

Variation in working memory capacity, fluid intelligence, and episodic recall: A latent variable examination of differences in the dynamics of free recall

NASH UNSWORTH

University of Georgia, Athens, Georgia

A latent variable analysis was conducted to examine the nature of individual differences in the dynamics of free recall and cognitive abilities. Participants performed multiple measures of free recall, working memory capacity (WMC), and fluid intelligence (gF). For each free recall task, recall accuracy, recall latency, and number of intrusion errors were determined, and latent factors were derived for each. It was found that recall accuracy was negatively related to both recall latency and number of intrusions, and recall latency and number of intrusions were positively related. Furthermore, latent WMC and gF factors were positively related to recall accuracy, but negatively related to recall latency and number of intrusions. Finally, a cluster analysis revealed that subgroups of participants with deficits in focusing the search had deficits in recovering degraded representations or deficits in monitoring the products of retrieval. The results are consistent with the idea that variation in the dynamics of free recall, WMC, and gF are primarily due to differences in search set size, but differences in recovery and monitoring are also important.

Free recall, wherein participants are presented with a list of items and are asked to recall the items in any order they wish, is one of the oldest and most heavily studied tasks in memory research (Crowder, 1976; Murdock, 1974; Tulving, 1968). Much previous research has been done on this task, demonstrating how various manipulations, such as list length, presentation rate, word frequency, and distractor-filled delays influence overall performance. In addition, a number of well-developed models have been created in order to explain the nature of free recall (see Raaijmakers & Shiffrin, 1992, for a review). Finally, a number of studies have shown that free recall measures are related to other memory measures and related to other well-known cognitive abilities, such as intelligence (Beier & Ackerman, 2004; Carroll, 1993). As such, this work points to the overall importance of free recall as a measure of memory that has not only constrained theories of memory, but has also been used to assess variation in overall memory performance.

The present study examined the dynamics of free recall at a latent-variable level in order to better understand the processes that drive recall and the nature of individual differences in recall. Recall accuracy, recall latency, and intrusion errors were examined at a latent level to determine the interrelations between different theoretical components of free recall. Specifically, on the basis of search models of free recall, individual differences in constraining the memory search, recovering degraded representations, and monitoring the products of the search were ex-

amined. Below, a brief review of search processes and the dynamics of free recall is given, followed by a brief review of individual differences in the dynamics of free recall. Finally, the rationale for the present study is given.

Search Processes and the Dynamics of Free Recall

As noted above, a number of detailed models of free recall have been developed over the last 40 years (Raaijmakers & Shiffrin, 1992). Some of the most successful models of free recall are search models that suggest that retrieval of information from memory involves a competitive search in order to locate the desired information among irrelevant and distracting information. In these search models, it is assumed that, during recall, a retrieval cue activates a subset of representations in memory that are related to the cue in some fashion. This delimited subset is known as the *search set*, and during recall, item representations are sampled (with replacement) from the search set on the basis of a relative strength rule (Raaijmakers & Shiffrin, 1980; Rohrer, 1996; Shiffrin, 1970b). Specifically, in search models of this type, the probability of sampling any particular item is equal to the strength of the item divided by the sum of all item strengths within the search set. After an item has been sampled, it must then be recovered into consciousness. In these search models, recovery of an item depends on the item's absolute strength rather than on its relative strength. Specifically, items whose strength ex-

N. Unsworth, nunswor@uga.edu

ceeds some critical threshold will be recovered and can be recalled, whereas weak items that do not exceed the threshold will not be recovered. Important for models of this type is the notion that all items can be sampled, but only those items whose strength exceeds the threshold can actually be recalled. Thus, it is possible to differentiate these two aspects of recall (sampling and recovery). Finally, after an item has been recovered, it is subjected to a monitoring and editing process that determines whether the item is correct and recalled or incorrect and not recalled.

These three components (sampling, recovery, and monitoring) make up the three primary components of free recall, and each can be affected by different variables and have different markers. For instance, in the simplest version of the relative strength model of sampling, the random search model, it is assumed that all items have the same probability of being sampled, and hence, the search is random (Bousfield, Sedgewick, & Cohen, 1954; Kaplan, Carvellas, & Metlay, 1969; McGill, 1963; Rohrer & Wixted, 1994; Wixted & Rohrer, 1994). In this model, item representations are randomly sampled from the search set at a constant rate, one item at a time (serial search). The search includes sampling with replacement, such that, after an item has been sampled, it has an equal chance of being selected on the next sample. Thus, target items that have been previously recalled, intruding items, or target items that are not recoverable all have an equal chance of being sampled. As the recall period progresses, the probability of recalling a new target item decreases, because each sample will likely generate an already sampled item or an intrusion. Assuming a constant sampling time per item, this model predicts that any increase in the size of the search set should lead to a concomitant increase in *recall latency* (Wixted & Rohrer, 1994), the average time during the recall period over which items are recalled. A number of studies have provided empirical evidence for this prediction. Specifically, Rohrer and Wixted (1994) showed that when list length increases, so does recall latency. Thus, as the size of the search set increases, the time needed to sample any given item also increases, leading to longer recall latencies. Furthermore, Wixted and Rohrer (1993) showed that recall latency increased as proactive interference increased, suggesting that the size of the search set increased because of the inclusion of previous list intrusions. These results suggest that recall latency provides an index of sampling on the basis of relative strength rule, which is directly related to the size of the search set (see also Rohrer, 1996).

Similarly, research has suggested that *recall total* (the total number of items recalled) provides an index of the number of recoverable targets within the search set. That is, recall total provides an index of the number of items in the search set whose strength exceeds some threshold and that are subsequently recovered. For instance, Rohrer and Wixted (1994) found that recall total increased as presentation duration increased. Presumably, increasing the presentation duration increased the strength of the items, leading to more items being recovered. Importantly, Rohrer and Wixted (1994) found that, although increasing the presentation duration increased the recall total, there was no change in recall latency. Thus, manipulations of presenta-

tion duration affected the recovery stage but not the sampling stage. This dissociation occurs because the size of the search set did not change (and hence did not change the relative strength of items), but each item's overall strength increased, leading to a higher proportion of target items whose strength exceeded the critical threshold. This work suggests that recall latency provides an index of the size of the search set, but recall total provides an index of the number of recoverable targets within the search set.

Finally, recall errors can be seen as an index of failures in the monitoring/editing processes that items are subjected to after they are recovered. Recall errors include items that were presented on previous lists (previous-list intrusions), items not presented in the experiment (extralist intrusions), as well as target items that are repeated (repetitions). Although these errors are generally rare, a number of studies have shown that they are quite systematic. For instance, previous-list intrusions typically come from the immediately preceding list (Murdock, 1974; Unsworth & Engle, 2007; Zaromb et al., 2006), extralist intrusions tend to be either semantically or phonologically related to one of the target items (Craik, 1968), and repetitions tend to occur late in the recall period and are recalled after several intervening items have been recalled since the original target item was recalled (e.g., Unsworth, 2008). Thus, recall errors provide an index of the monitoring/editing process that operates after recovery.

Collectively, the results suggest that recall can be envisioned as a multistage process whereby items are first sampled from a variable-size search set. Sampled items are then either recovered or not recovered, and a new item is sampled. After an item has been recovered, it is subjected to a monitoring/editing process; if deemed to be a target response, it is recalled, and if it is incorrect, it is not recalled. These three stages represent distinct components of free recall and are linked to separate measures (recall latency, recall total, and recall errors). At the same time, although these different components are affected by different variables, it is unlikely that they are completely independent. For instance, recall total is affected not only by the absolute strength of the items in the search set, but also by the number of irrelevant representations in the search set. The more intrusions that are included in the search set, the lower the probability of finding target representations will be (Unsworth, 2007; Wixted & Rohrer, 1993). In addition, recall errors come about because intrusions were first included in the search set, then were actually recovered, and, finally, the monitoring process failed and they were not edited out. Thus, although these different measures provide an index of the different components, the measures are likely interdependent on the other components as well. As such, these different measures should be related to one another at the level of individual differences.

