Image Cover Sheet

CLASSIFICATION	SYSTEM NUMBER 510027
UNCLASSIFIED	

TITLE

MENTAL MODEL RESEARCH: FROM THEORY TO OPERATIONAL APPLICATION

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to:

PARIS
25-29 MAY 98
34th iamps
international applied military psychology symposium

PROCEEDINGS

Hosted by the
Centre d'études en sciences sociales de la défense
24, rue de Presles
75015 PARIS

DCIEM No 98-P-40

Mental Model Research: From Theory to Operational Application

Megan M. Thompson, Ph.D.,
Defence and Civil Institute of Environmental Medicine,
Toronto, CANADA

Between 1980 and 1993 aircraft accidents destroyed seven Canadian Forces CC-130 Hercules and resulted in forty-four fatalities. An inspection of the accident investigation boards revealed that human error was a factor in most of these crashes. As a result, the Defence and Civil Institute of Environmental Medicine and Air Transport Group of the Canadian Forces were tasked to study the human factors relevant to improving the safety CC-130 operations. This paper summarizes the efforts of DCIEM scientists and Canadian Forces CC-130 trainers to design a simulator-based study creating a metric of effective crew performance (see Banks et al., 1996; Hendy et al., 1997, summaries of additional studies in this project).

Conceptual Basis: Mental Models

The conceptual basis that drove our efforts was the literature on mental models. Mental models are organized knowledge structures including objects, situations, events and the relationships between them represented in meaningful patterns (see Cannon-Bowers, Salas, & Converse, 1993; see also Rouse and Morris, 1986. They encompass past, present, and future flight parameters, goals, and considerations and "directly shape … responses" (Hendy, Laio & Milgram, 1997). Well-developed mental models are thought to lead to more efficient information processing, to decreased time pressure and workload, and to better performance (Hendy, et al, 1997; Hendy & Ho, 1997). Developing and maintaining a mental model become even more complicated when a task is to be completed by a team of individuals. At this level individual mental models need not be identical, but they must complement each other to increase team effectiveness (see Foushee, Lauber, Baetge, & Acomb, 1986; Kanki, Lozito, & Foushee, 1986; Kleinman & Serfaty, 1989; Leedom & Simon, 1993; McCallum, Oser, Morgan & Salas, 1989; Morgan, Glickman, Woodard, Blaiwes, & Salas, 1986; Oser, McCallum, Salas, & Morgan, 1989; Orasanu, 1993).

We used crewmember communications as an overt index of mental models. Although not a complete picture of an individual's mental model, communications do provide an adequate proxy of the important aspects of the individual's mental model and are a good measure of a formed crew's shared mental model. We focused our analyses upon Aircraft Commander (AC) communications because we expected ACs to have the biggest impact upon the performance of these hierarchically structured military aircrews. Indeed, it was the ACs who made the majority of statements during the flights.

Mental Model Metric: Accordingly, AC communications were coded according to the mental model domain and function they expressed (see Foushee, 1982; Foushee & Manos, 1981; Kanki & Foushee, 1989; Kanki, Lozito, & Foushee, 1989; Seigel & Federman, 1973 for important contributions in this area.). A mental model ***domain*** is the content area of an individual's statements concerning a flight. In our scenario the following domains were central: i) **aircraft systems**; ii) **procedures and checklists**; iii) **geography or air picture**; iv) the **mission**; and v) the changing **weather** picture. The number of mental model domains considered during a flight is indicative of the range of thought demonstrated by an individual. The ***function*** of a mental model refers to the meaning or purpose of a communication, and may vary in depth of thought. Functions progress in complexity from a simple **awareness** of the state of the world, through **cross-checking** (noting of deviations in any content domain and deviations of another crewmember's actions, planning, calculations, etc.) to the understanding **implications** and the development and implementation of plans (**preplanning**).

Additional Coding Categories. Prior research and our own preliminary observations has led to the inclusion of additional behavior categories into the metric. The first, **systems knowledge** was used by an AC demonstrated that he/she knew how to deal with a system malfunction, prior to consulting a checklist or manual. Two further categories relate most directly to resource management skills. **Task Prioritization** and **Crew Monitoring** (AC actively monitors other crewmembers' work and stress levels or progress on a specific demanding flight task) are essentially resource management skills. **Open-loop communication** (see Cannon-Bowers et al., 1993; Glickman, Zimmer, Monetro, Guerette, Campbell, Morgan, & Salas, 1978) indicated a failure to respond to another crewmember's statement or query. This category signals a lack of crew communication but is also a relatively good proxy measure for that crewmember's level of workload at that point in time. In essence, the crewmember simply does not have the resources to respond to all the inputs and demands at that moment. These mental model coding categories formed the basis of the communications analyses reported here.

