Roles of working memory capacity and long-term working memory skill in complex task performance

YOUNG WOO SOHN
Yonsei University, Seoul, Korea

and

STEPHANIE M. DOANE
Mississippi State University, Mississippi State, Mississippi

In the present research, we examined the relative roles of domain-general and domain-specific individual difference characteristics in complex cognitive task performance. Specifically, we examined the impact both of working memory (WM) capacity and of acquired skills used to encode presented information in an accessible form in long-term working memory (LTWM) on performance in a complex aviation task environment. Measures of WM capacity and LTWM skill served as performance predictors. A criterion measure of task performance was related to the predictor measures. The results indicated that an increase in LTWM skill decreases the role of WM capacity as the determinant of complex task performance, although both measures are important performance predictors. We discuss how the two distinct WM constructs coexist and interact to support complex task performance.

The importance of working memory (WM) has been demonstrated in various domains of complex cognition, such as physics (e.g., Larkin, McDermott, Simon, & Simon, 1980), language comprehension (e.g., Just & Carpenter, 1992; Kintsch, 1998), medicine (e.g., Groen & Patel, 1988), and computer problem solving (e.g., Anderson & Jeffries, 1985; Doane, McNamara, Kintsch, Polson, & Clawson, 1992; Doane, Sohn, McNamara, & Adams, 2000; Sohn & Doane, 1997). Individual differences in performance on such complex tasks are explained, in large part, by the ability to satisfy the WM demands that a given task imposes. When task demands exceed WM capacity, task performance is degraded. Of interest for the present research is the locus of the varied ability to satisfy the WM demands.

There are at least two contrasting loci that have been related to individual differences in complex task performance. The first is the capacity to compute and maintain previously presented information in short-term memory (STM; see, e.g., Baddeley, 1986; Daneman & Carpenter, 1980; Engle, Kane, & Tuholski, 1999; Shah & Miyake, 1996). The second locus is the acquired skill to encode the presented information in accessible form in long-term memory (LTM; see, e.g., Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995; Kintsch, 1998). In this view, the accessibility to the LTM information relevant to the given task is expected to functionally increase WM capacity. In the present research, we tested the validity of these two accounts in explaining and predicting individual differences in complex cognitive task performance.

Capacity-Based Account

Capacity, an important aspect of WM, refers to a limited amount of resources available for allocation to processing and storage functions (e.g., Baddeley & Hitch, 1974; Daneman & Carpenter, 1980; Hambrick & Engle, 2002; Radvansky & Copeland, 2001; Shah & Miyake, 1996). The amount of resources available varies between individuals and is thought to be the locus of individual differences in speed and accuracy performance on WM-demanding tasks. This account proposes that performance differences among individuals within a task domain can be explained, in large part, in terms of WM capacity. When task demands are high enough to strain capacity, individuals with a lesser capacity are predicted to be less able to process information in a timely fashion or to store intermediate products.

Individual differences in WM capacity are often measured with span tasks performed in a dual-task situation (e.g., Daneman & Carpenter, 1980; Shah & Miyake, 1996). Such tasks require participants to keep track of presented information while simultaneously doing mental operations. The capacity measures, based primarily on the retention of the information over time, have been shown to predict performance in a variety of cognitive tasks, such as text comprehension (Just & Carpenter,

1992), following directions (Engle, Carullo, & Collins, 1991), analytic computation (Carpenter, Just, & Shell, 1990), spatial visualization (Shah & Miyake, 1996), arithmetic (Adams & Hitch, 1997), reading graphs (Lohse, 1997), and solving logic problems (Kyllonen & Stephens, 1990).

Skill-Based Account

As an alternative to the fixed capacity theory, some aspects of WM are viewed as practiced skills (Chase & Ericsson, 1982; Ericsson & Kintsch, 1995). In this view, WM capacity is determined by an individual's skills in efficiently accessing task-relevant information in LTM. In contrast to capacity theory, the amount of information that can be maintained in accessible form in WM is not limited by a fixed capacity. Instead, WM capacity is thought to dynamically change within individuals as a function of their long-term access efficiency during a given task. This account of WM is referred to as long-term WM (LTWM) because it postulates a mechanism for extending WM that requires skilled use of both storage in and retrieval from LTM.

According to Ericsson and Kintsch (1995), LTWM is characterized as being mediated by retrieval structures used to encode and store information in LTM. Retrieval structures are composed of a set of retrieval cues that make relevant information in LTM more available. People can acquire retrieval structures through extensive knowledge built up from experience in a particular domain and can use them to overcome the capacity limits of WM. There is a substantial amount of evidence to support the construction and use of retrieval structures, including the recall of chess positions by expert chess players (Ericsson & Staszewski, 1989; Saariluoma, 1989), memory for meal orders in a restaurant (Ericsson & Polson, 1988), recall of patient information in medical diagnosis (Groen & Patel, 1988), mental abacus calculation (Hatano, Amaiwa, & Shimizu, 1987), and mental multiplication (Staszewski, 1988).

LTWM theory provides a compelling account of individual differences in WM capacity. The theorists explain differences in capacity measures (e.g., reading spans) through mechanisms that mediate LTWM (Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995; Kintsch, 1998). They suggest that capacity measures can predict performance differences because they reflect differences in the amount of relevant knowledge and the structure of encoding skills. In the LTWM account, high-span individuals are those with appropriate LTWM skills for efficient storage in LTM. Thus, LTWM theory proposes that WM capacity can be related to LTWM skills, whereas capacity theory states that these two constructs are distinct.

