Image Cover Sheet

CLASSIFICATION UNCLASSIFIED	**SYSTEM NUMBER** 510016

TITLE

ASSESSING MENTAL MODEL CONTENT AND FUNCTION: A METRIC OF EFFECTIVE CREW PERFORMANCE

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to:

DCIEM NO. 97-P-2

Proceedings of the

NINTH INTERNATIONAL SYMPOSIUM ON AVIATION PSYCHOLOGY

April 27-May 1, 1997

Sponsored by

The Department of Aviation
The Aviation Psychology Laboratory
The Ohio State University
Columbus, OH 43210

The International Journal of Aviation Psychology

The Association of Aviation Psychologists

Edited by

R. S. Jensen and L. Rakovan

Richard S. Jensen, General Chair
Lori Rakovan, Technical Chair

VOLUME 2

DCIEM No. 97-P-2

ASSESSING MENTAL MODEL CONTENT AND FUNCTION: A METRIC OF EFFECTIVE CREW PERFORMANCE

Dr. Megan M. Thompson, Dr. David W. Jamieson, and Mr. Keith C. Hendy
Defence and Civil Institute of Environmental Medicine
Toronto, Ontario, CANADA

INTRODUCTION

A key component of efficient decision making is the quality of the mental model that an individual holds and, in a team environment, the ability to effectively communicate that mental model. This research details the development of a generic measure of effective crew performance. Our approach incorporated the central components of the Information Processing/Perceptual Control Theory (IP/PCT) Model (Hendy, 1995) with the existing literature on mental models, team performance, and crew resource management. The resulting metric was refined and partially validated in a simulator-based medical evacuation mission. Moreover, the methodology used in this study allowed us to determine important diagnostic information about the characteristics that distinguish highly from less proficient aircrew. In order to identify these characteristics, check pilots rated videotapes of each of 23 crews on their proficiency, choosing 6 highly, nd 6 less proficient crews that were included in further analysis. Crewmember communications in hese 12 tapes were then coded according to mental model categories detailed below.

Mental Models

Mental models are organized knowledge structures hat include objects, situations, events and the elationships between them represented in meaningful patterns (see Cannon-Bowers, Salas, & Converse, 993; see also Rouse and Morris, 1986). In essence a mental model is 'the what, who, why, and the how of a task'. A mental model may encompass past, present, and future flight parameters, goals, and onsiderations. A mental model "directly shapes the operator's responses and determines the potential to perform in accordance with system or task demands" (Hendy, 1995, p. 4). Well-developed mental models re thought to lead to more efficient information processing, to decreased time pressure and workload, nd to better performance. Mental models become ven more complicated when a task is to be completed y a team of individuals. Researchers in the area uggest that it is the overlap in shared mental models that is chiefly responsible for the consequent effectiveness or the lack of effectiveness of a team (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)

Following from the work of Foushee and colleagues (Foushee, 1982; Foushee & Manos, 1981; Kanki & Foushee, 1989; Kanki, Lozito, & Foushee, 1989; see also Seigel & Federman, 1973), the raw data we used here were the communications each crewmember uttered through a simulator flight. Although not a complete picture of the entirety of an individual's mental model, it is expected that communications do provide an adequate proxy of the important aspects of the individual's mental model and certainly provide a good measure of a crew's shared mental model. Crew member communications were coded according to the mental model domain and function they expressed.

A mental model ***domain*** refers to the discernible content areas of an individual's thoughts concerning a flight. Although these may vary depending upon the flight, in our scenario the following domains were central: i) aircraft systems; ii) procedures and checklists; iii) geography or air picture; iv) the specific mission; and v) the changing weather picture. The ***function*** of a mental model refers to the meaning or purpose of a communication. 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**). This measurement of content and purpose amounts to a functional analysis of cockpit communications.

Mental model domains and functions can, of course, be considered together. Moreover, one can think of domain and function as reflecting the range and the depth of the mental model. For instance, the number of mental model domains considered during a flight is indicative of the range of thought demonstrated by an individual. Similarly, simple awareness statements, such as one indicating awareness of a system malfunction, would be classified as requiring less depth of thought than a statement noting the implications of a system malfunction, or a statement indicating preplanning in light of the implications of the system failure.

