retrieval process. The reason source memory ability may underlie individual differences in WM may not only be because accurate source memory helps to constrain retrieval to target items, but also because source information may be used to withhold recall of off-target, intruding items.

Recently, Unsworth and Brewer (2010b) reported data that also suggests the critical difference between high and low spans is in the ability to monitor the products of retrieval. They had high and low spans perform a standard free-recall task and an externalized free-recall task in which participants were instructed to press a key for each item that they knew was an intrusion. On the standard free-recall task, low spans recalled fewer targets and more intrusions (both prior-list and extralist). On the externalized free-recall task, high spans recalled as many extralist intrusions as low spans, but high spans were better at identifying recalled items that were intrusions. These findings converge with the present findings and suggest that high spans are better at withholding intrusions from output because they can identify the source of recalled items. Importantly, the present results provide novel support for this hypothesis because the nature of initial encoding was controlled and the measures of intrusion and source errors controlled for differences in the number of opportunities to commit such errors.

Relations Among WM, SM, and gF, and the Ability to Control Retrieval and Monitor Retrieval Output

The present findings showed that the number of recall intrusions and source errors on the delayed recall test provided the strongest associations with gF. The findings from this novel paradigm, which involves delayed recall and source memory of items initially from a WM task, converge with some recent findings, pointing to a relation between individual differences in WM and performance on other tasks measuring SM or source memory. For example, Unsworth and colleagues have shown that the number of intrusion errors on a variety of recall (SM) tasks were related and formed a single intrusion factor, and that this factor was negatively related to individual differences in both WM and gF (Unsworth, 2009; Unsworth & Brewer, 2010a). In addition, Unsworth and Brewer (2009) found support for a dual-process model in which the variance shared by item recognition, source recognition, and recall formed one factor (recollection) and the independent item recognition variance formed another factor (familiarity). They showed that recollection was the main source of the relation between WM and gF (Unsworth & Brewer, 2009), and a general source-monitoring ability (measured by two source recognition tasks) mediated the correlation between false recall and WM (Unsworth & Brewer, 2010a).

In the present study, the ability to withhold recall of irrelevant words (intrusions) on the delayed recall test captured an ability that was associated with performance on the gF test over and above the ability to recall source information about the words. Moreover, this was true regardless of whether individual differences in the number of opportunities to make an error were controlled. Although controlling for differences in veridical recall reduced the association between source memory and gF, the association between delayed recall intrusions and gF remained. Taken together, these findings suggest that individual differences in the efficiency of a specific retrieval process—monitoring retrieval from SM and suppressing irrelevant information—may be the key source of individual variation in WM, SM, and gF abilities. Future studies should focus on individual differences in this ability as the potential source of the association among WM, SM, and gF. A variety of encoding and retrieval conditions could be manipulated with the LOP span task to do so.

Résumé

Les différences individuelles quant à la mémoire de travail (MT) sont reliées au rendement à des tests de la mémoire secondaire (MS) et d'intelligence fluide (Gf). Toutefois, la source de la relation demeure peu claire, en partie parce que peu d'études ont contrôlé la nature de l'encodage. Ainsi, il n'est pas clair si les variations individuelles sont attribuables aux processus d'encodage, de maintien ou de récupération (rappel). Dans la présente étude, on a demandé aux participants d'effectuer une tâche de MT (tâche de niveaux de traitement; Rose, Myerson, Roediger III & Hale, 2010), ainsi qu'un test de MS, qui évaluaient à la fois les cibles et les mots de traitement distrayants de la tâche initiale de MT. Un traitement plus approfondi de l'encodage n'améliorait pas la MT, mais ultérieurement la MS, toutefois dans une moindre mesure pour les personnes ayant un empan inférieur de MT. Ce résultat suggère que, en dépit des indices d'encodage facilitant la récupération dans la MS, les sujets aux empans inférieurs auraient pu s'adonner à un traitement moins profond axé sur le maintien en vue de maintenir les mots dans la MT. Les sujets aux empans inférieurs se rappelaient d'un moindre nombre de cibles, d'un plus grand nombre de distracteurs et d'intrusions extra-liste que ceux aux empans supérieurs, mais ce résultat peut être en partie attribuable à leur rappel moins efficace des cibles, haussant ainsi le nombre d'occasions de commettre des erreurs de rappel. Le retard dans le rappel des intrusions et les erreurs relatives aux sources (décrire des cibles comme des mots de traitement, et vice versa) étaient des facteurs prédicteurs négatifs d'importance de Gf. Ces résultats suggèrent que la capacité d'utiliser des sources d'information pour rappeler des renseignements pertinents et d'éviter le rappel d'information non pertinente est une source déterminante des différences individuelles en matière de MT et dans la relation entre la MT, la MS et la Gf.

Mots-clés : mémoire de travail, mémoire secondaire, intelligence fluide, niveaux de traitement

References

Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working memory performance? *Psychology and Aging, 24,* 82–92. doi:10.1037/a0014078

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy-affordance hypothesis. *Memory & Cognition, 36,* 1383–1390. doi:10.3758/MC.36.8.1383

Barrouillet, P., Bernardin, S., & Camos, C. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133,* 83–100. doi:10.1037/0096-3445.133.1.83

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. doi:10.1037/0096-3445.104.3.268

De Beni, R., & Palladino, P. (2004). Decline in working memory updating through ageing: Intrusion error analyses. *Memory, 12,* 75–89. doi:10.1080/09658210244000568

Loaiza, V. M., McCabe, D. P., Youngblood, J. L., Rose, N. S., & Myerson, J. (2011). The influence of levels of processing on recall from working memory and delayed recall tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1258–1263. doi:10.1037/a0023923

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130,* 199–207. doi:10.1037/0096-3445.130.2.199

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. doi:10.1016/j.jml.2007.04.004

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. doi:10.1111/j.1467-9280.2008.02202.x

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General, 134,* 368–387. doi:10.1037/0096-3445.134.3.368

Palladino, P., & De Beni, R. (1999). Short term and working memory in aging: Maintenance and suppression. *Aging Clinical and Experimental Research, 11,* 301–306.