In summary, the capacity-based WM theory and the skill-based LTWM theory differ in their views about storage location, as well as about factors influencing individual differences in processing and storage. According to capacity theory, the maintenance of information in the span task is based solely on storage in STM, and the locus of individual performance differences that have been found in various domains is the domain-general capacity to maintain the information in STM. This capacity is assumed to be inherently different across individuals. In contrast, according to LTWM theory, the maintenance of information is based on storage in LTM in accessible form, and individual performance differences result from the acquired domain-specific skill in constructing better retrieval structures.

THE PRESENT RESEARCH

Research Objective

The objective of the present research was to investigate the respective roles of WM capacity and acquired LTWM skills in complex cognitive task performance. The preceding discussion suggests that both WM capacity and LTWM skills may contribute to the performance of complex cognitive tasks. What is not clear is whether the two constructs interact with each other and, if so, how. The present research addressed this question by comparing the ability of measures of WM capacity and LTWM skill to predict individual performance in an experimental context in which both measures assessed the cognitive resources required to complete the experimental task. By making this comparison, we examined whether individual differences in complex task performance were due to differences in WM capacity, acquired LTWM skills, or both. The specific question of interest was whether LTWM skills would reduce the importance of WM capacity during complex cognitive task performance. This interplay between WM capacity and LTWM skills has been hypothesized (Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995) but not empirically evaluated (but see Hambrick & Engle, 2002). The present research tests this important hypothesis and furthers our understanding of the respective roles of two memory mechanisms in complex task performance.

Situation Awareness as a Complex Cognitive Task

As was previously stated, we wanted to study performance on a task that imposed a heavy load on WM. Flight situation awareness (SA) meets this requirement. Although there has been no consensus regarding the definition of SA in the aviation psychology literature, most definitions cited can be incorporated into Endsley's (1995) information-processing view (Durso & Gronlund, 1999). Endsley's view defines three levels of SA in terms of component cognitive processes. The first level of SA involves the perceiving of elements in the environment, such as other aircraft, terrain, system status, and warning lights. The second level involves information integration, which structures the elements of the perceived situation into meaningful and recognizable configurations by activating LTM knowledge structures. The third level includes the projection of future flight status. This third level of SA uses the goal-relevant activated knowledge

structures formed in the second level of SA to predict the status of the aircraft. In Endsley's view, then, SA involves all the cognitive processes related to a complex cognitive activity.

The content relevant to SA is processed by and resides in WM. The accuracy of SA depends on the activation of knowledge structures in LTM that are capable of integrating incoming information into a situation model of the flight (Doane & Sohn, 2000; Endsley, English, & Sundararajan, 1997; Morrow, Menard, Stine-Morrow, Teller, & Bryant, 2001). The situation model is, in large part, organized spatially, and SA taps into the situation model level of comprehension (Endsley & Bolstad, 1994; Morrow et al., 2001). Understanding flight information from cockpit instruments and using this information to update a situation model should impose heavy demands on pilots' WM (Morrow & Rodvold, 1998). Thus, the complex cognitive processes involved in maintaining SA will either facilitate or detract from pilot performance.

The task posed to the participants in the present research was designed to reflect their ability to perform the three-level process of SA (Endsley, 1995). Important for the purposes of the present research, this task was presumed to impose both processing and storage demands on WM. The participants had to maintain the goal and the situational elements depicted in the cockpit snapshots in WM and then relate these elements to the situation that would be expected if the goal were to be achieved (details of task materials and procedures are provided in the Method section). Hence, this SA task was considered appropriate for studying the role of WM in the performance of a complex cognitive task.

Overview of the Experiment

To restate, the major questions addressed in this research were whether WM capacity, LTWM skill, or both would contribute to complex cognitive task performance and whether the two WM constructs would interact with each other in performance of a complex task and, if so, how. To address these questions, the present experiment consisted of three tasks. First, the span tasks were used to measure domain-general WM capacity. Many researchers hypothesize separate WM resources for different modalities of cognitive processes (e.g., Daneman & Tardif, 1987; Shah & Miyake, 1996). On the basis of this hypothesis, the span task assessed individual WM capacity for the computation and storage of spatial and verbal information.

Second, a situation recall task analogous to Chase and Simon's (1973) chess experiment was devised to measure individual pilots' domain-specific LTWM skills. In the present experiment, the elements to be reconstructed were cockpit situational indices (e.g., altitude, heading, airspeed), instead of chess pieces. To measure encoding in and retrieval from LTM, participant memory for a cockpit situation was tested after a 30-sec delay filled with an intervening task (see, e.g., Peterson & Peterson, 1959). The intervening task was considered to clear

short-term storage of scanned situational information. To measure the construction of retrieval structures, we presented the participants with both meaningful and nonmeaningful cockpit situations. Meaningfulness was defined as whether the cockpit situation represented a plausible state of an aircraft in the real world.

The difference in the reconstructions of meaningful and nonmeaningful situations is thought to reflect pilot ability to construct and use retrieval structures in LTM (see Stokes, Kemper, & Kite, 1997). In the meaningful trials, the pilots were expected to access their LTM retrieval structures to encode the situational information. In contrast, in the nonmeaningful trials, the pilots were not expected to have a matching retrieval structure that could be accessed to aid encoding. The benefit of accessing a LTM retrieval structure during encoding was calculated by comparing delayed-recall accuracy for meaningful situations with that for nonmeaningful situations as an index of the situational information accessible in LTWM.

