Event Reasoning as a Function of Working Memory Capacity and Long Term Working Memory Skill

Mark T. Jodlowski (mtj27@ra.msstate.edu)
Stephanie M. Doane, Ph.D. (sdoane@ra.msstate.edu)
Department of Psychology, P.O. Box 6161
Mississippi State, MS 39762 USA

Abstract

The present research examined the relationship of working memory (WM) capacity and long term working memory (LTWM) skill in complex task performance as a function of expertise. Individual differences in WM capacity and LTWM skills of 62 aviation pilots (Expert = 25, Novice = 37) were related to their performance on a task designed to measure components of flight situation awareness (SA). LTWM skill and WM capacity were not correlated, suggesting they are distinct constructs. Experts have higher LTWM skill compared to novices, and LTWM skill acts as a predictor of Expert SA task performance. The implications of these results are discussed.

There is currently a debate in the literature regarding the explanation of individual differences in complex task performance relative to WM capacity (e.g., Baddeley & Hitch, 1974) and LTWM skill (e.g., Ericsson & Kintsch, 1995). The present research addresses this debate. Prior to describing this research, a brief summary of the two theories and the debate is provided.

Working memory (WM), defined as a limited, temporary store for processing and storing information (Baddeley & Hitch, 1974), has been studied extensively in various cognitive tasks. Relevant to the present research, individual differences in cognitive task performance are often explained by WM capacity. For example, Just and Carpenter (1992) proposed that WM capacity constrains text comprehension. Once an individual's WM capacity is reached, the lack of available processing and storage hinders the ability to use and retain new information as well as intermediate products resulting from newly obtained information, resulting in decreased comprehension. Further research suggests a role of WM capacity in performing additional cognitive tasks, such as spatial visualization (Shah & Miyake, 1996), the ability to follow complex directions (Engle, Carullo, & Collins, 1991), and computer problem solving (e.g., Anderson & Jeffries, 1985; Doane, McNamara, Kintsch, Polson, & Clawson; 1992, Doane, Sohn, McNamara, & Adams, 2000; Sohn & Doane, 1997). In this view, WM is thought to reside in short term memory (STM); is believed to have a fixed capacity, although individuals differ in their capacities; and an individual's WM capacity remains stable over time (e.g., Baddeley & Hitch, 1974).

Alternatively, some researchers argue that WM capacity is not fixed (Ericsson & Kintsch, 1995) but instead varies as a function of expertise within a specific domain. This theory is referred to as Long Term Working Memory (Ericsson & Kintsch).

LTWM is a theory of a memory process that explains how individuals can extend WM capacity well beyond the proposed seven plus or minus two chunks (Miller, 1956). In LTWM theory, domain-specific knowledge and meaningful experiences are thought to increase individual ability to efficiently encode information into long term memory (LTM) and to create easily accessible retrieval structures. Ericsson and Kintsch (1995) defined a retrieval structure as an organization of meaningful data into a stable structure that can be used to rapidly encode information into and retrieve information from LTM. LTWM uses these retrieval structures as indices to situation-specific information that is temporarily stored in LTM. Because the information is temporarily stored in LTM, after a disruption, a task can still be completed by activating the necessary indices required to retrieve the situation-specific information. The indices change dynamically as a function of the task at hand and the individual's expertise for that particular task.

In summary, the WM capacity and LTWM theories differ in their views about whether WM capacity is fixed for a given individual and how WM capacity responds to the dynamics of task environments. In addition, at first glance they appear to offer competing views of the mechanisms that govern WM capacity and where they reside (STM or LTM).

However, some evidence suggests that LTWM and WM coexist, are independent constructs (Sohn & Doane, 2003), and serve distinct roles in cognitive task performance. Sohn and Doane devised a measure of LTWM skill, measured individual WM capacity and LTWM skill, and related these measures to performance on a complex cognitive task. They found that WM capacity and LTWM skill were not correlated, suggesting the two are independent constructs. WM capacity predicted novice task performance, whereas LTWM skill predicted expert task performance. In addition, LTWM skill appeared to have a compensatory role for experts with low WM capacity; experts with high LTWM skill tended to have lower WM capacity.