Individual Differences in Search and the Dynamics of Recall

Not only have the dynamics of recall across various manipulations been examined in previous research, but individual and group differences in the dynamics of free recall have been explored in more recent work (e.g., Rohrer,

Salmon, Wixted, & Paulsen, 1999; Wingfield, Lindfield, & Kahana, 1998). This is important, because as was previously noted, a number of studies have shown that free recall measures are related to a number of important cognitive abilities (Beier & Ackerman, 2004; Carroll, 1993). Specifically, within the confines of working memory capacity (WMC), it has been suggested in previous work that individual differences in WMC are related to aspects of recall (Unsworth, 2007; Unsworth & Engle, 2007). In this work, we argued that high- and low-WMC individuals (this distinction is based on complex span measures of working memory) differ in their ability to guide the search process, such that low-WMC individuals rely on noisier cues than high-WMC individuals do. This results in low-WMC individuals' including many more irrelevant representations (intrusions) in their search sets than high-WMC individuals do. The result is that low-WMC individuals recall fewer correct items, recall at a slower rate, and recall more errors than high-WMC individuals do. Evidence consistent with this interpretation came from studies in which immediate, delayed, and continual distractor free recall were examined (Unsworth, 2007; Unsworth & Engle, 2007). In each case, low-WMC individuals recalled fewer correct items, had longer recall latencies (and interresponse times), and recalled more intrusions (particularly, previous-list intrusions) than did high-WMC individuals. Similar results have also recently been obtained with cued recall (Unsworth, 2009). These results suggest that high- and low-WMC individuals differ only in the size of their search sets—and, hence, in the sampling component of recall—but did not differ in the number of available targets or in the effectiveness of the monitoring component. Indeed, simulations in which high- and low-WMC individuals differed either in overall search set size or in the number of recoverable targets suggested that only differences in search set size were consistent with the obtained pattern of results (Unsworth, 2007). Thus, this result suggests that individual differences in WMC and free recall might be best conceptualized as differences specifically in the sampling component, with low-WMC individuals searching through a larger set of items than high-WMC individuals did.

Some recent work in which the dynamics of recall in cued recall was examined qualifies this overall position, however (Unsworth, 2009). Specifically, in Unsworth (2009), high- and low-WMC individuals were tested either on a basic cued recall task with unrelated cue and target words (Experiment 1) or with specific cue phrases (Experiment 2). Like those in the free recall studies, low-WMC individuals tended to recall fewer correct items, to recall at a slower rate, and to recall more intrusions than high-WMC individuals did. Furthermore, giving participants more specific cue phrases increased overall proportion correct, decreased recall latency, and decreased the number of intrusions, suggesting that the use of specific cue phrases served to focus the search on the target items. However, when specific cue phrases were given, high- and low-WMC individuals differed in the number of overall omissions made. This suggests that specific cues focused the search for both high- and low-WMC individuals, but low-WMC individuals still performed more poorly because they had

more degraded/nonrecoverable representations than high-WMC individuals did. Additionally, differences between high- and low-WMC individuals in extralist intrusions in the specific cue experiment suggested differences in the monitoring component. Thus, although variation in WMC is most consistent with differences in search set size, it also seems likely that there are differences in both recovery and monitoring. Indeed, evidence consistent with this hypothesis comes from an examination of correlations between the different measures in cued recall. It was found that overall proportion correct was related both to the number of intrusions and to the number of omissions, and these two error types were not related. This result suggests that proportion correct is determined both by the number of recoverable targets and by the overall search set size (see above), but these two are not related. Furthermore, recall latency was related to proportion correct and number of intrusions, suggesting that it is a valid measure of search set size. Overall, these results provide preliminary support for the notion that individual differences in the dynamics of recall arise from differences in search set size, as well as from differences in recovery and monitoring.

The Present Study

The goal of the present study was to examine the notion that individual differences in the dynamics of free recall result from differences in search set size, as well as differences in recovery and monitoring abilities. Two primary research questions were addressed. First, to what extent are the three components of recall outlined above related to one another at an individual differences level? Although the dynamics of free recall has been examined in previous work and evidence has been shown for differences between recall accuracy, recall latency, and recall errors, no study has yet examined how these three measures/components are related at an individual differences level. This is important, because although it has been suggested that recall accuracy and recall latency can be experimentally dissociated (Rohrer & Wixted, 1994) and that there is no correlation between recall accuracy and recall latency (e.g., MacLeod & Nelson, 1984; Scheirer, 1971), search models of the type discussed here suggest that it is likely that these different measures will be related. As was previously noted, if a large number of intrusions are included in the search set, recall latency should increase and recall accuracy should decrease. Thus, there should be a negative correlation between recall accuracy and recall latency, a positive correlation between recall latency and number of intrusions, and a negative correlation between recall accuracy and number of intrusions.

Second, how do each of the recall components relate to individual differences in WMC and cognitive abilities? As noted above, it has been suggested in previous work that each of these components should be related at a broad level to individual differences in WMC (e.g., Unsworth, 2007). For more specific analyses concerning WMC differences in serial position functions and the cumulative recall functions, see Unsworth (2007). Furthermore, given the substantial correlation between WMC and fluid intelligence (gF) that has been found previously (e.g., Engle, Tuholski, Laughlin, & Conway, 1999), and the fact that recall and

gF tend to be correlated (Bors & Forrin, 1995), the pattern of results for the recall components and WMC should also be found for gF to the extent that the underlying processes are important for both. Similar to those in prior work (e.g., Unsworth, 2007, 2009), four different possible patterns of results were tested. In all four possibilities, low-ability (i.e., low-WMC, low-gF) participants should recall fewer correct target items than high-ability participants, but differences between the possibilities should appear in the other two measured components (i.e., recall latency and intrusions). Thus, what distinguishes the four possibilities is the pattern of results across the different measures.

On the basis of prior work (e.g., Unsworth, 2007), the first possibility (*low ability large*) is that low-ability individuals have larger search sets (because of the inclusion of more intrusions) than high-ability individuals. This should result in a positive correlation between the ability measures (WMC and gF) and recall accuracy, a negative correlation between the ability measures and recall latency, and a negative correlation between the ability measures and number of intrusions. Thus, low-ability participants should recall fewer target items, recall at a slower rate, and recall more intrusions than should high-ability participants, because of an inability to filter out irrelevant representations from their search sets. Importantly, in this view, low-capacity individuals actually activate more information in their search sets than high-capacity individuals do, contrary to the typical conceptualization of capacity (see also Vogel, McCollough, & Machizawa, 2005).

The second possibility (*low ability small*) is that low-ability participants have less capacity (or resources; Cantor & Engle, 1993; Just & Carpenter, 1992) to activate target representations, which leads to smaller search sets overall. This should result in a positive correlation between the ability measures (WMC and gF) and recall accuracy, because some of the target items were not included in the search set. However, because low-ability participants have smaller overall search sets, they should recall items that are actually in the search set at a faster rate than high-ability participants, leading to a positive correlation between the ability measures and recall latency. This possibility is consistent with prior work in semantic memory in which high-ability participants recalled at a slower rate than low-ability participants (e.g., Johnson, Johnson, & Mark, 1951). This possibility also suggests that there should likely be no correlation between the ability measures and intrusions, because low-ability participants should not have enough capacity to activate irrelevant representations.

The third possibility (*low ability nonrecoverable*) is that low-ability participants have fewer recoverable targets in their search sets than high-ability participants do. This could be due to differences in basic encoding abilities (e.g., differences in rehearsal, elaboration, or binding), which result in fewer correct items being recalled. Like the other possibilities, this possibility would predict a positive correlation between the ability measures and recall accuracy. However, this possibility would suggest no (or a very weak) correlation between the ability measures and recall latency. This is because, if participants have search sets of the same size, but simply differ in the number of

recoverable targets within the search sets, the participants should differ in overall recall accuracy, but they should not differ in how quickly they recall the correct items. This is consistent with work by Rohrer and Wixted (1994) demonstrating that increasing presentation duration, and presumably increasing the amount of rehearsal or attention devoted to items, leads to an increase in recall accuracy but to no change in recall latency.