METHOD

Subjects. 23 crews (Aircraft commander (AC), Co-pilot (CP), & Flight Engineer (FE)) undergoing normal continuation training, participated in the study. This number represents approximately one third of all CF Air Transport Group CC-130 crews. The instructor played the role of ATC, loadmaster, and any additional staff as required.

Procedure. After completing preparations for a local training flight each crew was asked to participate in a videotaped simulator session. After their consent, participant aircrew were told their mission was now an emergency medical evacuation (Medevac) between Trenton and Toronto. The mission was time critical: each crew had one hour to arrive in Toronto for a donor organ to be viable for transplant. Efforts were made to make the simulator session as realistic as possible through the use of a real-time scenario, as well as videotaped mission and weather brief employing actual CF operations personnel.

The CC-130 Flight Simulator Scenario: an *event set* approach (see Hammon, Seamster, Smith & Lofaro, 1993; Seamster, Hammon & Edens, 1995) was used to design the scenario. First, we specified the key behavioral dimensions of the mental models to be assessed, especially resource management skills, decision making and selected systems knowledge. We collaborated closely with CC-130 trainers in the design of a simulator scenario having event triggers specifically designed to elicit predetermined ACT behaviors at particular times in the mission. This collaborative effort resulted in a scenario that had precision in terms of our research needs but also possessed significant training value for aircrews. Thus, our research sessions were incorporated directly into the regular CC-130 training schedule.

The mission itself was a Trenton (Ont.) - Toronto (Ont.) (approx. 1 hr flying time), winter deteriorating weather, night mission. The weather brief indicated few available alternates (Trenton weather was expected to close shortly after departure). During the flight systems malfunctions and changes to ATC procedures were expected to take up a great deal of the crews' attention (see Table 1). A critical performance measure was whether the AC would continue to monitor more 'discretionary' aspects of the flight such as the mission status and the weather. Care was taken to ensure that the scenario did not present an unrealistic level of workload for the crews. All 23 crews completed the simulator flight, although with varying degrees of difficulty.

Expert Rating Assessments of Highly and Less Proficient Crews. Reviewing the videotapes of the simulator sessions, three pilots (one civilian, and two military CC-130 ICPs) provided proficiency rankings for each of the 23 crews based upon independent multidimensional assessments (see Clothier, 1993). Raters debated their assessments until agreeing upon groups of ACs representing higher (n=6) and less proficiency (n=6).

Communications from each of the twelve video tapes were coded by two coders (who different from the 'proficiency' raters and were not aware of the proficiency group assignment) according to the mental model categories outlined above (see also Table 2).

RESULTS

Preliminary Analyses: *T*- tests revealed that the highly proficient ACs had a greater number of hours on crewed aircraft (H: 3800.83 vs. L: 1819.17, $t = 2.35$, $p = .05$) and had spent more hours as ACs than did ACs in the less proficient group, although this latter result is only marginally significant (H: 2425.0 vs. L: 883.33, $t = 1.88$, $p = .11$).

Mental Model Analyses: Essentially our research asks: "Do communication patterns reliably differentiate ACs who are highly proficient from those who are less proficient?" The number of communications in each coding category made by the AC, divided by the total number of communications made by that AC provided a measure of the proportion of communications that fell within each of our coding categories.

Results of one-way ANOVAS, presented in Table 2, indicated that highly proficient (HP) ACs (relative to less proficient (LP) ACs) showed greater depth of thought, making a higher proportion of preplanning statements ($t = 1.73$, $p = .05$), especially concerning procedures and checklists ($t = 2.05$, $p = 0.04$), air picture ($t = 2.35$, $p = .02$), weather ($t = 1.55$, $p = .07$), and the mission ($t = 1.26$, $p = .12$). HP ACs also noted the implications of changes in wind direction ($t = 1.76$, $p = .05$). As anticipated, HP ACs demonstrated a greater range of thought. Their statements covered more mental model domains relevant to this flight scenario, but especially concerning the mission (awareness: $t = 1.40$, $p = .09$, total proportion of mission-related statements: $t = 1.45$, $p = .06$) and the weather (awareness: $t = 1.56$, $p = .08$, total proportion of weather-related statements: $t = 1.77$, $p = .06$). HP ACs tended to prioritize their tasks ($t = 1.34$, $p = .10$) and were more likely to be aware of, and concerned about the workload and stress levels of their crews (crew monitoring: $t = 1.41$, $p = .09$). Less proficient ACs engaged in greater cross-checking in terms of checklists/procedures ($t = 2.57$, $p = .01$), reflecting that LP ACs were less certain of aircraft systems and checklists. LP ACs also had less system knowledge without referring to checklist and manuals ($t = 2.40$, $p = .02$). Finally, LP ACs evidenced three times the instances of open loop communication (ACs ($t = 2.65$, $p = .01$).