The present study extends Sohn and Doane's (2003) research paradigm to further examine the roles of WM capacity and LTWM skill in complex cognitive task performance. Specifically, the relationships of WM capacity and LTWM skill are related to aviation pilot performance on a flight situation awareness (SA) task. SA is a term that can be broadly defined as a pilot's ability to understand his or her current situation and anticipate future status (Endsley, 1995). SA has been cited as a leading cause in military aviation mishaps involving human error (e.g., Hartel, Smith, & Prince, 1991; Salas, Prince, Baker, & Shreshta, 1995). Building SA is hypothesized to require both WM and LTWM resources (Durso & Gronlund, 1999), and, as previously mentioned, initial findings support this hypothesis (Sohn & Doane).

To summarize, the purpose of this research is to examine the relationship between WM capacity and LTWM skill in complex task performance. Two different theories account for performance differences. Capacity theory suggests that WM is limited, fixed in size, and resides in STM. In contrast, LTWM theory suggests that WM capacity is unlimited, changes in size, resides in LTM, and is a function of expertise.

Method

Participants

Fifty-two U.S. Navy student and instructor pilots and 25 local pilots were recruited to participate in the experiment. Each local pilot was paid $20 per hour for his/her participation in the two-session (2.5-hour) experiment. During the first session, pilot WM capacity and LTWM skill were assessed. In the second session, pilots completed a task designed to measure SA. Each of the three different tasks is described in the subsequent sections.

In order to obtain information about piloting expertise, each pilot completed a questionnaire before the experiment began. Pilots were classified into different levels of experience with a discriminant analysis based on questionnaire data. As a result of the analysis, the pilots were classified into two groups: expert (M = 2030 flight hours, SD = 1235) and novice (M = 55 flight hours, SD = 60) groups contained 36 and 41 pilots, respectively. Eleven of the expert pilots and 4 of the novice pilots' data were not used because of their failure to complete the SA task (i.e., skipping numerous trials).

WM Assessment. In order to assess individual WM capacity, pilots completed an aviation-based WM capacity assessment task. This task was developed as a combined analog measure of both verbal WM (Daneman & Carpenter, 1980) and spatial WM (Shah & Miyake, 1996). In this task, pilots viewed a series of attitude indicator displays positioned in different flight orientations (see Figure 1a). Each attitude indicator displayed one of 14 different positions. The attitude indicators displayed an aircraft as being pitched up or down with a bank angle of either 30, 60, 90, or 0 degrees to the left or right of straight and level flight. In addition, a number between 15 and 33, corresponding to common headings on the heading indicator, appeared immediately below the attitude indicator (see Figure 1a). The possible numbers were 15, 18, 21, 24, 27, 30, and 33. Each attitude indicator was presented for 2,200 ms. Upon presentation, pilots were asked to say aloud whether the aircraft was pitched up or down.

Pilots were asked to remember the orientation of the horizon line displayed on the attitude indicator and the number below the attitude indicator (see Figure 1b). After viewing all of the attitude indicators in sequence, pilots were asked to keyboard in the number presented beneath the first attitude indicator. They were then asked to indicate their memory for the orientation of the horizon line in each attitude indicator by selecting the square that indicated the direction the horizon line was pointing towards (see Figure 1b). Although this does not seem to be an easy task to non-pilots, pilots have no difficulty completing this task. The direction the horizon is pointed towards and the selected square are shown in Figure 1b for the reader; they were not shown in the experiment. Pilots first viewed a series of two attitude indicators for five trials and progressed through a series length of three and four, each containing five trials.

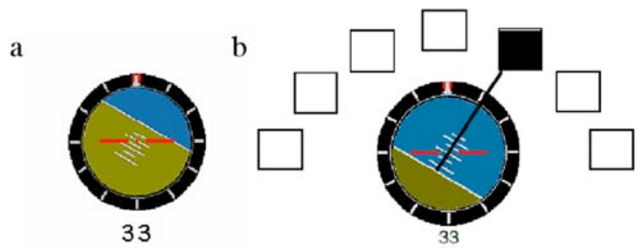

Figure 1. (a) Pitched down attitude indicator. (b) Pitched up attitude indicator that depicts the direction the horizon is pointing towards.