Finally, we obtained criterion measures of complex task performance by asking pilots to perform the aforementioned SA tasks. The methods and results of the three tasks will be detailed in the subsequent sections.

METHOD

Participants

Fifty pilots from the University of Illinois and the University of Connecticut participated in this experiment. All the participants were paid $25 for their participation. Twenty-five student pilots (novices) and 25 flight instructors (experts) had, on average, 60 h ($SD = 70$) and 907 h ($SD = 1,501$) of flight time, respectively.

Materials and Procedure

After signing an informed consent form, the participants completed the flight background questionnaire. Then they completed the three tasks (span, situation recall, and SA), administered on a Macintosh G4 computer using a PsyScope program in the following order: the span task, the situation recall task, and the SA task. The experiment took place in a single session of 2–2.5 h.

Span task. The span task consisted of two parts, spatial and verbal. In the first part (spatial), the participants were presented with a set of English capital letters (F, J, L, P, and R) and their mirror images one at a time, each appearing in different orientations. The task was to remember the orientation of each letter in the correct order, while deciding whether each letter was normal or its mirror image as quickly and accurately as possible. Each letter was presented for 2,200 msec in one of seven possible orientations in 45º increments, excluding the upright orientation. The participants were asked to respond aloud to indicate whether the orientation was normal or a mirror image. After the entire set of letters had been presented in a trial, the participants were asked to report the correct orientations in serial order by clicking on small buttons that indicated the possible orientations (for more details, see Shah & Miyake, 1996).

The structure of the second part (verbal) was identical to that for the spatial part, except that the participants were presented with a set of different letters and were asked to recall the letters, rather than the orientations. We added two letters (G and Q) to the letter set (F, J, L, P, or R) that was used for the spatial span task in order to increase the variability of the possible set of letters to be recalled. Letter presentation was constrained such that the same letter could appear only once in a given trial. The participant's task was to re-

member the letters and their order of presentation while deciding whether each letter was normal or a mirror image. At the end of a trial, the participant was prompted to type the letters on a keyboard in the order presented. This task is unique, as compared with existing verbal span tasks (e.g., the reading span task), since it creates a dual-task condition of spatial processing and verbal storage, whereas the others create a condition of verbal (or math) processing and verbal storage. The present task was particularly designed to tap into the general ability used in analogous tasks involving the reading of displays with different orientations of values and figures. Although this task is not one of the conventional verbal span tasks, it provides a measure of reading span that is more appropriate when one is required to interact with a display-based system.

Each part of the span task included a total of 20 letter sets (5 sets at each size, ranging from two to five letters), and the participants were presented with increasingly longer sets of letters. After the participants had finished five trials with a letter set of a particular size, they were informed of how many letters would appear in the next set of trials.

Scoring of the spatial span was done in the same manner as in Shah and Miyake (1996). For the scoring of verbal span task, however, a higher criterion was used, because this task was less difficult than the spatial task. A participant's verbal span score was defined as the highest level of set size for which all of the letters were recalled in the correct order for at least four, rather than three (used for the spatial span scoring), of the five sets. Although the scoring criteria differed across span tasks, due to the variance in their difficulty, they were applied equitably across participants and resulted in comparable span scores.

Situation recall task. In this task, the participants were presented with a pair of consecutive cockpit snapshots for 40 sec and then completed an intervening task for a 30-sec delay period, during which they counted backward by threes as fast as possible. For example, a prompt such as "Count backward by 3: 528" was presented on the screen following the disappearance of the cockpit snapshots, and the participants counted backward aloud by threes from 528: 525, 522, 519, and so on. This intervening task was used to interrupt maintenance and computation of the display indications in WM. After the 30-sec delay filled with the intervening task, they had to recall the flight situation and manually fill in, on a sheet of paper, the blank indications of display instruments shown within a cockpit frame. The participants were asked to reconstruct the value indications that were displayed on one of the two cockpit snapshots, and the choice of snapshot to be recalled was randomly selected. They were allowed as much time as they wished to fill in the indications.

As was previously stated, we manipulated the meaningfulness of the cockpit situation. For the meaningful situations, each snapshot displayed a routine flight situation, and the pair of snapshots in sequence depicted consecutive routine flight situations. For the non-meaningful situations, each snapshot displayed an implausible flight situation that pilots would never encounter, and the pair of snapshots did not depict consecutive flight situations.[1] The situation recall task included a total of nine trials: six trials with meaningful situations and three trials with nonmeaningful situations. The participants were first presented with a set of trials with meaningful situations and then with a set of trials with nonmeaningful situations. This was done to avoid any possible changes in viewing strategy that might be caused by exposure to unexpected nonmeaningful situations. We gave the participants two practice trials with meaningful situations in order to familiarize them with the procedure. They received feedback regarding recall accuracy for the practice trials. No accuracy feedback was provided for the experimental trials.

Participant indications for each of the eight flight elements (i.e., airspeed, heading, altitude, pitch, bank, power, rate of climb, and rate of turn) indicated by the cockpit instruments were scored by the experimenter according to explicit rules. Responses were scored as correct and full points were given if they matched exactly all the corresponding indications. The responses for pitch, bank, rate of climb, and rate of turn, which provide both directional and value indications, were given partial credits if only the directional indications were correct. However, no credit was given to these responses if only their value indications were correct, because value information in the wrong direction would indicate a more severe loss of memory for the situation than would memory loss for the exact value but retention of the correct direction.