Additional Coding Categories. Prior research and our own preliminary observations also led to the inclusion of additional behavior categories into the measurement battery. The first category, **systems knowledge** was used only when aircrew demonstrated that they knew exactly and immediately how to deal with a system malfunction, prior to consulting any checklists or manuals. Two further categories relate most directly to resource management skills. **Task prioritization** requires no further explanation. **Crew monitoring** refers to instances in which a crew member actively and closely monitors other crewmembers' work and stress levels or their progress on a specific demanding flight task. A final category, **open-loop communication** (see Cannon-Bowers et al., 1993; Glickman, Zimmer, Monetro, Guerette, Campbell, Morgan, & Salas, 1978), was used in instances in which a crewmember failed to respond to another crewmember's statement or query. This category is important as it certainly signals a lack of crew communication. It 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.

METHOD

Subjects. 23 crews, consisting of an aircraft commander (AC), Co-pilot (CP), and Flight Engineer (FE) undergoing normal continuation training participated in the study. This is a significant number as it represents approximately one third of Canadian Forces Air Transport Group CC-130 crews. During the flight task, the simulator instructor played the role of ATC, loadmaster, and any additional staff as required. The experimenter flew all simulator sessions but did not interact with the crews during the flight itself.

Procedure. Each crew had completed preliminary preparations for a 'local trainer' flight in the simulator. Just prior to the beginning of the simulator session the crew was brought into the briefing room, told the general purpose of the study and asked to participate. Participants were told that, with their permission, their simulator session would be videotaped for later review at DCIEM, were assured of the confidentiality of their videotapes, that they might decline from participation at any time. After their consent, participant aircrew were told that the nature of their mission had been altered. Instead of the local trainer they had expected to fly, they would fly a medical evacuation (Medevac). The mission was time critical: crews were briefed that they had approximately 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 videotaped mission and weather brief employing operations personnel from CFB Trenton.

The CC-130 Flight Simulator Scenario. Constructed with the cooperation of CC-130 trainers to have significant training value for the participating aircrew, the scenario was devised to test several aspects of the crews' mental model, especially selected systems knowledge and resource management skills. The flight occurred in winter and was a night, poor weather, Trenton Ontario to Toronto Ontario IFR mission. The weather brief indicated that there were few alternates available, including Trenton where the weather was expected to close in soon after departure. During the mission a number of aircraft system failures were simulated and there were changes to ATC procedures and weather that required preplanning (see Table 1). While the scenario was busy, care was taken to ensure that it did not present an unrealistic level of workload for most crews. Systems malfunctions were expected to take up a great deal of the crews' attention. In fact, a critical measure was the ability of the AC to continue to monitor and assess more 'discretionary' aspects of the flight such as the mission status and the weather. All 23 crews completed the simulator flight, albeit with varying degrees of difficulty.

Expert Rating Assessments of Highly and Less Proficient Crews. We adapted the Aircrew Observation and Evaluation Scale (Clothier, 1991b) as the metric used by our subject matter experts to evaluate the performance of crews. Three experienced pilots (one civilian, and two military CC-130 ICPs) provided proficiency rankings for each of the 23 crews based upon independent multidimensional assessments of each videotape (e.g., assessments of safety concerns, decision making, and workload management).

We expected that in hierarchically structured teams such as flight deck aircrew that crew effects would be largely driven by the behavior of the AC. Indeed, it was the ACs who made the majority of statements throughout the flights. Thus, 6 highly-proficient crews and 6 less-proficient crews were selected based upon the three raters' consensual assessments of AC proficiency. To make this final selection, the three raters met as a group and reviewed their relative scorings for the 23 crews. The crews that all three raters had independently selected as representative of the higher and lesser proficiency groups were automatically included in our test group of crews. The raters then debated their assessments of the remaining crews to achieve a consensus regarding the crews that were to be included in the highly and less proficient groups. This process yielded two groups which the raters agreed on average represented a more proficient group and a less proficient group. As one might expect, t- tests revealed that the highly proficient ACs had a greater number of hours on crewed aircraft (high vs. low mean hrs. = 3800.83 vs. 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 statistically significant (high vs. low mean hrs. = 2425.0 vs. 883.33, $t = 1.88, p = .11$).