LTWM Assessment. To assess LTWM skill, a domain-specific piloting task similar to chess experiments (e.g., Charness, 1976; de Groot, 1965) was developed. In this task, pilots simultaneously viewed two cockpits for 40 seconds (see Figure 2a-b). One cockpit was displayed on the top half of the screen, with the second cockpit displayed directly below the first. After 40 seconds elapsed, the computer presented a number and asked the pilot to count backwards aloud by threes for 30 seconds, starting from the presented number. For example, if the number presented was 342, pilots should respond aloud: 339, 336, 333, 330, 327, and so on. After counting backwards by threes for 30 seconds, the computer prompted the pilots to recall the situation-specific values displayed in either the top or the bottom cockpit. Pilots did not know in advance which cockpit they would be asked to recall. Pilots used a sheet of paper containing the seven instruments with no values and a pen to fill in the situation-specific values for each instrument. No time limits were imposed on pilot recall of the instrument values. When the pilots completed the first trial, they pressed the "Return" key on the keyboard, and the next trial was automatically presented.

In six of the trials, the two cockpits were related. That is, the bottom cockpit represented the future state of the aircraft 5 to 10 seconds after applying one or two control movements to the top cockpit (see Figure 2a). Three trials consisted of two unrelated cockpits; both depicting nonmeaningful flight configurations (see Figure 2b). For example, the attitude indicator would display the aircraft pitched up and banked to the right, whereas the turn coordinator would display the aircraft as being banked to the left. In addition, the vertical speed indicator might display the aircraft as descending.

The LTWM task was designed to demonstrate the effect of retrieval structures. Use of both meaningful and nonmeaningful situations allowed us to differentiate between retrieval originating from the use of LTWM retrieval structures and retrieval originating from the use of STM, respectively. Because WM capacity is thought to be temporary and limited in size, counting backwards by three from a given number requires processing and storage that would use WM processes, thus WM capacity could not account for pilot ability to recall cockpit information. Furthermore, because there are an infinite number of cockpit situations, it is difficult to argue that a specific situation is retrieved as a pattern stored in LTM.

SA Task. In the SA task, each trial consisted of a series of four screens that depicted a desired flight status, a current flight situation (cockpit 1), questions about methods to achieve the desired status (control movement selection), a future flight situation intermediate to the desired status (cockpit 2), and one of two types of inquiries about a change in flight status. Figure 4 depicts the names of each screen and a schematic of a trial. The present paper focuses on the consistency judgment that takes place after viewing cockpit 2. The task required for response to the fourth screen of each trial and the data from those trials will not be discussed further. What follows is a detailed description of the task as it relates to the first three screens of each trial.

The first screen contained text described as a desired heading, altitude, and airspeed. It also depicted a flight situation. Pilots were asked to assess the flight situation and the desired flight status specified in the goal statement and click "Next" when they felt they knew what flight control movements would be required to reach the goal. For example, Figure 4 shows an altitude of 3,530 feet and airspeed of 95 knots. The goal is to get to an altitude of 3,430 feet and airspeed of 105 knots. To reach the goal, the pilot would need to descend by moving the elevator forward. Therefore, on the control movement selection screen, the pilot would select "Forward pressure on elevator."

After selecting a control movement(s), pilots clicked "Next" to view the third screen (cockpit 2). The third screen depicted a future flight situation that resulted from the application of one or two control movements to the starting situation (cockpit 1). Cockpit 2 was not affected by the pilots' control movement selection. The pilots' task was to determine if cockpit 2 accurately depicted a flight situation that would reach the goal described in the first screen (cockpit 1) within the next 5 to 10 seconds of mentally simulated flight. Pilots pressed the keyboard key labeled "C" if the flight situation was consistent with obtaining the goal in the next 5 to 10 seconds. If the situation was inconsistent with obtaining the goal within the next 5 to 10 seconds, they pressed the key labeled "I." Referring again to cockpit 2 in Figure 4, it appears that the control movement applied to cockpit 1 was "Forward pressure on elevator" (the attitude indicator and VSI indicate that the plane is flying level in cockpit 1 and descending in cockpit 2). The pilot should indicate that cockpit 2 is consistent with obtaining the goal within the next 5 to 10 seconds of simulated flight by pushing key labeled "C" on the keyboard.