SA task. The participants viewed consecutive screens that showed a goal description and two consecutive cockpits and then judged whether an aircraft depicted by the consecutive cockpit snapshots would reach the specified goal state in the next 5 sec. The first screen displayed the goal description at the top of the screen and, immediately below it, the cockpit snapshot that depicted the initial flight situation (at Time 1) for 20 sec, at which time the first cockpit snapshot disappeared and the second cockpit snapshot appeared. The goal description indicated the desired state that an aircraft should reach in the near future (i.e., in approximately 5 sec) with respect to the three flight performance elements (e.g., "Altitude 3,500 ft, Heading 180°, Airspeed 90 kt"). The second snapshot depicted the "current" state of the aircraft (at Time 2) following changes caused by control movements (not specified to the participants) executed at Time 1. This second snapshot depicted the state of the aircraft approximately 5 sec following the status depicted in the first snapshot, and the goal description remained at the top of the screen.

The participants were asked to determine whether the aircraft depicted in the cockpit snapshots would achieve the goal initially specified in the next 5 sec without further control movements. In essence, they had to mentally predict the state of the aircraft in the next 5 sec, with the constraint that no further control movements would be applied. The participants indicated whether the aircraft was moving in a manner consistent with achieving the goal or in a manner inconsistent with achieving the goal by pressing the key marked "C" for consistent or "I" for inconsistent. Inconsistent trials were created by manipulating one of three flight elements (i.e., altitude, heading, or airspeed) depicted in the second snapshot to render the "current" (Time 2) flight situation inconsistent with the specified goal. Once the participants pressed a response key, the goal description and the second snapshot disappeared, and the next trial was presented.

The participants were presented with seven practice trials in order to familiarize them with the procedure prior to viewing the task trials. The experimental materials included 50 (23 consistent and 27 inconsistent) snapshot sequences. They received accuracy feedback during the practice trials, but feedback was not provided during the experimental trials. Consistency judgment accuracy and the time that elapsed between the presentation of the second snapshot and the entry of the consistency judgment were recorded by the computer.

It is important to note that the situation recall task and the SA task shared a need for memory resources. The situation recall task required storing two entire cockpit displays in memory and retrieving the information for one of the displays during recall. The SA task also required storage in memory. However, the information to be stored was dictated by the specified goal. In addition, the SA task required the integration of display elements with the specified goal, in order to project a future state of the aircraft. The SA task, then, was a measure of how the pilots constructed a memory representation according to their goals and used this to build a situation model of a future flight state. This distinction between the two tasks was supported by the lack of a consistent correlation between the two measures; the novices showed a low correlation ($r = .24$, $p < .3$), whereas the experts showed a high correlation ($r = .54$, $p < .01$). If they had substantially overlapped, the correlation should have been equally high for both groups.

RESULTS AND DISCUSSION

WM Capacity Measure

A summary of the descriptive statistics for the span measures for the novices and the experts is presented in

Table 1
Summary of Descriptive Statistics for Span Measures

Measure	Novice				Expert			
	M	SD	Min	Max	M	SD	Min	Max
Spatial span	3.03	1.11	1.00	5.00	3.50	1.07	2.00	5.00
Verbal span	3.49	0.91	1.50	5.00	3.68	1.00	1.50	5.00

Table 1. As was previously mentioned, the capacity account proposes that individual differences in WM capacity are inherent in the individual and do not vary with expertise. Two-tailed t tests resulted in no significant difference between the two groups in either of the span measures (ps < .2). This is consistent with the hypothesis that WM capacity does not change for given individuals as they acquire expertise. The lack of a significant relationship between WM capacity and expertise also rules out the idea that WM capacity influences the development of expertise in a definite direction—that is, that those with different WM capacities are better or less able to become experts in the complex task domain.

LTWM Skill Measure

A summary of the descriptive statistics for delayed-recall accuracy and LTWM skill scores is presented in Table 2. As was previously discussed, LTWM skill scores were calculated by subtracting delayed-recall accuracy for the nonmeaningful situations from that for the meaningful situations, indicating the skill with which pilots encoded situational information in an accessible form in LTWM. The experts showed higher mean recall accuracy than did the novices for meaningful situations [$t(48) = 2.06, p < .04$], but this was not true for nonmeaningful situations [$t(48) = 0.32, p < .7$]. As was expected, the LTWM skill scores were higher for the experts than for the novices [$t(48) = 2.20, p < .03$]. These results are consistent with the LTWM hypothesis that expert pilots, who possess extensive relevant knowledge and experience with meaningful cockpit instrument values, will generate and use more efficient retrieval structures.

Although these results support the assumption that underlies the validity of using meaningfulness as an indication of retrieval structure, the measure of retrieval structure construction was indirect, and the specific forms of the retrieval structures of the pilots were not identified in this study. However, according to previous literature (Stokes, Belger, & Zhang, 1990; Stokes et al., 1997), pilots are likely to structure cockpit situational information in LTM by reference to the large repertoire of specific flight scripts (e.g., straight-and-level, cruise climb, approach descent) that pilots have internalized. Meaningful configurations and coherent sequences of instrument indications should pertain to the scripts in the pilot's repertoire and, thus, serve to organize the presented cockpit information in an accessible form.