Compensatory Behaviors. The copilots of LP ACs tended to make more awareness statements ($t = 1.82$, $p = .05$), and more directives ($t = 1.63$, $p = .07$) concerning systems malfunctions and issues. They also made more awareness statements ($t = 2.56$, $p = .01$), and were more proactive ($t = 2.14$, $p = .03$) regarding checklists and procedures. Similarly, the flight engineers of the LP ACs were more proactive concerning systems malfunctions ($t = 1.56$, $p = .07$), and engaged in more preplanning of checklists and procedures ($t = 1.51$, $p = .08$). There was also more open loop communication among the crews of the less proficient ACs (CPs: $t = 2.70$, $p = .02$; FEs: $t = 1.58$, $p = .08$).

DISCUSSION

This study was designed to measure effective crew performance. The results of this research indicated that several aspects of AC communication reliably distinguish highly proficient from less proficient ACs. Specifically, highly proficient ACs demonstrated a greater range and depth of thought, and were better able to keep in mind the more discretionary portions of the total flight mental model. Systems knowledge and resource management skills also tended to differentiate highly from less proficient ACs.

Further analyses also determined the behaviors that distinguished the copilots and flight engineers of the less proficient ACs. The proactive nature of most of these behaviors suggested that these crews tried to assist the LP ACs, perhaps even attempting to compensate for a lack of preplanning. Finally, LP ACs and their crews had a much higher incidence of open-loop communication, suggestive of less efficient crew level information exchange and a higher level of information or work overload for the less proficient ACs and their crews.

Operational Applications of Mental Model Research

Our measurement of mental model content and function holds the potential for a functional analysis of cockpit communications. Grounded in easily coded, observable behaviors, this approach to scenario design and crew analyses holds the potential for more quantifiable assessments and increased inter-rater reliability, serving as a more rigorous basis for on-line training and standards assessments.

This work is clearly preliminary in nature. Our future work includes i) including preflight briefings and post-mission debriefings as part of the metric; ii) expanding our efforts to include all crew members; iii) providing quantifiable markers for instructor check pilot use in establishing and maintaining CC-130 CRM standards; iv) determining the applicability of this assessment tool of crew resource management (CRM) behaviors on the CC-130 flightdeck (vs. exclusively simulator-based assessment); and v) providing quantifiable "value-added" assessments of CRM course modifications presently being developed at DCIEM (see Hendy & Ho, 1997). This metric, in combination with the event set approach to scenario design, should benefit both scientists and aircrew trainers in performing more accurate, consistent and standardized assessments of CRM performance via the non-intrusive yet systematic assessment of key observable behaviors indicative of complex decision making in ecologically valid simulations.

REFERENCES

Banks, R. B., Hendy, K. C., Fraser, W. F., Thompson, M. M., Jamieson, D. W. (1996). Human factors Study of CC-130 Operations. (DCIEM Report 96-R-66). Toronto, Ontario, CANADA

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In J. Castellan (Ed.), *Current Issues in Individual and Group Decision Making*. Hillsdale, NJ:Erlbaum, 221-246.

Clothier, C. C. (1991b). Behavioural interactions across various aircraft types: Results of systematic observations of line observations and simulations. *In Proceeding of the Sixth International Symposium on Aviation Psychology*, 332-337.

Foushee, H. C. (1982). The role of communications, socio-psychological and personality factors in the maintenance of crew coordination. *Aviation, Space,and Environmental Medicine*, *53*, 1062-1066.

Foushee, H. C., Lauber, J. K., Baetge, M. M., & Acomb, D. B. (1986). *Crew factors in flight operations III: The operational significance of exposure to short-haul air transport operations* (NASA Technical Memoradum 88322). Moffett Field, CA: NASA-Ames Research Center.

Foushee, H. C., & Manos, K. L. (1981). Information transfer within in cockpit: Problems in intracockpit communications. In C. E. Billings & E. S. Cheney (Eds.), *Information transfer problems in the aviation system* (NASA Technical Memoradum 1875, pp. 63-71). Moffett Field, CA: NASA-Ames Research Center.

Glickman, A. S., Zimmer, S., Montero, R. C., Guerette, P. J., Campbell, W. J., Morgan, B. B., & Salas, E. (1987). *The evolution of teamwork: An empirical assessment with implications* *for training*. NASA Technical Memoradum 87-016). Orlando FL: Naval Training Systems Center.

Hammon, W. R., Seamster, T. L., Smith, K. M., & Lofaro, R. J. (1993). The future of LOFT scenario design and validation. In the *Proceedings of the 7th Annual International Symposium on Aviation Psychology*. Columbus OH: That Aviation Psychology Laboratory, Ohio State University, 589-594.