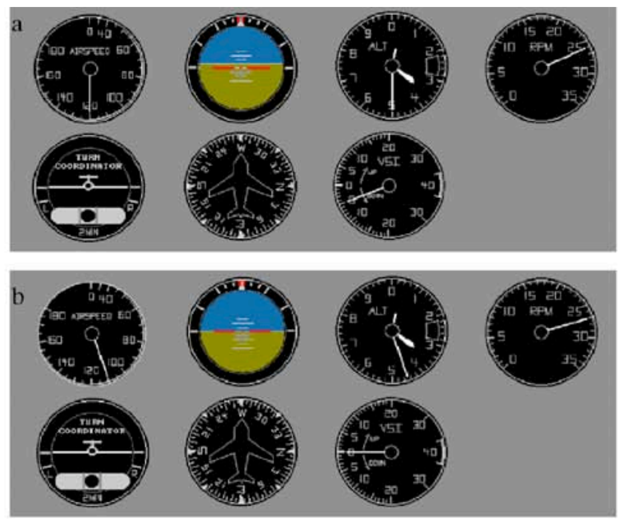

Figure 2. (a) Meaningful starting flight situation. (b) Meaningful future status after a control movement is executed.

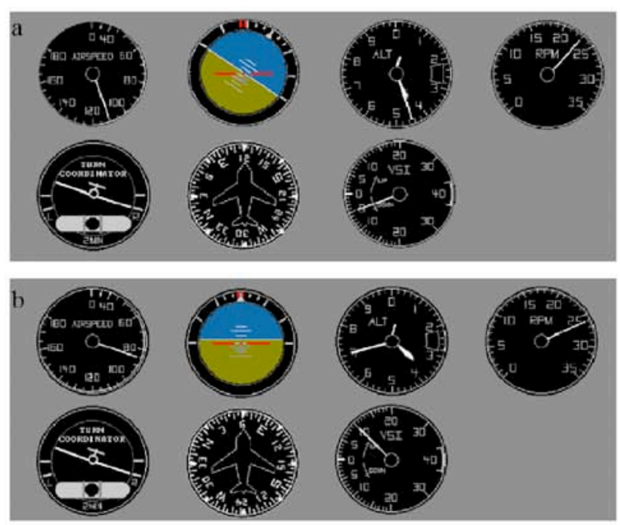

Figure 3. (a) Nonmeaningful starting situation. (b) Non-meaningful future status (instruments are in conflict in both cases).