SA Performance Measures

To measure pilot SA performance, we used the accuracy of consistency judgments to calculate hits (correct judgments for consistent stimuli) and false alarms (incorrect judgments for inconsistent stimuli) and used these to determine pilot judgment sensitivity (d') and bias (beta) (for details of the procedures, see Green & Swets, 1966).[2] Judgment latency was also measured as another index of SA performance. Table 3 shows mean judgment sensitivity, bias (a 0 bias score indicates no bias; positive and negative scores indicate a bias to respond *inconsistent* and *consistent*, respectively), and latency for the novices and the experts. Judgment sensitivity was higher for the experts than for the novices [$t(48) = 2.25, p < .03$], but there was no group difference in bias or latency (ps < .7). The results suggest that the SA task differentiated between novices and experts, although group performance differences were due to judgment sensitivity, not to judgment criterion or latency differences. Because judgment sensitivity varied as a function of expertise, we used this performance measure as the criterion measure to be predicted by the WM capacity and LTWM skill measures.

Predicting SA Performance

The major question of interest in this research was whether there are main and interactive effects of WM capacity and LTWM skill on complex task performance. To address this question, hierarchical regression analyses were conducted using WM capacity, LTWM skill, and their interaction as predictors of SA performance. These analyses allowed us to determine whether the effect of one predictor variable on SA performance is significant when the other predictor variables are controlled for (Cohen & Cohen, 1983).

We created a composite measure representing WM capacity by averaging the z scores for the spatial and the verbal spans. This was justified because the correlation between the two measures of WM capacity ($r = .42, p < .01$) was strong, which is suggestive of convergent validity. In addition, there was a weak correlation of the composite measure with the LTWM skill measure ($r = .20, p < .2$), which is suggestive of discriminant validity.

The hierarchical regression analyses included the composite WM capacity measure, the LTWM skill measure, and a cross product of these two measures (LTWM skill × WM capacity), representing their interaction as independent variables. The analyses were conducted in three steps. LTWM skill was entered in the first step, WM capacity was entered in the second step, and the cross product was entered in the third step. This hierar-

Table 2
**Summary of Descriptive Statistics
for Situation Recall Accuracy**

Situation Meaningfulness	Novice		Expert	
	M	SD	M	SD
Meaningful situation	0.56	0.16	0.65	0.13
Nonmeaningful situation	0.35	0.12	0.37	0.13
LTWM score*	0.21	0.14	0.28	0.08

Note—LTWM, long-term working memory.   *Difference in recall accuracy between meaningful and nonmeaningful situations.

Table 3
Mean Judgment Sensitivity, Bias, and
Correct Judgment Latency for Novices and Experts
in the Situation Awareness (SA) Task

SA Measures	Novice		Expert	
	M	SD	M	SD
Judgment sensitivity (d')	1.24	0.54	1.55	0.42
Judgment bias (log-beta)	−0.31	0.58	−0.26	0.52
Judgment latency (seconds)	8.82	2.53	8.30	2.88

chical order for entry reflected what we presumed to be the direction of influence in the relation of LTWM skill and WM capacity. Our hypothesis for the direction of the influence was that an increase in LTWM skill would decrease the pilot's reliance on WM capacity as the determinant of SA performance. The cross product was entered only after the two variables that might have been a source of interaction effect on SA performance had been entered. A significant increment in variance (referred to as Inc. R^2 for the purpose of exposition) accounted for by the variable entered in each step would indicate the unique contribution of that variable to SA performance. For example, a significant Inc. R^2 in the third step would indicate a LTWM skill × WM capacity interaction, independent of the main effects.

The results of the hierarchical regression analysis showing whether LTWM skill, WM capacity, and their interaction predict pilot SA performance are summarized in Table 4. There were significant main effects of both LTWM skill and WM capacity, which suggests that the higher the LTWM skill or the WM capacity, the better the performance on the SA task. Also, there was a significant LTWM skill × WM capacity interaction. For the purpose of exposition, the relation of WM capacity to SA performance as a function of LTWM skill is characterized in Figure 1. In this figure, the low-skill participants have z scores of 0 or lower for the LTWM skill measure, and the high-skill participants have z scores higher than 0. The figure suggests that the effect of WM capacity on SA performance was prominent for the low LTWM skill participants, but not for the high LTWM skill participants. That is, for those participants with high levels of LTWM skill, the reliance on WM capacity should be low. This finding is consistent with the hypothesis that LTWM skills reduce the importance of WM capacity during the performance of complex cognitive tasks.

Also of interest was whether the effects of WM and LTWM measures on SA performance differed as a function of expertise. To address this question, we conducted the attribute × treatment interaction analysis (see Pedhazur & Schmelkin, 1991), which determines the significance of differences in regression lines as a function of expertise. There were no significant interactions between novice and expert regression lines with this analysis (ps < .2). However, we examined the effects of WM and LTWM measures on SA performance qualitatively by analyzing them separately for the novices and the ex-

perts, using the hierarchical regression analysis. In this context, WM capacity was predictive of novice SA performance [Inc. R^2 = .17, B = .28; F change (1,22) = 4.90, p < .04], whereas LTWM skill was predictive of expert SA performance [Inc. R^2 = .29, B = .29; F change (1,23) = 9.39, p < .01]. In addition, LTWM skill and WM capacity interacted to predict expert SA performance [Inc. R^2 = .13, B = −.21; F change (1,21) = 4.85, p < .04]. Although this must be interpreted with caution because of the separation of groups, it does suggest that WM capacity facilitates novice performance and LTWM skill facilitates expert performance. Most interesting is the suggestion that an increase in LTWM skill reduces the role of WM capacity in SA performance, particularly among experts.