The final possibility (*low ability monitoring*) is that low-ability participants may have specific deficits in their ability to monitor and catch errors before they are recalled but have no differences in the size of their search sets or in the number of recoverable targets within their search sets. This would result in a positive correlation between the ability measures and recall accuracy, because low-ability participants would incorrectly recall many intrusions rather than correct target items. However, because all of the individuals search through roughly the same number of items, there should be no (or a relatively weak) correlation between the ability measures and recall latency. Importantly, this possibility predicts a negative correlation between the ability measures and the total number of intrusions recalled.

Each of these possibilities predicts a unique pattern of correlations among the different recall components and WMC and gF, which can be examined at an individual differences level. Specifically, although each possibility predicts a positive correlation between ability and recall accuracy, the possibilities differ in terms of the correlations among ability, recall latency, and intrusions. Thus, the overall pattern of correlations is what is important rather than the correlation between recall accuracy and ability, as has been previously examined.

Additionally, it may be possible to find specific deficits for low-ability participants whereby some low-ability participants have specific deficits in focusing their search sets, whereas others have deficits in terms of the number of degraded representations in their search sets or in the efficiency of their monitoring/editing processes. This would result in subgroups of participants who have specific deficits in one component but not necessarily global deficits manifested on all components. Collectively, these results should provide strong evidence that search models, which have been extremely successful in accounting for a number of findings in free recall (e.g., Raaijmakers & Shiffrin, 1980), can also account for individual differences in free recall, which have been shown to be important predictors of a number of cognitive abilities, including gF (e.g., Beier & Ackerman, 2004; Carroll, 1993).

In order to address these questions, a latent-variable approach was used to examine the relations between the different components with each other and with WMC and gF. Specifically, the latent-variable approach was used for three reasons. (1) Typically, only a single task at a time has been examined in previous studies, and thus, these results may be due to idiosyncratic task effects and, therefore, may not provide the best evidence for more general principles across different recall measures. (2) No study has yet examined whether recall latency and recall errors extracted from different recall tasks actually measure the same thing. That is, previous researchers on the dynamics of free recall have

typically only examined performance on a single task at a single time. If recall latency reflects a stable process or set of processes, this process should be needed on all free recall tasks, and the recall latency measures extracted from different tasks should be related to one another and should form a single factor. The same goes for recall errors. Intrusions in one recall task should be related to intrusions in another recall task if they reflect the same underlying process. Although this seems like a straightforward prediction, no study has yet actually demonstrated this result. (3) Previous work in which individual differences in the dynamics of free recall have been examined has only been done with extreme groups (e.g., Unsworth, 2007; Unsworth & Engle, 2007), which can possibly lead to biases in the results (e.g., Conway et al., 2005; Preacher, Rucker, MacCallum, & Nicewander, 2005). Thus, it is important to examine these issues using the full range of participants, and it is important to examine these predictions for other cognitive ability constructs, such as gF. Therefore, in order to derive latent variables of WMC, gF, and the three recall components, three WMC tasks (operation span [ospan], reading span [rspan], and symmetry span [symspan]), three gF tasks (Raven advanced progressive matrices, number series, and verbal analogies), and three free recall tasks (delayed free recall with unrelated words, delayed free recall with semantically related words, and list-before-last recall) were used. Note that standard immediate free recall was not used in the present study, because previous work has suggested that two components (primary and secondary memory) underlie performance on this task, and the two components have different retrieval dynamics (e.g., Unsworth & Engle, 2007). Thus, only tasks thought to rely primarily on secondary memory search processes were used. For these free recall tasks, number of items correctly recalled, recall latency, and number of intrusions were recorded.

METHOD

Participants

A total of 186 participants were recruited from the University of Georgia. The participants were between 18 and 35 years old and received course credit for their participation. Each participant was tested individually in two laboratory sessions lasting approximately 2 h each.

Materials and Procedure

All of the participants completed the ospan task, the symspan task, the rspan task, a delayed free recall task with semantic category switches, a delayed free recall task with unrelated words, and a version of the list-before-last task (Ward & Tan, 2004), Raven advanced progressive matrices (Raven, Raven, & Court, 1998), a version of Thurstone's (1962) number series test, and a verbal analogies test. All of the tasks were administered in the order above.

Tasks

WMC Tasks

Ospan. The participants were required to solve a math operation, and, after solving the operation, they were presented with a letter from a set of unrelated letters (F, H, J, K, L, N, P, Q, R, S, T, and Y) for 1 sec. Immediately after the letter was presented, the next operation was presented. Three trials of each list length (3–7) were presented, with the order of list length varying randomly. At recall, the letters from the current set were recalled in the correct order by

clicking on the appropriate letters (see Unsworth, Heitz, Schrock, & Engle, 2005, for more details). The participants received three sets (of list length 2) of practice. For all of the span measures, items were scored if the item was correct and in the correct position. The score was the proportion of correct items in the correct position.

Symspan. In the symmetry judgment task, participants were shown an 8×8 matrix with some squares filled in black. Participants decided whether the design was symmetrical about its vertical axis. The pattern was symmetrical half of the time. Immediately after determining whether the pattern was symmetrical, the participants were presented with a 4×4 matrix, with one of the cells filled in red, for 650 msec. At recall, participants recalled the sequence of red-square locations in the preceding displays, in the order in which they appeared, by clicking on the cells of an empty matrix. There were three trials of each list length, with list length ranging from 2 to 5. The same scoring procedure as that for the ospan task was used.

Rspan. The participants were required to read sentences while trying to remember the same set of unrelated letters as in the ospan task. For this task, participants read a sentence and determined whether the sentence made sense or not (e.g., "The prosecutor's dish was lost because it was not based on fact."). Half of the sentences made sense, whereas the other half did not. Nonsense sentences were made by simply changing one word (e.g., "dish" from "case") from an otherwise normal sentence. Participants were required to read the sentence and to indicate whether it made sense or not. After participants gave their response, they were presented with a letter for 1 sec. At recall, the letters from the current set were recalled in the correct order by clicking on the appropriate letters. There were three trials of each list length, with list length ranging from 3 to 7. The same scoring procedure as that for the ospan task was used.

gF Tasks

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

Number series. In this task, participants saw a series of numbers and were required to determine what the next number in the series should be (Thurstone, 1962). That is, the series follows some unstated rule that the participants are required to figure out in order to determine which number should be next in the series. Participants selected their answers out of five possible numbers. Following five practice items, participants had 4.5 min to complete 15 test items. A participant's score was the total number of items solved correctly.

Verbal analogies. In this task, participants read an incomplete analogy and were required to select the one word out of five possible words that best completed the analogy. After one practice item, the participants had 5 min to complete 18 test items. These items were originally selected from the Air Force Officer Qualifying Test (Berger, Gupta, Berger, & Skinner, 1990), and we used the same subset of items used in Kane et al. (2004). A participant's score was the total number of items solved correctly.

Free Recall Tasks

Delayed free recall, unrelated words. In this task, the participants were given six lists of 10 words each. All of the words were common nouns and were presented for 1 sec each. After list presentation, participants engaged in a 16-sec distractor task before recall: Participants saw eight three-digit numbers appear for 2 sec each and were required to write the digits in ascending order. After the distractor task, participants saw "???" which indicated that they should type as many words as they could remember from the current list in any order they wished.¹ Participants had 45 sec for recall. Proportion

of items recalled correctly, recall latency, and the total number of intrusion errors were recorded. Here, *recall latency* refers to the time point in the recall period when a given response was emitted. Thus, if responses were emitted 5, 10, and 15 sec into the recall period, the mean recall latency would be 10 sec.

List-before-last recall. This task was a variant of the list-before-last task developed by Shiffrin (1970a) and modified by Ward and Tan (2004). On each trial in this task, participants were presented with 2 lists of 10 words each. There were six trials (12 lists total). All of the words were common nouns that were presented for 1 sec each. Each list was labeled as either List 1 or List 2, and the list labels preceded each list for 3 sec. Participants were told to remember both lists, and, at recall, they would be cued to recall either List 1 or List 2. During recall, participants had 45 sec to type as many words from the cued list as possible. Proportion of items recalled correctly, recall latency, and total number of intrusion errors were recorded.