Rose, N. S., & Craik, F. I. M. (2012). A processing approach to the working memory/long-term memory distinction: Evidence from a levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1019–1029. doi:10.1037/a0026976

Rose, N. S., Myerson, J., Roediger, III, H. L., & Hale, S. (2010). Similarities and differences between working memory and long-term memory: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 471–483. doi:10.1037/a0018405

Shelton, J. T., Elliott, E. M., Matthews, R. A., Hill, B. D., & Gouvier, W. D. (2010). The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 813–820. doi:10.1037/a0019046

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language, 49,* 446–468. doi:10.1016/S0749-596X(03)00095-0

Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 1020–1034. doi:10.1037/0278-7393.33.6.1020

Unsworth, N. (2009). Examining variation in working memory capacity and retrieval in cued recall. *Memory, 17,* 386–396. doi:10.1080/09658210902802959

Unsworth, N., & Brewer, G. A. (2009). Examining the relationships among item recognition, source recognition, and recall from an individual differences perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1578–1585. doi:10.1037/a0017255

Unsworth, N., & Brewer, G. A. (2010a). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language, 62,* 19–34. doi:10.1016/j.jml.2009.08.002

Unsworth, N., & Brewer, G. A. (2010b). Variation in working memory capacity and intrusions: Differences in generation or editing? *European Journal of Cognitive Psychology, 22,* 990–1000. doi:10.1080/09541440903175086

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2009). There's more to the working memory capacity–fluid intelligence relationship than just secondary memory. *Psychonomic Bulletin & Review, 16,* 931–937. doi:10.3758/PBR.16.5.931

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2011). Variation in working memory capacity and episodic memory: Examining the importance of encoding specificity. *Psychonomic Bulletin & Review, 18,* 1113–1118. doi:10.3758/s13423-011-0165-y

Unsworth, N., & Engle, R. W. (2006a). A temporal-contextual retrieval account of complex span: An analysis of errors. *Journal of Memory and Language, 54,* 346–362. doi:10.1016/j.jml.2005.11.004

Unsworth, N., & Engle, R. W. (2006b). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54,* 68–80. doi:10.1016/j.jml.2005.06.003

Unsworth, N., & Engle, R. W. (2007). 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. doi:10.1037/0033-295X.114.1.104

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention, memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62,* 392–406. doi:10.1016/j.jml.2010.02.001

Wechsler, D. (1997). *WAIS-III/WMS-III technical manual*. San Antonio, TX: The Psychological Corporation.

Appendix

Correlations Among All Measures for the Whole Sample

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Mean span	1.00																		
2. Immed. word span	.94**	1.00																	
3. Immed. phono. LOP Span	.64**	.65**	1.00																
4. Immed. seman. LOP span	.93**	.80**	.61**	1.00															
5. Immed. # PLI	−.24	−.29*	−.45**	−.21	1.00														
6. Immed. prop. PLI	.15	.17	.05	.20	.56**	1.00													
7. Immed. # ELI	−.37*	−.38**	−.40**	−.27	.32*	.07	1.00												
8. Immed. prop. ELI	.31	.29*	.26	.33*	.10	.66**	.48**	1.00											
9. Immed. # processing	−.49**	−.47**	−.27	−.43**	.30*	−.02	−.02	−.27	1.00										
10. Immed. prop. processing	.05	.01	.17	.12	.07	−.11	−.17	−.17	.73**	1.00									
11. Delay word span	.41**	.48**	.44**	.36*	−.12	.13	−.30*	.12	−.07	.20	1.00								
12. Delay phono. LOP span	.01	.02	−.05	.06	.07	.03	−.14	−.15	.14	.15	.22	1.00							
13. Delay seman. LOP span	.24	.35*	.21	.23	−.24	−.09	−.36*	−.08	.00	.11	.49**	.52**	1.00						
14. Delay # processing	.08	.17	.11	.03	.06	.09	−.01	.11	.00	−.03	.33*	.35*	.59**	1.00					
15. Delay prop. processing	−.23	−.23	−.16	−.39**	.04	−.05	.17	−.02	.07	−.12	−.30*	−.20	−.18	.31*	1.00				
16. Delay # intrusion	−.32*	−.28*	−.13	−.30*	.14	.00	.17	.01	.08	−.01	−.02	.02	.17	.33*	.29*	1.00			
17. Delay prop. intrusion	−.41**	−.42**	−.18	−.36*	.12	−.13	.25	−.05	.13	.03	−.02	.15	.12	.21	.23	.86**	1.00		
18. Delay # source errors	−.32*	−.30*	−.11	−.32*	.26	.09	.23	.13	.24	.03	−.14	−.23	−.07	.11	.33*	.45**	.53**	1.00	
19. Delay prop. source errors	−.41**	−.40**	−.21	−.36*	.22	.02	.24	.04	.19	.00	−.33*	−.41**	−.35*	−.27	.42**	.27	.30*	.67**	1.00
20. Matrix reasoning	.26	.19	.14	.25	−.14	.03	−.32*	−.06	.03	.15	.19	−.09	−.01	−.12	−.23	−.51**	−.56**	−.45**	−.27

Note. Immed. = immediate; phono. = phonological; seman. = semantic; LOP = levels of processing; PLI = prior list intrusion; ELI = extra list intrusion; prop. = proportion.
$^{*}p < .05$. $^{**}p < .01$.

Received December 14, 2012
Accepted August 9, 2013 ■