Hendy, K. H., Thompson, M. M., Fraser, W. R., & Jamieson, D. W. (1997). Human factors of CC-130 Operations. Vol. 1: Final report of DCIEM/ATG study team. DCIEM Report 97-R-14, Toronto, CANADA.

Hendy, K. H. & Ho, G. (1997). Human Factors of CC-130 Operations Vol. 5.: Human factors of decision making, (DCIEM 98-R-18) Toronto, ON, Canada.

Hendy, K. C, Liao, J., & Milgram, P. (1997). *Combining time and intensity effects in assessing operator information processing load. Human Factors, 39(1), 30-47.*

Kanki, B. G., & Foushee, H. C. (1989). Communication as a group process mediator of aircrew performance. *Aviation, Space, and Environmental Medicine*, *60*, 402-410.

Kanki, B. G., Lozito, S., & Foushee, H. C. (1989) Communication indices of crew coordination. *Aviation, Space and Environmental Medicine*, *60*, 56-60.

Kleinman, D. S. & Serfaty, D. (1989) Team performance assessment in distributed decision making. In P. K. Gilson & B. Glodiez (Eds.), *Proceedings of the Interactive Networked Simulation for Training Conference.*

Leedom, D. K. & Simon, R. (1993). *Improving team coordination: A case for behavior-based training.* U. S. Army Institute.

McCallum, G. A., Oser, R., Morgan, B. B., & Salas, E. (1989). *An investigation of the behavioral components of teamwork.* In *Proceedings of the American Psychological Association.*

Morgan, B. B., Glickman, A. S., Woodard, E. A., Blaiwes, A. S. & Salas, E. (1986). *Measurement of team behaviors in a navy environment.* (U.S. Government Report NTSC TR-86-014) Orlando, FL: Naval Training Systems Center.

Orasanu, J. M. (1993). Decision-making in the cockpit. In B. G. Kanki, E. L. Weiner, & R. L. Helmreich (Eds.), *Cockpit Resource Management.* Orlando FL: Academic Press, 137-168.

Oser, R., McCallum, G. A., Salas, E., & Morgan, B. B. (1989). *Toward a definition of teamwork: An analysis of critical team behaviors* (89-004). Orlando, FL: Naval Training Systems Center.

Rouse, W. B. & Morris, N. M. (1986). On looking in the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*, 359-363.

Seamster, T. L., Hammon, W. R., & Edens, E. S. (1995). Specification of observable behaviors within LOFT/LOE event sets. In the *Proceedings of the 8th Annual International Symposium on Aviation Psychology.* Columbus OH: That Aviation Psychology Laboratory, Ohio State University, 613-618.

Seigel, A. I. & Federman, P. J. Communications content training as an ingredient in effective team performance. *Ergonomics*, *16*, 403-416.

Table 1. *System malfunctions, flightpath and weather updates in the CC-130 flight simulator scenario.*

Aircraft system malfunction 1:	On take-off landing gear will not retract (touchdown relay failure)
Aircraft system malfunction 2:	#2 EDHP light (pump malfunction)
Toronto ATC malfunction:	Toronto radar goes down and is on procedural control, the aircraft directed to Simcoe to hold
Aircraft System malfunction 3:	#4 Generator light (generator failure)
Aircraft System malfunction 4:	#4 Generator bearing light (generator bearing failure)
Approaching Toronto wind updates:	Wind on arrival runway 24 at Toronto approaches crosswind limits
Aircraft system malfunction 5:	#1 reduction gearbox failure

Table 2: *Pattern Of Results Of Mental Model Domains And Functions Among Highly And Less Proficient Aircraft Commanders And Pattern Of Results For Additional Coding Categories.*

DEPTH OF THOUGHT	<u>MENTAL MODEL FUNCTION</u>	< ------------ *RANGE OF THOUGHT* ------------ > <u>MENTAL MODEL CONTENT</u>				
		Mission	**Geography (Air Picture)**	**Systems**	**Procedures & Checklists**	**Weather**
	Awareness	H > L			H > L	H > L
	Cross-Checking		H > L		*L > H*	
	Implications					H > L
	Preplanning	H > L	H > L		H > L	H > L

	SYSTEMS KNOWLEDGE	H > L
RESOURCE MANAGEMENT	CREW MONITORING	H > L
	TASK PRIORITIZATION	H > L
	OPEN-LOOP COMMUNICATION	*L > H*

KEY: H>L = High Proficiency ACs made a greater proportion of these statements relative to Low Proficiency AC's
Note only statistically significant or marginally significant results are reported.

510027