Delayed free recall, semantically related words. Participants received six lists of 10 words each, broken down into two blocks (three lists per block). All of the words in each block came from the same semantic category (e.g., professions or fruits). The first three lists allowed for proactive interference to accrue, and the first list in the next block allowed for a release from proactive interference. Following the last word in a list, participants were required to count backward by threes as quickly and accurately as possible from a three-digit number onscreen for 15 sec and to write the numbers down as they went. After the distractor task, participants saw “???” which indicated that they should type as many words as they could remember from the current list in any order they wished. Participants had 45 sec for recall. Proportion of items recalled correctly, recall latency, and total number of intrusion errors were recorded.

RESULTS AND DISCUSSION

Preliminary Analyses

Descriptive statistics for the memory measures are shown in Table 1. As can be seen in Table 1, all of the measures had generally acceptable values of internal consistency, and most of the measures were approximately normally distributed, with values of skewness and kurtosis under the generally accepted values (i.e., skewness < 2 and kurtosis < 4; see Kline, 1998), except for the intrusion error measures, which were highly skewed. This is not surprising, given that, in general, so few errors were actually made. Additionally, consistent with Wixted and Rohrer (1993), there was sub-

Table 1
Descriptive Statistics and Reliability Estimates for the Memory and Intelligence Measures

Measure	<i>M</i>	<i>SD</i>	Skew	Kurtosis	<i>α</i>
Ospan	60.26	10.60	-1.29	1.67	.79
Symspan	29.32	7.60	-0.51	-0.43	.77
Rspan	57.26	12.34	-1.04	0.96	.78
Raven	10.30	2.50	-0.34	-0.22	.73
NS	9.44	2.48	-0.05	-0.43	.71
Analogy	11.31	3.03	0.24	1.68	.68
DFRUacc	0.54	0.11	0.17	0.24	.70
LBLacc	0.42	0.12	0.30	0.42	.77
DFRSacc	0.58	0.09	-0.11	-0.33	.63
DFRUlat	15,700	3,268	0.72	0.47	.66
LBLlat	12,994	3,724	0.57	0.44	.70
DFRSlat	16,977	2,687	0.01	0.28	.69
DFRUIntru	3.03	3.17	2.52	10.47	.74
LBLIntru	5.63	4.49	1.30	2.08	.67
DFRSIntru	2.35	2.19	2.29	10.65	.72

Note—Ospan, operation span; Symspan, symmetry span; Rspan, reading span; Raven, Raven advanced progressive matrices; NS, number series; Analogy, verbal analogy; DFRU, delayed free recall unrelated words; LBL, list-before-last recall; DFRS, delayed free recall semantically related words; acc, recall accuracy; lat, recall latency; Intru, intrusions.

stantial proactive interference in the delayed free recall task with semantically related words in both proportion correct and recall latency. Specifically, proportion correct dropped from Trial 1 to Trial 3, collapsed across the two blocks, by 30% (*SE* = .01), and recall latency increased from Trial 1 to Trial 3 by 1,788 msec (*SE* = 400) (both *t*s > 4.45, both *p*s < .01). Finally, correlations for all of the measures are shown in Table 2.

Next, an exploratory factor analysis was done to examine the initial structure of the data for the recall measures. In particular, the exploratory factor analysis was done (1) to examine the communality estimate of each recall measure, which estimates the proportion of variance accounted for in a measure by the other measures in the factor analysis; (2) to examine the loading of each recall measure on the first nonrotated factor, which provides an indication of the shared variance among the measures on a common factor; and (3) to examine the initial factor structure of the data to

Table 2
Correlations for All of the Memory and Intelligence Measures

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Ospan	—														
2. Symspan	.49	—													
3. Rspan	.61	.44	—												
4. Raven	.26	.29	.34	—											
5. NS	.23	.27	.24	.26	—										
6. Analogy	.13	.10	.24	.34	.27	—									
7. DFRUacc	.30	.23	.24	.17	.24	.11	—								
8. LBLacc	.12	.17	.15	.08	.19	.21	.50	—							
9. DFRSacc	.12	.02	.16	.06	.15	.24	.38	.37	—						
10. DFRUlat	-.29	-.30	-.26	-.22	-.19	-.21	-.29	-.17	-.16	—					
11. LBLlat	-.14	-.19	-.17	-.09	-.10	-.13	.03	.02	.47	—					
12. DFRSlat	-.18	-.21	-.25	-.18	-.17	-.13	-.14	-.15	-.05	.43	—				
13. DFRUIntru	-.10	-.17	-.12	-.14	-.07	-.12	-.25	-.12	.05	.21	.13	—			
14. LBLIntru	-.15	-.25	-.17	-.16	-.15	-.21	-.22	-.29	.02	.18	.34	.30	—		
15. DFRSIntru	-.06	-.09	-.18	-.12	-.01	-.10	-.16	-.08	-.04	.10	-.03	.13	.42	—	

Note—Ospan, operation span; Symspan, symmetry span; Rspan, reading span; Raven, Raven advanced progressive matrices; NS, number series; Analogy, verbal analogy; DFRU, delayed free recall unrelated words; LBL, list-before-last recall; DFRS, delayed free recall semantically related words; acc, recall accuracy; lat, recall latency; Intru, intrusions. Correlations > .15 are significant at the *p* < .05 level.

Table 3A
Exploratory Factor Analysis for the Recall Measures

Measure	<i>h</i> ²	Nonrotated 1st	Rotated Factor		
			1	2	3
DFRU	.57	-.52			.72
LBL	.45	-.45			.65
DFRS	.33	-.22	.19		.59
DFRUlat	.48	.54		.64	-.21
LBLlat	.62	.43		.81	.19
DFRSlat	.34	.47		.54	
DFRUIntru	.67	.63	.85		
LBLIntru	.65	.72	.74	.16	
DFRSIntru	.25	.38	.52		

Note—DFRU, delayed free recall unrelated words; LBL, list-before-last recall; DFRS, delayed free recall semantically related words; acc, recall accuracy; lat, recall latency; Intru, intrusions; *h*², communality estimate; Nonrotated 1st, the first nonrotated factor. Factor loadings <.15 have been omitted for clarity.

Table 3B
Correlations for the Recall Measures

Factor	Factor		
	1	2	3
1	—		
2	.37	—	
3	-.27	-.19	—

see how the recall measures would load on certain factors and how these factors would be related. Therefore, in order to examine the factor structure of the recall measures, a principal factor analysis with promax rotation (oblique rotation) was conducted on the recall measures. As is shown in Table 3A, the factor analysis yielded three factors (Factor 1 eigenvalue = 2.74, Factor 2 eigenvalue = 1.64, Factor 3 eigenvalue = 1.44), accounting for 48.27% of the variance. The scree plot also suggests the presence of three factors. The first factor consists primarily of the intrusion error measures from the three free recall tasks, along with a weak loading of the delayed free recall with semantically related words recall accuracy measure. The second factor consists primarily of the recall latency measures from the three free recall tasks, along with a weak loading of the list-before-last intrusion error measure. Finally, the third factor consists of the recall accuracy measures from the three recall tasks, along with weak loadings from the delayed free recall and list-before-last recall latency measures. In general, the results suggest that recall accuracy, recall latency, and intrusions from the three different recall tasks clustered together, and, thus, these three components accounted for the basic structure of the data. In addition, the three factors were weakly to moderately related, and the direction of the correlations was in the predicted direction. These results provide initial support for the three components and their possible relations.

Also shown in Table 3A are the communality estimates (*h*²) for each measure. As can be seen, nearly all of the measures demonstrated fairly strong communality estimates, suggesting that the measures share a great deal of variance with the other measures included in the factor analysis. This notion is further supported by an examination of each measures' loading on the first nonrotated factor shown in

Table 3A. Again, all of the measures demonstrated moderate to strong loadings on the first nonrotated factor, suggesting a good deal of common variance among the measures. This result provides preliminary evidence for the relations among the recall components and suggests that each of the measures from the different tasks can be grouped together.

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) was used to further examine the underlying structure of the data and the relations among the three recall components with each other and with WMC and gF. Specifically, CFA was used to examine the relation among the three components at a latent level, similar to the exploratory factor analysis, but, here, each measure was forced to load on only one factor, and the fit of the model was assessed. In addition, the three WMC measures and the three gF measures were included to examine the relation among the three recall components with WMC and gF at the latent level. As was previously noted, the pattern of correlations between the recall components and the ability latent variables should provide evidence in favor of one of the four possibilities.