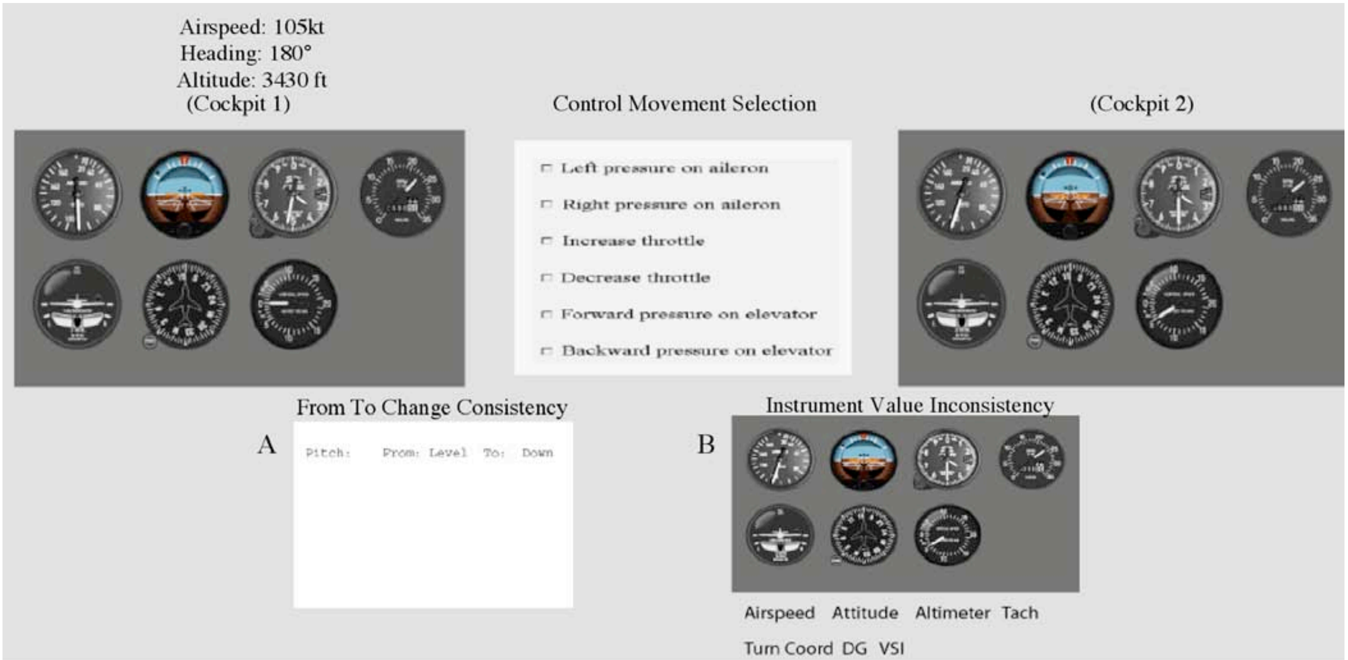

Figure 4. Flow from start to finish of an SA trial. The pilot views the cockpit and the goal statement, chooses the controls to reach the goal, views a second cockpit, and determines if the second cockpit will reach the goal in the next 5 to 10 seconds of simulated flight. Screen A or B is displayed after a consistent or inconsistent judgment, respectively.

Results

WM and Expertise

Individual WM span was scored using a modification of Daneman and Carpenter's (1980) reading span measures. WM capacity scores for novices (M = 2.05, SD = 0.89) and experts (M = 2.14, SD = 0.99) did not differ as a function of expertise, $F(1, 61) = 0.13$, $p < .72$. This is consistent with the capacity theory view that individual differences in the amount of resources available to process and store information in WM are independent of acquired skills.

LTWM and Expertise

The difference between the recall accuracy scores for meaningful and nonmeaningful flight situations for the delayed recall task was used to measure LTWM skill. Recall that this difference allows us to differentiate between retrieval originating from STM (nonmeaningful) and that originating from LTWM (meaningful). We obtained the expected expertise effects. Expert accuracy for meaningful situations (M = 0.68, SD = 0.10) was greater than novice accuracy (M = 0.61, SD = 0.12), $F(1, 61) = 5.95$, $p < .02$. For the nonmeaningful trials, expert accuracy (M = 0.34, SD = 0.13) was not statistically different from novice accuracy (M = 0.40, SD = 0.12), $F(1, 61) = 3.30$, $p < .07$. As expected, mean LTWM scores for experts (M = 0.34, SD = 0.09) were greater than for novices (M = 0.21, SD = 0.12), $F(1, 61) = 19.63$, $p < .01$, suggesting that expert pilots are better able to create retrieval structures. One interpretation of these results is that expert pilots create retrieval structures that enable rapid access to LTM, which in turn enables them to recreate the situation. This result was obtained after a 30-second delay, in which pilots had to process and store information in WM.