Communications from each of the twelve video tapes were coded by two independent coders (who were different from the raters of 'proficiency' and who were 'blind' to the proficiency group assignment of the crews) according to the mental model categories outlined above (see also Tables 2).

RESULTS

The specific unit of analysis used here was the number of communications in each coding category made by the AC, divided by the total number of communications made by that AC, providing a measure of the proportion of communications that fell within each of our coding categories. Essentially we asked the questions: "Out of all the communications (statements, commands, questions etc.) made by an individual, what proportion of statements reflect each of our categories?" and, more importantly, "Does the pattern of these communications reliably differentiate highly from less proficient ACs?"

To make this determination, we conducted a series of one-way ANOVAS. As the results presented in Table 2 indicate, the overall pattern of results substantiated our hypotheses. Highly proficient ACs (relative to less proficient ACs) demonstrated a greater depth of thought, as evidenced by a higher proportion of preplanning communications during the simulator flight ($t = 1.73$, $p = .05$), especially concerning procedures and checklists ($t = 2.05$, $p = 0.04$), geography or air picture ($t = 2.35, p = .02$), the weather ($t = 1.55$, $p = .07$), and the mission ($t = 1.26, p = .12$). Highly proficient ACs were also more likely to note the implications of changes in wind direction ($t = 1.76$, $p = .05$). Also as anticipated, highly proficient ACs demonstrated a greater range of thought. Their statements encompassed a greater number of the mental model domains relevant to this flight scenario, but most particularly concerning the mission (awareness: $t = 1.40$, $p = .09$, total proportion of statements concerning the mission: $t = 1.45, p = .06$) and the weather (awareness: $t = 1.56$, $p = .08$, total proportion of statements concerning the weather: $t = 1.77, p = .06$). These results suggest that highly proficient ACs were better able to keep in mind these more discretionary portions of the total flight mental model.

Our results also indicated that less proficient ACs engaged in greater cross-checking in terms of checklists/procedures ($t = 2.57$, $p = .01$). This result simply reflects the fact that less proficient ACs were less certain of the relevant aircraft systems and related checklists. This result is further substantiated by the fact that they demonstrated less system knowledge without referring to checklist and manuals ($t = 2.40$, $p = .02$). Thus, they simply had less system knowledge at their fingertips.

Results concerning resource management skills also differentiated the highly from the less proficient ACs. Highly proficient ACs showed a tendency to prioritize their tasks ($t = 1.34, p = .10$) and were more likely to engage in crew monitoring, that is, to be aware of, and concerned about the workload and stress levels of their crews ($t = 1.41$, $p = .09$). Finally, less proficient ACs showed a greater proportion of open-loop communications than did highly proficient ACs ($t = 2.65$, $p = .01$). Indeed, less proficient ACs evidenced three times the instances of open loop communication, suggesting, as expected, less efficient information exchange at the crew level and a higher level of information or work overload for the less proficient ACs.

Compensatory Behaviors. We also analyzed the communications of the other crewmembers of the highly and less proficient ACs (i.e., CPs and FEs). Overall, we saw an interesting pattern of behaviors emerge for these crews. Specifically, the copilots of less proficient ACs tended to make more awareness statements ($t = 1.82, p = .05$), and attempted to take a more directive ($t = 1.63, p = .07$) role concerning systems malfunctions and issues. Furthermore, the copilots of the less proficient ACs made more awareness statements ($t = 2.56, p = .01$), and took a more proactive role ($t = 2.14, p = .03$) regarding checklists and procedures. Similarly, the flight engineers of the less proficient ACs tended to take a more proactive role concerning systems malfunctions ($t = 1.56, p = .07$), and to engage in more preplanning concerning checklists and procedures ($t = 1.51, p = .08$). Perhaps most descriptive of communication problems, there were greater also instances of open loop communication among the crews of the less proficient ACs (CPs: $t = 2.70, p = .02$; FEs: $t = 1.58, p = .08$).