Model fits were assessed via a combination of several fit statistics. These include chi square, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), the non-normed fit index (NNFI), and the comparative fit index (CFI). The chi-square statistic reflects whether there is a significant difference between the observed and reproduced covariance matrices. Therefore, non-significant values are desirable. However, with large sample sizes, even slight deviations can result in a significant value; therefore, the ratio of chi square to the number of degrees of freedom is also reported. Ratios of two or less usually indicate an acceptable fit. Also reported are the RMSEA and the SRMR, both of which reflect the average squared deviation between the observed and reproduced covariances. In addition, the NNFI and the CFI, both of which compare the fit of the specified model to a baseline null model. NNFI and CFI values greater than .90 and RMSEA and SRMR values less than .08 are indicative of an acceptable fit (Kline, 1998).

Shown in Figure 1 is the resulting CFA model. In this model, each measure extracted from the three recall tasks loaded on one specific factor with no cross loadings (i.e., DFR recall accuracy loaded on the accuracy factor only). Furthermore, the three WMC measures loaded only on the WMC factor and the three gF measures loaded on the gF factor. Correlations among all of the factors varied freely. As can be seen in Figure 1, all measures loaded significantly and substantially on their respective factors, and the overall fit of the model was acceptable [$\chi^2(80) = 134.82, p < .01, \chi^2/df = 1.69, RMSEA = .06, SRMR = .06, NNFI = .92, CFI = .96$]. Furthermore, as is shown in Figure 1, all of the factors were significantly correlated with one another. Specifically, as with the exploratory factor analysis, recall accuracy and recall latency were negatively related, recall accuracy and intrusions were negatively related, and recall latency and intrusions were positively related. In addition, both WMC and gF were positively related to recall accuracy but were negatively related to both recall latency and intrusions. This pattern of results is consistent with the

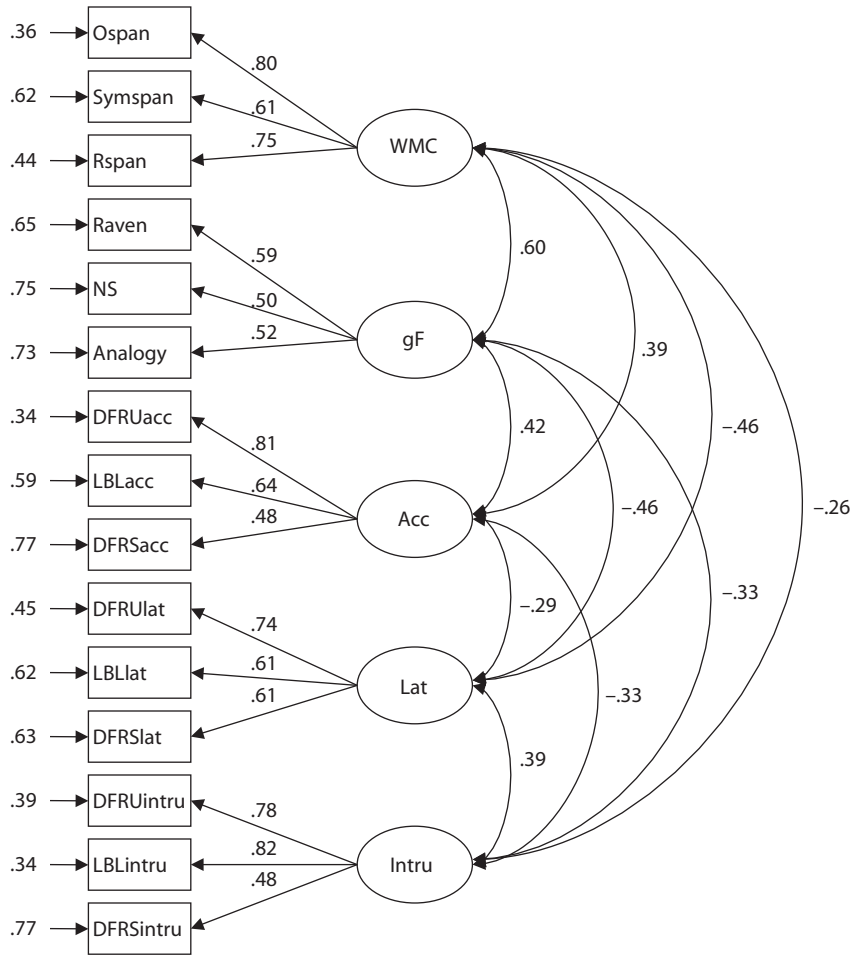


Figure 1. Confirmatory factor analysis for working memory capacity (WMC), general fluid intelligence (gF), recall accuracy (Acc), recall latency (Lat), and intrusions (Intru). Paths connecting latent variables (circles) to each other represent the correlations between the constructs, the numbers from the latent variables to the manifest variables (squares) represent the loadings of each task onto the latent variable, and numbers appearing next to each manifest variable represent error variance associated with each task. All paths are significant at the $p < .05$ level.

first possibility outlined in the introduction, suggesting that high-ability (both WMC and gF) individuals recall a greater proportion of target items, a lower proportion of errors, and recall at a faster rate than low-ability individuals, possibly because of the fact that low-ability individuals include more irrelevant representations in their search sets than high-ability individuals do. This pattern is inconsistent with the other possibilities, mainly because there was a significant negative correlation between the ability factors and recall latency. The other possibilities outlined in the introduction suggest either that there should be no correlation (if there are differences only in encoding or monitoring) or a positive correlation (if low-ability individuals do not have enough capacity to activate the correct target representations). Thus, overall, the results are consistent with Unsworth (2007), but, importantly, these results are for the full range of participants and for multiple measures of WMC, gF, and recall.

Given that all of the recall latent variables were interrelated and the fact that all of the measures loaded strongly on

the first nonrotated factor in the exploratory factor analysis, with the next CFA, we examined the extent to which a higher-order recall factor could be extracted from the three lower-order recall factors and how this higher-order factor would be related to WMC and gF. That is, this model tests the notion that there is substantial common variance between the three recall latent variables and that this common variance is related to WMC and gF. All of the latent variables from the prior CFA remained the same, but here, a higher-order factor was specified, such that the three recall latent variables loaded onto a single higher-order factor (recall) and this higher-order factor was allowed to correlate with WMC and gF. The loadings of all of the measures remained the same as those in Figure 1. The fit of the model was acceptable [$\chi^2(84) = 138.51, p < .01, \chi^2/df = 1.65, RMSEA = .06, SRMR = .06, NNFI = .92, CFI = .94$]. Shown in Figure 2 is the resulting model. As can be seen, the three lower-order factors loaded strongly on the higher-order factor, suggesting the presence of a good deal of com-

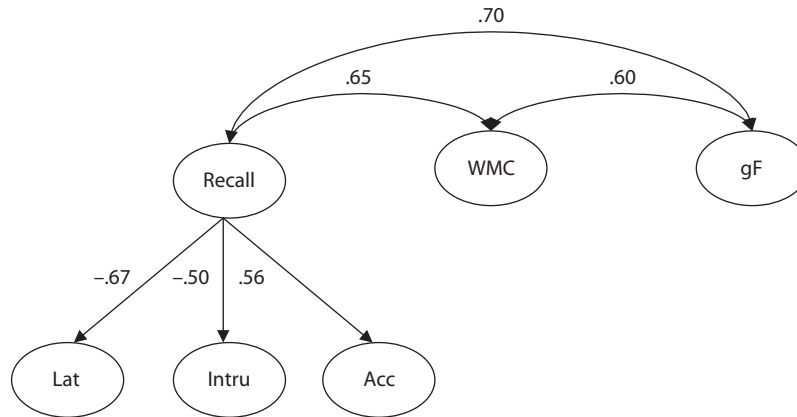


Figure 2. Confirmatory factor analysis for higher order recall factor (Recall) based on lower order factors composed of recall accuracy (Acc), recall latency (Lat), and intrusions (Intru) and the relation among the higher-order factor, working memory capacity (WMC), and general fluid intelligence (gF). Single-headed arrows connecting latent variables (circles) to each other represent standardized path coefficients indicating the unique contribution of the latent variable. Double-headed arrows connecting latent variables indicate the correlation between the latent variables. All paths are significant at the $p < .05$ level.

mon variance that was strongly related to both WMC and gF. Thus, this result suggests that a single factor based on longer recall latencies, more intrusions, and a lower proportion of correct responses was strongly related to WMC, consistent with previous theorizing (Unsworth, 2007), and this factor was related to overall intellectual functioning in gF.