GENERAL DISCUSSION

The present research represents a theoretical contribution to the understanding of the roles of WM capacity and LTWM skill in a complex and knowledge-intensive task performance. Our findings serve to distinguish WM capacity and LTWM skill as distinct WM constructs and further advance our theoretical understanding of how they coexist and interact to support complex task performance.

Distinguishing the Two Constructs

Capacity theory and LTWM theory propose contrasting accounts for differences in human performance. According to capacity theory, the most plausible locus of individual differences in performance is the domain-general capability to maintain presented information in an activated state (e.g., Daneman & Carpenter, 1980; Engle et al., 1999; Just & Carpenter, 1992; Shah & Miyake, 1996). This capability is assumed to differ inherently across individuals. In contrast, according to LTWM theory, the locus of the performance differences is the acquired domain-specific skill at encoding the presented information efficiently in accessible form in LTM (e.g., Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995; Kintsch, 1998). Whereas the former theory proposes that the amount of information that can be maintained in accessible form in STM is limited by a fixed capacity, the latter theory proposes that WM capacity dynamically changes within individuals as a function of LTWM accessibility during a given task.

To determine whether WM capacity and LTWM skill are distinct constructs, we examined the correlation be-

Table 4
Results of Hierarchical Regression Analyses Predicting Pilot
Situation Awareness (SA) Performance

Predictor Variable	R^2	Inc. R^2	B	F change
LTWM skill	.15	.15	.20	8.49**
WM capacity	.24	.09	.19	5.79*
LTWM skill × WM capacity	.31	.07	−.19	4.33*

Note—LTWM, long-term working memory.   *p < .05.   **p < .01.
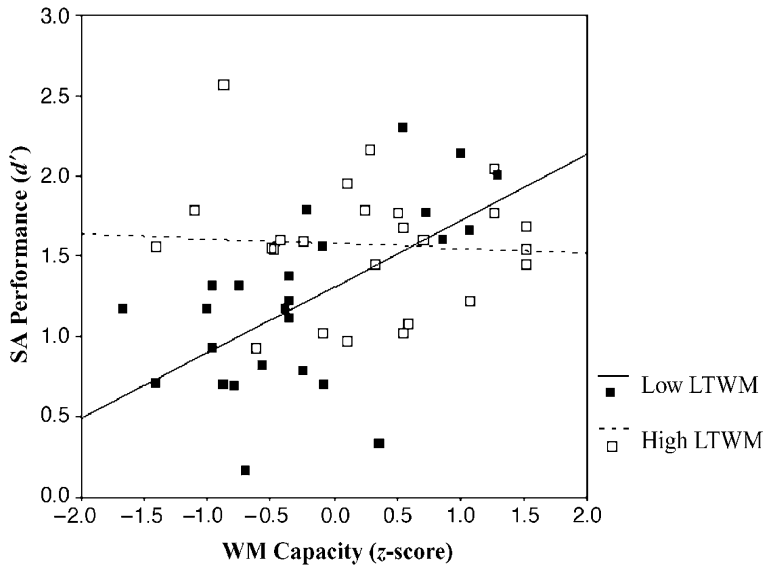
Figure 1. Relation of working memory (WM) capacity to situation awareness (SA) performance as a function of long-term WM (LTWM) skill.

tween the two measures. If individual differences in WM capacity reflected differences in acquired LTWM skill or vice versa, we should have observed a high correlation between the span and the LTWM skill measures. A significant correlation between the two measures was not observed for experts ($r = .26$, $p < .3$), who have acquired reliable skills in LTM encoding and retrieval, or for novices ($r = .10$, $p < .6$), indicating a distinction between the two WM constructs.

The lack of a significant correlation between WM capacity and LTWM skill measures is important because it addresses the theoretical issue raised by Ericsson and Kintsch (1995). As was discussed previously, capacity theory emphasizes a domain-general ability that is separate from any acquired domain-specific skill. In contrast, LTWM theory emphasizes acquired domain-specific skill and further proposes a variable correlation between the two measures as a function of the commonality of mental representations. Ericsson and Kintsch argued that measures of WM capacity (e.g., reading span) capture acquired domain-specific skill (e.g., retrieval structures possessed by readers). However, the present results suggest that the measures of WM capacity and LTWM skill used in this study are distinct. This fact provided a firm ground on which to evaluate the relative contributions of domain-general and domain-specific individual difference characteristics to complex task performance.

Coexistence

Although capacity and LTWM accounts of individual differences in complex task performance are competing accounts, our findings suggest that both memory mechanisms exist and play a significant role in complex task performance. In the previous literature, WM has been

examined with methods that minimize the role of expertise in order to examine inherent capacities (e.g., Daneman & Carpenter, 1980; Engle et al., 1999; Just & Carpenter, 1992; Shah & Miyake, 1996) or that minimize the role of such capacities in order to focus on the role of expertise (e.g., Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995).

In the present research, we simultaneously examined the respective roles of the two WM resources (WM capacity and LTWM skill) in an experimental context in which both were important to perform the experimental task. In this context, both WM capacity and LTWM skill affected complex task performance, although their respective roles might have varied as a function of expertise. WM capacity was more predictive of performance in the earlier stage, whereas LTWM skill was more predictive of performance in the later stage, of the development of expertise. This is consistent with the prediction of skill acquisition theories, which states that general cognitive ability plays a greater role in the initial stages of learning than do practiced skills (e.g., Ackerman, 1988). This explication can harmonize the two accounts that propose different explanations for individual differences in complex task performance. In sum, the two WM mechanisms coexist and play roles in complex performance, but which is more prominent depends, in some part, on the extent to which expertise has been developed.