DISCUSSION

The communication measures developed and tested here were designed to capture known and hypothesized differences in crew efficiency. Their validity was demonstrated in a known-groups design (i.e., a high versus less proficient AC comparison). The measurement battery presented here is reliable, capable of yielding scientifically defensible results based on theory, and is applicable to a wide range of operational issues of concern to the aviation industry.

The methodology we selected also allowed us to begin to investigate important differences between highly and less proficient aircrew behavior. Highly proficient ACs possessed strong systems and procedural knowledge. They were more likely to demonstrate a superior range and depth of thought concerning important aspects of the flight and were more able to address discretionary aspects such as weather and the mission. Highly proficient ACs also demonstrated greater resource management skills at both the team and task levels. Proficient ACs facilitated teamwork because their communications maximized the planning of flight-related tasks and goals. Conversely, less proficient ACs had less knowledge of the aircraft systems. This likely increased the overall workload of the AC and the rest of the crew. Less proficient ACs showed evidence of work or information overload being less likely to respond to the questions and statements of their crews (an indirect indicator of workload). Just as importantly, less proficient ACs also demonstrated less range and depth of thought, engaging in less preplanning than did their more proficient counterparts, focusing primarily upon systems-related, procedural checking and rechecking of information. It may be that the lack of systems knowledge simply saturated the mental capabilities of the less proficient group of ACs: they simply did not have additional mental resources to address aspects of the flight that were more discretionary (a restricted opportunity hypothesis). Alternatively, these findings may also reveal that those ACs deemed less proficient simply do not typically demonstrate great range or depth of thought (a restricted capacity hypothesis). The present design does not determine whether restricted opportunity or restricted capacity is at the root of our findings.

With respect to the theoretical foundations of our work, our results indicate that more proficient ACs have better articulated mental models as evidenced by their ability to quickly identify and correct aircraft system errors, to understand the consequences and implications of flight anomalies, and to preplan the remainder of the flight in light of these implications. In effect, high proficiency ACs are better dynamic decision makers and better purveyors of information to their crews. Highly proficient ACs may use preplanning to both decrease the time taken to process decision-relevant information and increase the time available to devise, coordinate, and action plans. Our results also begin to illustrate the interactive and systemic or dynamic nature of crewwork. We found some evidence that the crews of the less proficient ACs crews evidenced behaviors that appeared to be compensatory in nature. Importantly however, these attempted behaviors did not fully compensate for the deficits of the less proficient ACs.

The behavioral and communications-based distinctions illuminated here are particularly relevant to training, by identifying those positive behaviors that should be particularly highlighted and modeled in the training system, as well as those specific behaviors that contribute to ineffectiveness and less safe practices among aircrew.

REFERENCES

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.

Hendy, K. C. (1995). *An information processing approach to crew resource management*. North York ON, Canada, Defence and Civil Institute of Environmental Medicine.

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.

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, Toronto is on procedural control, the aircraft is directed to Simcoe to hold
Aircraft System malfunction 3:	#4 Generator light (generator failure)
Aircraft System malfunction 4:	#4 Generator bearing light (bearing failure)
Approaching Toronto wind updates:	Wind on arrival runway 24 at YYZ 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*

<table>
<tr><td></td><td></td><td colspan="5"><------------ R A N G E ------------></td></tr>
<tr><td rowspan="5">(less)
D
E
P
T
H
(more)</td><td><i>Mental Model Content</i>
<i>Mental Model Function</i></td><td>Mission</td><td>Geography (Air Picture)</td><td>Systems</td><td>Procedures & Checklists</td><td>Weather</td></tr>
<tr><td>Awareness</td><td>H > L</td><td></td><td></td><td>H > L</td><td>H > L</td></tr>
<tr><td>Cross-Checking</td><td></td><td>H > L</td><td></td><td><i>L > H</i></td><td></td></tr>
<tr><td>Implications</td><td></td><td></td><td></td><td></td><td>H > L</td></tr>
<tr><td>Preplanning</td><td>H > L</td><td>H > L</td><td></td><td>H > L</td><td>H > L</td></tr>
</table>

SYSTEMS KNOWLEDGE	**H > L**
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.

#510016