Variance Partitioning

Given that each of the three recall latent variables was related to both WMC and gF and the fact that the higher-order latent variable was related to WMC and gF, variance partitioning (e.g., Chuah & Maybery, 1999) was used to examine the relations among the recall components and WMC and gF. Variance partitioning attempts to allocate the overall R^2 of a particular criterion variable (here, WMC

and gF) into portions that are shared and unique to a set of predictor variables (here, recall accuracy, recall latency, and intrusions). A series of regression analyses was carried out to obtain R^2 values from different combinations of the predictor variables (see Table 4) in order to partition the variance. For each variable entering into the regression, the latent correlations from the previous CFA were used. Separate analyses were carried out for WMC and gF to determine how the three recall components would account for both shared and unique variance in each of the criterion variables (WMC and gF).

As is shown in Figure 3A, the results suggest that 29% of the variance in WMC was accounted for by the three recall components. Specifically, 4% was shared by all three components, whereas 1%–3% was shared between two of the components independently of the third component. For instance, 3% was shared between recall accuracy and recall latency independently of intrusions. Finally, 7% was uniquely accounted for by recall accuracy, and 12% was uniquely accounted for by recall latency. This suggests that these three components (recall accuracy, recall latency, and intrusions) represent somewhat distinct recall components, with both recall accuracy and recall latency accounting for independent variance in WMC. The WMC variance accounted for by intrusions, however, was largely due to shared variance with both recall accuracy and recall latency. This suggests that WMC is independently related to both differences in recovery and differences in overall search set size. Differences in editing out intrusions (or monitoring abilities) were not uniquely related to WMC, suggesting that these abilities are based, in part, on differences in recovery and overall search set size. Overall, these results are consistent with previous work suggesting that these components are in many ways distinct, with some variables affecting one component while not affecting another component (e.g., Rohrer & Wixted, 1994).

Table 4
 R^2 Values for Regression Analyses Predicting Working Memory Capacity (WMC) and General Fluid Intelligence (gF) for Various Predictor Variables

Predictor Variables	R^2	F
WMC		
Recall accuracy, recall latency, intrusions	.29	24.31
Recall accuracy, recall latency	.29	36.59
Recall accuracy, intrusions	.17	18.93
Recall latency, intrusions	.22	25.70
Recall accuracy	.15	33.01
Recall latency	.21	49.38
Intrusions	.07	13.34
gF		
Recall accuracy, recall latency, intrusions	.31	27.24
Recall accuracy, recall latency	.30	39.45
Recall accuracy, intrusions	.22	25.43
Recall latency, intrusions	.24	28.63
Recall accuracy	.18	39.41
Recall latency	.21	49.38
Intrusions	.11	22.48

Note—All R^2 values are significant at $p < .01$.

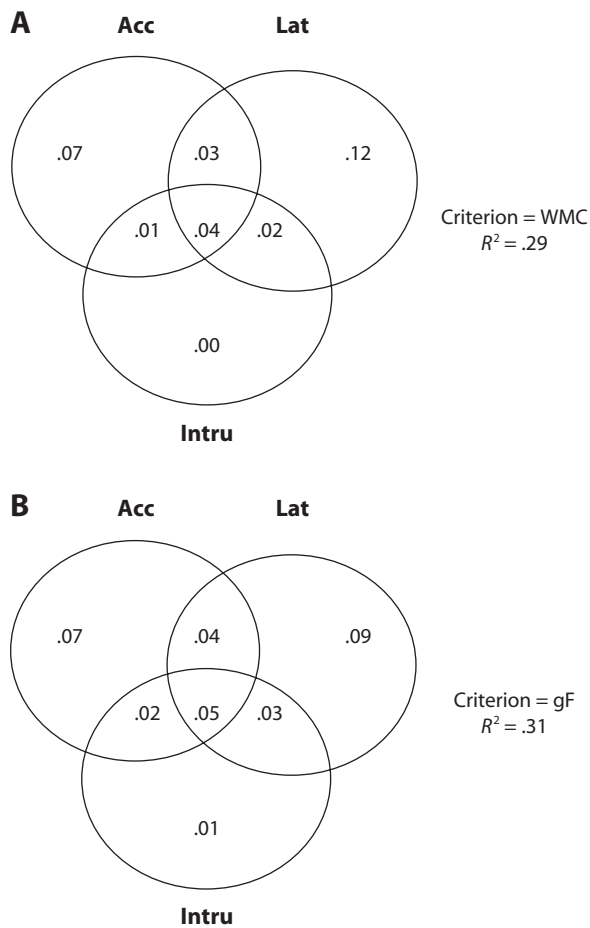


Figure 3. (A) Venn diagrams indicating the shared and unique variance accounted for in WMC by recall accuracy (Acc), recall latency (Lat), and intrusions (Intru). (B) Venn diagrams indicating the shared and unique variance accounted for in gF by Acc, Lat, and Intru. Numbers are based on regressions from Table 4.

As is shown in Figure 3B, similar results were obtained when examining gF as the criterion variable. Specifically, 31% of the variance in gF was accounted for, with 5% accounted for by the shared variance across all three components, 2%–4% accounted for by the shared variance across two of the components, and 7%–9% uniquely accounted for by either recall accuracy or recall latency. Thus, similar to those of the WMC analyses, these results suggest that differences in recovery abilities and differences in overall search set size (as indicated by recall accuracy and recall latency, respectively) are independently related to gF, whereas differences in editing out intrusions was not.

Cluster Analysis

In the final set of analyses, we utilized cluster analytic techniques to determine whether subgroups of participants based on differences in the three component processes were present in the data. Specifically, as was previously noted, it is possible that some low-ability participants have specific problems related to sampling in which some low-ability participants have deficits in focusing the search

set, whereas others have deficits in terms of the number of degraded representations in their search sets, or in the efficiency of their monitoring/editing processes. In order to examine the possibility of subgroups of participants who have specific deficits in one component, rather than global deficits manifested on all components, a cluster analysis was used. Cluster analysis is a tool used to determine group membership by minimizing within-group differences and maximizing between-groups differences (Everitt, Landau, & Leese, 2001; Kaufman & Rousseeuw, 1990). Groups are formed where individuals in the group are very similar to one another but unlike individuals in other groups. It should be noted that these methods are largely atheoretical, and group membership is merely based on similarities within a cluster and differences across clusters.

In order to examine possible subgroups in the three recall component processes, composites for recall accuracy, recall latency, and intrusions were formed on the basis of the three recall tasks. Specifically, recall accuracy in each of the three recall tasks were averaged together to form the recall accuracy composite. The same was done for recall latency and intrusions. Next, the three composites were entered into a two-step cluster analysis. The two-step cluster analysis was performed using SPSS Version 16. In this analysis, cases were first grouped into preclusters at the first step by constructing a cluster feature tree (see Zhang, Ramakrishnan, & Livny, 1996). For each case, the algorithm determined whether the case should be included with a previously formed precluster or whether a new precluster should be created on the basis of the cluster feature tree. In the second stage, an agglomerative hierarchical clustering method was used on the preclusters and allowed for an exploration of different numbers of clusters. In this stage, clusters were recursively merged until the desired number of clusters was determined. In these analyses, the distance between clusters was based on a log-likelihood measure, whereby distance was related to the decrease in log likelihood as the clusters were formed into a single cluster.