Interaction

If the capacity- and LTWM-based mechanisms coexist, how do the two mechanisms work together? This research provides an insight into their interplay. Our findings suggest that these two memory mechanisms interact in such a way that performers with higher LTWM
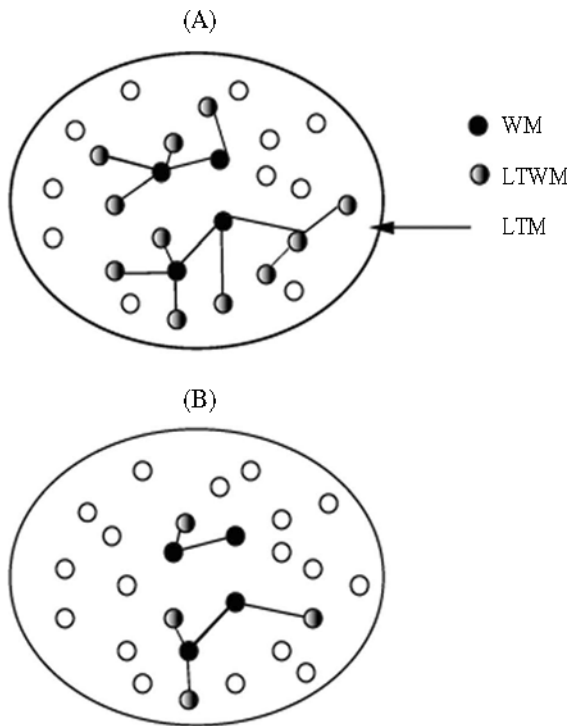
Figure 2. Hypothetical representation of relationships between working memory (WM) capacity and long-term WM (LTWM) for (A) those with high levels of LTWM skill and (B) those with low levels of LTWM skill. From *Comprehension: A Paradigm for Cognition* (p. 218), by W. Kintsch, 1998, Cambridge: Cambridge University Press. Copyright 1998 by Cambridge University Press. Adapted with permission.

skill rely less on WM capacity during the performance of complex tasks. Stated differently, the effect of WM capacity is smaller for performers with higher LTWM skill. This suggests that LTWM can be utilized as an efficient extension of WM to overcome the inherent limits of WM.

To elaborate, the hypothesized relationship between the two WM constructs is illustrated in Figure 2. As is shown in the figure, WM (depicted by black circles) is the activated part of LTM, and LTWM (depicted by gray circles) is the part of LTM linked to WM elements by retrieval structures. The number of WM elements (i.e., WM capacity) is limited, containing no more than 7 ± 2 chunks, although it varies across individuals (Miller, 1956). The elements (black circles) available in the capacity-limited WM serve as retrieval cues to make accessible a set of LTM elements (gray circles) that are linked by retrieval structures. Hence, the WM capacity can be expanded functionally, containing not only items already in WM, but also as many LTWM elements as are made accessible by retrieval structures. In this way, the capacity of LTWM is constrained by the extent and efficiency of the retrieval structures that can be accessed through the WM elements.

For the purpose of exposition, hypothetical memory representations varying as a function of WM capacity and LTWM skill are illustrated in Figures 2A and 2B. As compared with Figure 2B, Figure 2A shows that the relative portion of WM elements (black circles) declines as the portion of LTWM elements (gray circles) increases. Our findings suggest that such an increase in LTWM should reduce one's reliance on WM elements as the determinant of complex task performance.

Finally, our explanation for the relation of LTWM skill and WM capacity is consistent with that found in some research on expertise and aging (e.g., Charness, 1991; Clancy & Hoyer, 1994; but see Hambrick & Engle, 2002; Meinz, 2000). The research suggests that increases in domain-relevant knowledge and experience reduce the effect of age-related declines in general cognitive ability on complex task performance. Older adults with high levels of expertise rely on cognitive resources to a lesser extent (e.g., Morrow et al., 2001). To explain such relationships between expertise and aging, researchers have speculated that there is an interplay of domain-specific skill and general ability (e.g., Clancy & Hoyer, 1994; Morrow et al., 2001). The present research provides empirical evidence for this interplay and, in so doing, advances our understanding of the relationships between the two cognitive constructs in the context of complex task performance.

REFERENCES

ACKERMAN, P. L. (1988). Determinants of individual differences during skill acquisition: A theory of cognitive abilities and information processing. *Journal of Experimental Psychology: General*, **117**, 299-329.

ADAMS, J. W., & HITCH, G. J. (1997). Working memory and children's mental addition. *Journal of Experimental Child Psychology*, **67**, 21-38.

ANDERSON, J. R., & JEFFRIES, R. (1985). Novice LISP errors: Undetected losses of information from working memory. *Human–Computer Interaction*, **1**, 133-161.

BADDELEY, A. D. (1986). *Working memory*. New York: Oxford University Press.

BADDELEY, A. D., & HITCH, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

CARPENTER, P. A., JUST, M. A., & SHELL, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, **97**, 404-431.

CHARNESS, N. (1991). Expertise in chess: The balance between knowledge and search. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 39-63). Cambridge: Cambridge University Press.

CHASE, W. G., & ERICSSON, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 1-58). New York: Academic Press.

CHASE, W. G., & SIMON, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

CLANCY, S. N., & HOYER, W. J. (1994). Age and skill in visual search. *Developmental Psychology*, **30**, 545-552.

COHEN, J., & COHEN, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

DANEMAN, M., & CARPENTER, P. A. (1980). Individual differences in

working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, **19**, 450-466.