Predicting Performance

Mean accuracy of consistency judgments served as a measure of performance. Signal detection theory (Green & Swets, 1966) was used to calculate observer sensitivity (d'). Correct judgments for consistent trials and incorrect judgments for inconsistent trials represented hits and false alarms, respectively. Overall, there were no significant differences in observer sensitivity as a function of expertise $F(1, 61) = 0.18$, $p < .68$. However, there were differences in performance as a function of expertise. As can be seen in Figure 5, when expert pilots chose the correct controls (control movement selection), their consistency judgment

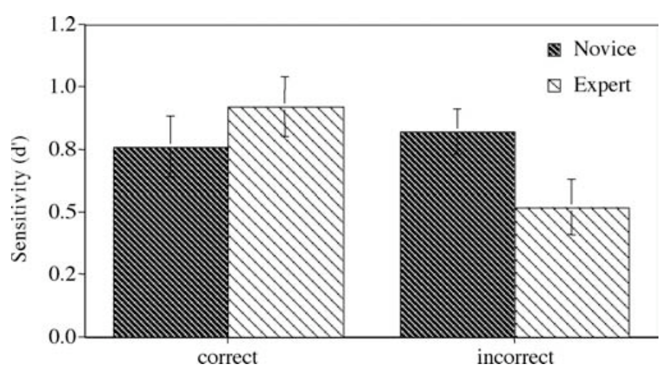

Figure 5. Accuracy of consistency judgments based on selecting the correct versus incorrect controls as a function of expertise.

Criterion Variable	Predictor Variable	R^2	Inc. R^2	B	F Change
Expert d'					
	LTWM	.19	.15	2.27	5.35*
	WM	.22	.15	0.09	0.90
	WM x LTWM	.25	.14	1.27	0.85
Novice d'					
	LTWM	.04	.01	-0.72	1.43
	WM	.04	-.02	0.01	0.01
	WM x LTWM	.04	-.05	0.24	0.08

Note. *$p < .05$.

Table 1. Results of hierarchical regression analyses predicting novice and expert pilot performance.

accuracy increased compared to when they chose the incorrect controls. This was not true for novices, as evidenced by a significant expertise x performance interaction as a function of control selection accuracy $F(1, 60) = 6.59$, $p < .02$. We state these findings to justify the analyses of WM capacity, LTWM skill, and SA performance as a function of expertise.

The focus of the present paper is to highlight the role of WM capacity and LTWM skill in predicting complex task performance. To test this, hierarchical regression analyses tested whether LTWM skill, WM capacity, or both predicted performance on the flight consistency judgment. 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. Hierarchical regression analyses examine the effect of one variable, while controlling for the effects of remaining variables (Cohen & Cohen, 1983).

The results of the hierarchical regression analyses for LTWM skill, WM capacity, and their interaction for predicting pilot accuracy on flight change consistency judgments are summarized in Table 1. Looking at the table, it appears that LTWM skill is a performance predictor for experts [Inc. $R^2 = 0.15$, $B = 2.27$; F change $(1, 23) = 5.35$, $p < .03$]. As can be seen in Figure 6, as LTWM skill increases for experts, d' increases as well. This effect was not observed in novice pilots [Inc. $R^2 = 0.01$, $B = -0.72$; F change $(1,35) = 1.43$, $p < .24$].

Previous research has implicated WM capacity as a predictor of complex task performance for novices. This result was not replicated and is left unexplained. However, one hypothesis is that the present task involved extensive processing and storage, resulting in a WM capacity overload for all individuals.

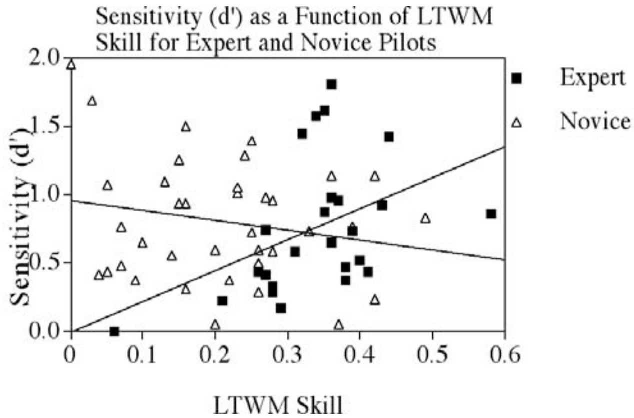

Figure 6. Relation of LTWM skill to d' as a function of expertise.