The cluster analysis specified four groups in the data, consisting of 74, 45, 46, and 21 participants each. Shown in Table 5 are the composite scores for each of the four groups, as well as the overall average scores collapsed across groups. Corresponding *F* values from the omnibus ANOVAs for each measure are also presented. Bonferroni post hoc comparisons suggested that there were significant differences (all $ps < .01$) between all of the groups on recall accuracy (except for Groups 2 and 3, which did not differ [$p > .78$]). There were significant differences (all $ps < .01$) between all of the groups on recall latency (except for Groups 1 and 4, which did not differ [$p > .11$]). There were also significant differences between all of the groups on intrusions (all $ps < .05$). Importantly, the pattern of results across the three recall components suggested that some of the groups demonstrated specific deficits on one component rather than global deficits. Specifically, Group 1 seemed to be composed of relatively good performers, who recalled the most target items, had a shorter than average recall latency, and recalled few intrusions. Group 2 was composed of individuals who seemed to have deficits primarily in focusing their

Table 5
Descriptive Statistics and Omnibus ANOVA Results for Each Group Defined by the Cluster Analysis and Overall Performance Collapsed Across Groups

Measure	Group 1		Group 2		Group 3		Group 4		Overall		<i>F</i>	η^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
RecallAcc	.59	.07	.46	.06	.45	.05	.51	.05	.51	.08	69.1	.53
RecallLat	14,424	1,890	18,278	1,761	13,410	1,667	15,473	1,702	15,224	2,546	64.8	.52
RecallIntru	2.15	1.19	4.40	2.11	3.09	1.62	8.75	3.07	3.67	2.71	75.4	.55
WMC	.27	.61	-.28	.91	.18	.79	-.22	.84	.00	.80	6.2	.09
gF	.25	.71	-.20	.72	.03	.67	-.11	.63	.00	.72	4.3	.07

Note—RecallAcc, proportion correct; RecallLat, recall latency; RecallIntru, average number of intrusions; WMC, *z* score composite of the three working memory capacity tasks; gF, *z* score composite of the three general fluid intelligence measures.

search sets, which resulted in the longest recall latencies, poor recall performance, and greater than average number of intrusions. Group 3 was composed of individuals who seemed to have deficits primarily in recovery processes, as indicated by their low recall scores, but shorter than average recall latency and fewer than average intrusions. Finally, Group 4 was composed of individuals who seemed to have deficits primarily in monitoring abilities, as indicated by a larger number of intrusions, but average recall latency and less than average recall.

Furthermore, these groups tended to differ in their levels of WMC and gF. As is shown in Table 5, there were significant differences between the groups in a *z* score composite of WMC. Specifically, Group 1 had higher WMC scores than either Group 2 or Group 4 (both $ps < .05$), and Group 2 had lower WMC scores than Group 1 or Group 3 (both $ps < .05$). Thus, not only did the groups differ with respect to the recall components, but the groups also differed with respect to WMC. In particular, Group 1 had higher WMC scores, recalled more items, recalled fewer intrusions, and recalled at a faster rate than Group 2, consistent with the notion that high- and low-WMC individuals differ primarily in the number of irrelevant items within their search sets (e.g., Unsworth, 2007). In addition, the present results suggested that other low-ability participants seem to have deficits not in the size of their search sets, but rather in the number of recoverable targets within their search sets (Group 3) or in the effectiveness of their monitoring component (Group 4). In terms of gF, Group 1 had higher gF scores than either Group 2 or Group 4 (both $ps < .05$). None of the other groups differed significantly from one another with respect to gF. Overall, these results suggest that there are subgroups of individuals who have specific deficits in recall because of problems in focusing the search set, recovering degraded items from the search set, or monitoring and editing errors.

GENERAL DISCUSSION

The present study examined the dynamics of free recall in a large sample of participants. Measures of recall accuracy, recall latency, and intrusions were extracted from three different free recall tasks. This is the first study in which recall accuracy, recall latency, and intrusions were simultaneously examined in the same study, with multiple measures of each. As such, in the present study, we extended prior work by demonstrating a number of important findings. For instance, the results suggest that measures from the

different tasks loaded onto their respective factors in both the exploratory and the confirmatory factor analyses. That is, accuracy scores across the three recall tasks tended to group together, as did recall latency and intrusions. This is important because, although previous work has shown that recall accuracy across tasks is related, no study has shown that recall latency and intrusions across different tasks are related and can be accounted for by a single construct. This provides strong evidence for the systematic nature of recall latency and intrusions across various free recall tasks.

The results also suggest that the three recall factors in both the exploratory and the confirmatory factor analyses were related to one another and were related in the expected direction. Specifically, recall accuracy was negatively related to both recall latency and intrusions, and recall latency and intrusions were positively related. Thus, although the results of prior experimental work has suggested that the different components (especially recall accuracy and recall latency) can be dissociated, the present results suggest that the components are related at an individual differences level. This points to the interrelated nature of these components, in that individuals who recall few correct items also tend to recall at a slower rate and to recall more intrusions than individuals who recall many correct items. Indeed, it was possible to extract a higher-order factor accounting for the shared variance from the three recall factors. As was previously noted, this result suggests that poorer performers may be searching through a larger set of items than good performers. At the same time, the different components also accounted for unique variance in the cognitive ability constructs, and the cluster analysis suggested that there were subgroups within the data with very specific deficits. Thus, at a broad level, the three components are interrelated but are also somewhat distinct.

In terms of WMC, the results suggest that each of the components were related to individual differences in WMC, consistent with prior work (e.g., Unsworth, 2007). Specifically, high-WMC individuals recalled a higher proportion of correct items, recalled at a faster rate, and recalled fewer intrusions than low-WMC individuals. Furthermore, the higher-order recall factor was strongly related to WMC. These results are consistent with the notion that low-WMC individuals are poorer at focusing their search sets on only the current list items and thus search through a larger set of items (including more intrusions) than high-WMC individuals do. These results are inconsistent with the other three possibilities discussed, because

those possibilities suggest either no differences in recall latency (low ability nonrecoverable, low ability monitoring) or a positive correlation between ability and recall latency (low ability small). Importantly, these results extend previous work by demonstrating that these effects are apparent on several free recall tasks and occur when examining the full range of participants, rather than merely looking only at high- and low-WMC individuals. Thus, the present results provide strong supporting evidence for the notion that individual differences in WMC are partially due to differences in search set size. In addition, the variance partitioning analyses suggested that both recall accuracy and recall latency were independent related to WMC, but the number of intrusions was related to WMC only through shared variance with recall accuracy and recall latency. Thus, despite the fact that each of the components was related to one another, the components were also somewhat distinct (see also Rohrer & Wixted, 1994). This suggests that WMC is related to both overall search set size and recovery independent of differences in monitoring processes.

Similar results were found for gF, suggesting that high-gF participants were better at focusing their search on only relevant target items than were low-gF participants. Thus, the substantial correlation between measures of recall and aspects of intelligence that has been found previously (e.g., Beier & Ackerman, 2004; Carroll, 1993) is partially due to the fact that low-gF participants are poorer at using cues to focus their search sets to the same extent as high-gF participants. In addition, the variance partitioning results suggested that the ability to recover degraded representations independently of differences in search set size was also an important predictor of gF.

Finally, the cluster analysis results suggest that not all participants could be classified in terms of overall deficits. Rather, these results suggest that there were subgroups of participants within the data, with each group having specific deficits in search set size, recovery, or monitoring. Specifically, one group was composed of high-ability participants who recalled a high proportion of items, recalled at a fast rate, and emitted few intrusions. Another group was composed of individuals who recalled fewer items, recalled at a slower rate, and recalled more intrusions than average, consistent with deficits in search set size. Additionally, another group was composed of individuals who recalled few correct items, but recalled at a normal rate and had few intrusions, consistent with deficits in recovery. Finally, the last group of participants recalled few correct target items and recalled at a normal rate, but emitted far more intrusions than any of the other groups, consistent with deficits in monitoring. Furthermore, the groups tended to differ in baseline WMC and gF, with some of the groups having lower WMC and gF scores than the other groups. This suggests that there are likely multiple subgroups of low-ability individuals. Some low-ability individuals seem to have problems focusing their search sets, whereas others have problems in monitoring the products of the search process.