DANEMAN, M., & TARDIF, T. (1987). Working memory and reading skill reexamined. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 491-508). Hillsdale, NJ: Erlbaum.

DOANE, S. M., MCNAMARA, D. S., KINTSCH, W., POLSON, P. G., & CLAWSON, D. M. (1992). Prompt comprehension in UNIX command production. *Memory & Cognition*, **20**, 327-343.

DOANE, S. M., & SOHN, Y. W. (2000). ADAPT: A predictive cognitive model of user visual attention and action planning. *User Modeling & User-Adapted Interaction*, **10**, 1-45.

DOANE, S. M., SOHN, Y. W., MCNAMARA, D. S., & ADAMS D. (2000). Comprehension-based skill acquisition. *Cognitive Science*, **24**, 1-52.

DURSO, F. T., & GRONLUND, S. D. (1999). Situation awareness. In F. T. Durso, R. Nickerson, R. Schvaneveldt, S. Dumais, S. Lindsay, & M. T. H. Chi (Eds.), *The handbook of applied cognition* (pp. 283-314). New York: Wiley.

ENDSLEY, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, **37**, 32-64.

ENDSLEY, M. R., & BOLSTAD, C. A. (1994). Individual differences in pilot situation awareness. *International Journal of Aviation Psychology*, **4**, 241-264.

ENDSLEY, M. R., ENGLISH, T. M., & SUNDARARAJAN, M. (1997). The modeling of expertise: The use of situational models for knowledge engineering. *International Journal of Cognitive Ergonomics*, **1**, 119-136.

ENGLE, R. W., CARULLO, J. J., & COLLINS, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research*, **84**, 253-262.

ENGLE, R. W., KANE, M. J., & TUHOLSKI, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and the functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). New York: Cambridge University Press.

ERICSSON, K. A., & DELANEY, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 257-297). New York: Cambridge University Press.

ERICSSON, K. A., & KINTSCH, W. (1995). Long-term working memory. *Psychological Review*, **102**, 211-245.

ERICSSON, K. A., & POLSON, P. G. (1988). Memory for restaurant orders. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 23-70). Hillsdale, NJ: Erlbaum.

ERICSSON, K. A., & STASZEWSKI, J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235-267). Hillsdale, NJ: Erlbaum.

GREEN, D. L., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

GROEN, G. J., & PATEL, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. T. H. Chi, R. J. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 287-310). Hillsdale, NJ: Erlbaum.

HAMBRICK, D. Z., & ENGLE, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, **44**, 339-387.

HATANO, G., AMAIWA, S., & SHIMIZU, K. (1987). Formation of a mental abacus for computation and its use as a memory device for digits: A developmental study. *Developmental Psychology*, **23**, 832-838.

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

KINTSCH, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

KYLLONEN, P. C., & STEPHENS, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning & Individual Differences*, **2**, 129-160.

LARKIN, J., MCDERMOTT, J., SIMON, D. P., & SIMON, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, **208**, 1335-1342.

LOHSE, G. L. (1997). The role of working memory in graphical information processing. *Behavior & Information Technology*, **16**, 297-308.

MEINZ, E. J. (2000). Experience-based attenuation of age-related differences in music cognition tasks. *Psychology & Aging*, **15**, 297-312.

MILLER, G. A. (1956). The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psychological Review*, **63**, 81-97.

MORROW, D. G., MENARD, W. E., STINE-MORROW, E. A. L., TELLER, T., & BRYANT, D. (2001). The influence of expertise and task factors on age differences in pilot communication. *Psychology & Aging*, **16**, 31-46.

MORROW, D. G., & RODVOLD, M. (1998). Issues in air traffic control communication. In M. Smolensky & E. Stein (Eds.), *Human factors in air traffic control* (pp. 421-456). New York: Academic Press.

PEDHAZUR, E. J., & SCHMELKIN, L. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.

PETERSON, L. R., & PETERSON, M. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, **58**, 193-198.

RADVANSKY, G. A., & COPELAND, D. E. (2001). Working memory and situation model updating. *Memory & Cognition*, **29**, 1073-1080.

SAARILUOMA, P. (1989). Chess players' recall of auditorily presented chess positions. *European Journal of Cognitive Psychology*, **1**, 309-320.

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

SOHN, Y. W., & DOANE, S. M. (1997). Cognitive constraints on computer problem-solving skills. *Journal of Experimental Psychology: Applied*, **3**, 288-312.

STASZEWSKI, J. J. (1988). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 71-128). Hillsdale, NJ: Erlbaum.

STOKES, A. F., BELGER, A., & ZHANG, K. (1990). *Investigation of factors comprising a model of pilot decision making: Anxiety and cognitive strategies in expert and novice aviators* (Tech. Rep. ARL-90-8/SCEEE-90-2). Champaign: University of Illinois, Institute of Aviation.

STOKES, A. F., KEMPER, K., & KITE, K. (1997). Aeronautical decision making, cue recognition, and expertise under time pressure. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 183-196). Mahwah, NJ: Erlbaum.

NOTES

1. We confirmed this fact with flight instructors serving as consultants to the research.

2. To develop a bias measure, beta values were calculated for each participant according to Green and Swets (1966). The beta values between 0 and 1 indicate a bias to respond *consistent,* whereas the values between 1 and infinity indicate a bias to respond *inconsistent.* Because the distribution of these values is asymmetric around 1, we analyzed log-beta, which is distributed symmetrically around zero, and used log-beta values as bias scores for this study.