General Discussion

The present research contributes to understanding the role of WM capacity and LTWM skill in complex task performance. What follows is a discussion of how our findings support WM capacity and LTWM skill as distinct cognitive constructs and the implications of our findings.

WM Capacity and LTWM Skill

There is a clear theoretical difference between how WM capacity and LTWM skill account for human performance differences. Capacity theory postulates that performance is based on the ability to actively maintain presented information (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992; Shah & Miyake, 1996). The information is temporarily stored in STM. In addition, capacities are fixed and stable within an individual but differ between individuals. In contrast, LTWM theory suggests that performance differences are a function of the ability to temporarily and efficiently store information in and retrieve information from LTM (Ericsson & Kintsch, 1995). The ability to store and access information quickly results in a flexible WM capacity that changes as a function of expertise.

Previous research has indicated that WM capacity and LTWM skill are distinct constructs in an experimental task (e.g., Sohn & Doane, 2003) and our results support this finding. If individual differences in LTWM skill reflected differences in WM capacity or vice versa, a high correlation between LTWM skill and WM capacity measures would be expected. Correlations between WM capacity and LTWM skill were weak and unreliable. Thus, the present results provide further evidence that WM capacity and LTWM skill coexist.

Further support of WM capacity and LTWM skill as distinct constructs was seen in the findings from the LTWM task. Expert pilots outperformed novices in meaningful situation re-creation. However, novice and expert pilot nonmeaningful situation re-creation were equivalent. This is consistent with the hypothesis that expert pilots can create and use retrieval structures (a function of LTWM).

It could be argued that the expert pilots were better able to chunk the instrument values. However, it is believed that STM is time sensitive (e.g., Baddeley & Hitch, 1974). Therefore, it is likely that the processing and storage (counting backward by three) would replace the contents of STM. It could also be argued that the recall of the meaningful situation is a result of activating specific patterns stored in LTM. However, because there are an infinite number of situations possible, it is unlikely that a pattern with specific values would reside in LTM.

One final piece of evidence suggesting that WM and LTWM coexist as distinct constructs comes from the hierarchical regression analyses. LTWM skill predicted complex task performance for expert pilots. This was an expected finding. If pilots have the ability to use retrieval structures to re-create situations, it is reasonable to assume that they can use retrieval structures to enhance performance. No additional predictors were found. If WM capacity was a function of LTWM, WM capacity would be expected to be a predictor of complex task performance.

In conclusion, this study provides further evidence of the roles WM capacity and LTWM skills serve in supporting complex cognitive task performance.

References

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., & 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.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 641-653.

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

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

de Groot, A. D. (1965). Thought and choice in chess. The Hague, Netherlands: Mouton.

Doane, S. M., McNamara, D. S., Kintsch, W., Polson, P., & Clawson, D. M. (1992). Prompt comprehension in UNIX command production. *Memory & Cognition, 20,* 327-43.

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. Schvaneeldt, S. Dumais, S. Lindsay, & M. 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*(1), 32-64.

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.

Ericsson, K. A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211-243.

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

Hartel, C. E. J., Smith, K., & Prince, C. (1991). *Defining air crew coordination: Searching mishaps for meaning.* Paper presented at the Sixth International Symposium on Aviation Psychology, Columbus, OH.

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

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

Salas, E., Prince, C., Baker, D. P. & Shreshta, L. (1995). Situation Awareness in team performance: Implications for measurement and training. *Human Factors, 37*(1), 123-136.

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*(1), 4-27.

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

Sohn, Y. W., & Doane, S. M. (2003). Role of working memory capacity and long term working memory skill in complex task performance. *Memory & Cognition, 31*(3), 458-466.

Acknowledgments

This research was supported by a grant from the Office of Naval Research (ONR) #N000140210898. The authors wish to thank ONR Program Manager Dr. Tanja Blackstone for her support of this effort. The authors also wish to thank LTjg. Phil Fatolitis, USN of Navy Submarine Medical Research Lab for his data collection effort and Mr. Tommy Sledge of Delta State University for his pilot recruitment efforts.