Overall, the results of the present study demonstrate that search models of free recall can account for individual differences in free recall. In particular, the results are consistent with prior work suggesting that individual differ-

ences in WMC and the dynamics of recall are due in part to differences in the size of the set of items that individuals search through (Unsworth, 2007). In this framework, it is argued that low-WMC (and low-gF) individuals are less able to focus their search on only the correct targets items and, instead, include many irrelevant representations (intrusions) in their search sets relative to high-WMC (and high-gF) individuals. This results in a lower probability of selecting target items, a slower rate of recall, and a higher incidence of intrusion errors. Furthermore, this framework predicts that these effects should occur on a number of free recall tasks. As was noted above, the results from the present study are very much in line with these overall predictions and provide strong support for the view of differences in search set size outlined in previous work (Unsworth, 2007). Thus, this view suggests that low-capacity (low-WMC) individuals actually activate more representations than high-ability individuals, but some of these representations are irrelevant to the task at hand. Thus, here, *capacity* refers to the ability to filter out irrelevant information, rather than to how much information can be maintained (see also Vogel et al., 2005).

At the same time, however, the results go beyond prior results by suggesting that there are not only deficits in overall search set size, but there are also deficits in the number of degraded targets within the search set, as well as deficits in monitoring abilities. Thus, the present results suggest that the prior model (Unsworth, 2007), in which differences were due exclusively to search set size, was too simplistic, and current models must account for differences in search set size, as well as differences in the strength of representations and monitoring abilities. Thus, rather than localizing the effects to a single component of recall, the present results suggest that differences arise because of a mixture of processes, with some individuals having deficits in search set size, others having deficits in the recovery of degraded representations, and others having deficits in monitoring the products of retrieval for possible errors. Each of these seems to be important for individual differences in WMC and gF. In order to have a fuller understanding of the dynamics of free recall, as well as individual differences in free recall, an examination of all components is needed. Future work should be devoted to examining not only variation in search set size, but also variation in recovery and monitoring processes. Combining explicit mathematical process models of free recall (e.g., Raaijmakers & Shiffrin, 1980) with individual differences models should provide a powerful framework to better understand the dynamics of recall.

AUTHOR NOTE

Thanks to Greg Spillers and Gene Brewer for comments on an earlier version of the article. Correspondence concerning this article should be sent to N. Unsworth, Department of Psychology, University of Georgia, Athens, GA 30602-3013 (e-mail: nunsworth@uga.edu).

REFERENCES

- BEIER, M. E., & ACKERMAN, P. L. (2004). A reappraisal of the relationship between span memory and intelligence via "best evidence synthesis." *Intelligence*, *32*, 607-619.

- BERGER, F. R., GUPTA, W. B., BERGER, R. M., & SKINNER, J. (1990). *Air Force officer qualifying test (AFOQT) Form P: Test manual* (AFHRL-TR-89-56). Brooks Air Force Base, TX: Manpower and Personnel Research Division, Air Force Human Resources Laboratory.
- BORS, D. A., & FORRIN, B. (1995). Age, speed of information processing, recall, and fluid intelligence. *Intelligence*, **20**, 229-248.
- BOUSFIELD, W. A., SEDGEWICK, C. H. W., & COHEN, B. H. (1954). Certain temporal characteristics of the recall of verbal associates. *American Journal of Psychology*, **67**, 111-118.
- CANTOR, J., & ENGLE, R. W. (1993). Working memory capacity as long-term memory activation: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1101-1114.
- CARROLL, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- CHUAH, Y. M. L., & MAYBERY, M. T. (1999). Verbal and spatial short-term memory: Common sources of developmental change? *Journal of Experimental Child Psychology*, **73**, 7-44.
- 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.
- CRAIK, F. I. M. (1968). Types of error in free recall. *Psychonomic Science*, **10**, 353-354.
- CROWDER, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- 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.
- EVERITT, B. S., LANDAU, S., & LEESE, M. (2001). *Cluster analysis*. London: Arnold.
- JOHNSON, D. M., JOHNSON, R. C., & MARK, A. L. (1951). A mathematical analysis of verbal fluency. *Journal of General Psychology*, **44**, 121-128.
- JUST, M. A., & CARPENTER, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, **99**, 122-149.
- KANE, M. J., HAMBRICK, D. Z., TUHOLSKI, S. W., WILHELM, O., PAYNE, T. W., & ENGLE, R. W. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, **133**, 189-217.
- KAPLAN, I. T., CARVELLAS, T., & METLAY, W. (1969). Searching for words in letter sets of varying size. *Journal of Experimental Psychology*, **82**, 377-380.
- KAUFMAN, L., & ROUSSEEUW, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.
- KLINE, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.
- MACLEOD, C. M., & NELSON, T. O. (1984). Response latency and response accuracy as measures of memory. *Acta Psychologica*, **57**, 215-235.
- MCGILL, W. J. (1963). Stochastic latency mechanism. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 309-360). New York: Wiley.
- MURDOCK, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.
- PREACHER, K. J., RUCKER, D. D., MACCALLUM, R. C., & NICEWANDER, W. A. (2005). Use of extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods*, **10**, 178-192.
- RAAIJMAKERS, J. G. W., & SHIFFRIN, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207-262). New York: Academic Press.
- RAAIJMAKERS, J. G. W., & SHIFFRIN, R. M. (1992). Models for recall and recognition. *Annual Review of Psychology*, **43**, 205-234.
- RAVEN, J. C., RAVEN, J. E., & COURT, J. H. (1998). *Progressive matrices*. Oxford: Oxford Psychologists Press.
- ROHRER, D. (1996). On the relative and absolute strength of a memory trace. *Memory & Cognition*, **24**, 188-201.
- ROHRER, D., SALMON, D. P., WIXTED, J. T., & PAULSEN, J. S. (1999). The disparate effects of Alzheimer's disease and Huntington's disease on semantic memory. *Neuropsychology*, **13**, 381-388.
- ROHRER, D., & WIXTED, J. T. (1994). An analysis of latency and inter-response time in free recall. *Memory & Cognition*, **22**, 511-524.
- SCHEIRER, C. J. (1971). Effect of cueing, modality, and effective contiguous time on response latency in short-term memory. *Journal of Experimental Psychology*, **88**, 429-432.
- SHIFFRIN, R. M. (1970a). Forgetting: Trace erosion or retrieval failure? *Science*, **168**, 1601-1603.
- SHIFFRIN, R. M. (1970b). Memory search. In D. A. Norman (Ed.), *Models of human memory* (pp. 375-447). New York: Academic Press.
- THURSTONE, T. G. (1962). *Primary mental abilities*. Chicago: Science Research Associates.
- TULVING, E. (1968). Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 2-36). Englewood Cliffs, NJ: Prentice Hall.
- 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, & Cognition*, **33**, 1020-1034.
- UNSWORTH, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory & Language*, **59**, 223-236.
- UNSWORTH, N. (2009). Examining variation in working memory capacity and retrieval in cued recall. *Memory*, **17**, 386-396.
- 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.
- 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.
- VOGEL, E. K., MCCOLLOUGH, A. W., & MACHIZAWA, M. G. (2005). Neural measures reveal individual differences in controlling access to visual working memory. *Nature*, **438**, 500-503.
- WARD, G., & TAN, L. (2004). The effect of the length of to-be-remembered lists and intervening lists on free recall: A re-examination using overt rehearsal. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 1196-1210.
- WINGFIELD, A., LINDFIELD, K. C., & KAHANA, M. J. (1998). Adult age differences in the temporal characteristics of category free recall. *Psychology & Aging*, **13**, 256-266.
- WIXTED, J. T., & ROHRER, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1024-1039.
- WIXTED, J. T., & ROHRER, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, **1**, 89-106.
- ZAROMB, F. M., HOWARD, M. W., DOLAN, E. D., SIROTTIN, Y. B., TULLY, M., WINGFIELD, A., & KAHANA, M. J. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 792-804.
- ZHANG, T., RAMAKRISHNAN, R., & LIVNY, M. (1996). BIRCH: An efficient data clustering method for very large databases. In *Proceedings of the ACM SIGMOD Conference on Management of Data*. Montreal: ACM.

NOTE

1. Prior to performing the free recall tasks, all of the participants completed a brief typing exercise (typing the words *one* to *ten*) to estimate their typing efficiency. Typing speed was weakly related to recall latency ($r = .19$) but was not related to recall accuracy, intrusions, WMC, or gF (all r s < .09). Importantly, partialing out typing speed did not change any of the results. Thus, individual differences in typing speed did not account for any of the results of the present study.

(Manuscript received November 18, 2008;
revision accepted for publication March 26, 2009.)