MEASURING C-5 CREW COORDINATION PROFICIENCY IN AN OPERATIONAL WING

V. Alan Spiker
Anacapa Sciences, Inc.
Santa Barbara, CA

Steven J. Tourville
Lockheed Martin Information Services
Albuquerque, NM

MAJ John Bragger and **TSG Duane Dowdy**
709 Airlift Squadron
Dover AFB, DE

Robert T. Nullmeyer
Warfighter Training Research Division
Air Force Research Laboratory
Mesa, AZ

ABSTRACT

Air Force Instruction (AFI) 11-290 requires that all operating units use proficiency data to measure the effectiveness of their cockpit/crew resource management (CRM) training programs. In response, the 512th Airlift Wing, Dover AFB and the Air Force Research Laboratory (AFRL) developed an approach whereby the Wing supplied qualified data collectors while AFRL developed process and performance instruments, "calibrated" Wing observers in the use of the instruments, and analyzed the data. Data were collected from 16 C-5 aircrews during a challenging, simulated nighttime airlift mission involving poor weather, post-takeoff landing gear malfunctions, and eventual engine failure. Building on AFRL's established methodology with the MC-130P, two expert observers independently rated CRM proficiency and mission performance using behaviorally anchored, C-5 specific scale elements. A significant correlation (r=.58) was obtained between rated overall CRM proficiency and mission performance, extending the validity of AFRL's approach to another weapon system and mission. The study also yielded a wealth of qualitative data capturing the specific CRM behaviors of successful aircrews (e.g., pilots and flight engineers directly interact to mission plan and solve in-flight problems). Data from the study were briefed to wing leaders who have already implemented the study's major recommendations as a set of training initiatives to improve the mission performance of all aircrews. The study demonstrates that, with nominal outside research support, an operational Wing can establish a valid CRM proficiency measurement program. Lessons learned from this research can be applied across major commands to ensure that all units are able to comply with the CRM proficiency data requirements of the AFI 11-290.

ABOUT THE AUTHORS

V. Alan Spiker is a principal scientist at Anacapa Sciences, Inc., Santa Barbara, CA. He is responsible for human factors research with advanced technology and has conducted field studies on mission planning and C2 systems for all branches of service. He has a Ph.D. in Experimental Psychology from the University of New Mexico and has been with Anacapa Sciences since 1982.

Steven Tourville is a senior instructional designer with Lockheed Martin Information Services, Albuquerque, NM. He is an experienced MC-130P instructor navigator. This research was completed while he was a senior training analyst for Raytheon under contract with the Air Force Research Laboratory. He received his Ed.D in Adult Education from Nova Southeastern University.

John Bragger is a C-5 Instructor Flight Examiner pilot in the 709th Airlift Squadron at Dover AFB, DE as well as an A-320 Captain with Transmeridian Airlines. He is currently the CRM Program Manager for his operations group and has served as glass cockpit consultant to SRC Inc., designing simulation studies for the FAA. He has a BS in Aerospace Engineering from Purdue.

Duane Dowdy is a senior C-5 Instructor Flight Engineer in the 709th Airlift Squadron at Dover AFB, DE and a certified flight instructor with commercial and multi-engine ratings. His duties include teaching civilians to fly, instructing new C-5 flight engineers, and managing his operations group's CRM program. He has a BS in Professional Aviation from Embry-Riddle Aeronautical University.

Robert T. Nullmeyer is a research psychologist with the Warfighter Training Research Division of the Air Force Research Laboratory at Mesa, AZ (AFRL/HEA), where he has been since 1981. He has conducted research on training system design and evaluation, and simulator training effectiveness. He earned a Ph.D. in Experimental Psychology from the University of New Mexico.

MEASURING C-5 CREW COORDINATION PROFICIENCY IN AN OPERATIONAL WING

V. Alan Spiker
Anacapa Sciences, Inc.
Santa Barbara, CA

INTRODUCTION

This paper describes the results of a study that examined the relationship between crew coordination (CRM) proficiency and mission performance during a simulated training mission for the C-5 Galaxy aircraft. The work was performed under a partnership between the 512th Airlift Wing (512 AW) at Dover AFB and the Warfighter Training Research Division of AFRL.

Previous CRM Research

During the past 20 years, CRM has become a widely used component of aircrew training for both the civil and military communities (Gregorich&Wilhelm, 1993). In his landmark study of crew coordination, Ruffell Smith (1979) reported that the behaviors which most differentiate effective crews from weaker ones involve leadership, decision making, and resource management, thus establishing the need for training "softer" as opposed to more technically-oriented skills. While the call for a more interpersonal, attitudinal focus to CRM training was endorsed by the airlines as a way to "fix" pilots who resist receiving information from other crewmembers, application to the military environment was met with considerable skepticism (Spiker, Tourville, Silverman, & Nullmeyer, 1996). In the past few years, there has been a growing awareness that for CRM to have a useful place in the combat mission training curriculum, it must become operationally relevant, specific to the weapon system of interest, and have a target audience broader than just the aircraft commander—other crewmembers and the entire combat team (air traffic control, logistics, intelligence, etc.) must be included as well (Spiker, Silverman, Tourville, & Nullmeyer, 1998). To support training, it is necessary to identify observable CRM behaviors that can be practiced and reinforced during simulator and flightline training. It is important to delineate those CRM skills having a direct bearing on mission performance, as training time and resources are limited.

Despite recognition that a behaviorally-based approach to CRM is important for military training, there has been surprisingly little evidence to support a direct relationship with mission performance, particularly in a tactically-oriented context (Spiker et al., 1996). In a proof of concept study, Silverman, Spiker, Tourville, & Nullmeyer (1997) examined the relationship between CRM behaviors and mission performance for 11 Air Force Special Operations Command (AFSOC) MC-130P aircrews during the preparation and execution of a highly complex mission scenario using a high fidelity weapon system trainer (WST). Six crewmembers comprised each crew—two pilots, two navigators, a flight engineer, and a communications specialist. Their CRM behaviors and mission performance were rated independently by a former MC-130P navigator and an experienced researcher, respectively. To structure data collection, CRM was divided into five processes based on front-end analyses with tactical experts: time management, function allocation, situation awareness (SA), tactics employment, and command-control-communications (C3) (Spiker et al., 1996). These processes, along with various mission performance variables (e.g., chart and briefing quality, navigation accuracy), were rated during mission preparation and four execution phases: low-level navigation, aerial refueling, air drop, and airland. A strong, positive correlation was observed between overall ratings of CRM process and mission performance ($r=.86$), with mission planning quality also a significant predictor of mission performance ($r=.60$). Probing the data further, four of the five CRM processes were found to be significantly related to mission performance, the exception being C3. Since previous research has often equated CRM with quality and quantity of crew communications (Spiker et al., 1996), this finding may partly explain why CRM-performance links have not been more consistently observed.

The MC-130P study provides solid evidence that CRM is important in a tactical context. The findings replicated an earlier Air Force study (Povenmire, Rockway, Bunecke, & Patton, 1989) which used a similar methodology to show a direct relationship between CRM processes and mission performance in B-52 aircrews. Besides the quantitative results, the MC-130P study identified a number of CRM behaviors consistently exhibited by the most effective crews. These included: having more consideration of the "big picture" and viewing the crew as part of the larger tactical team (*SA)*; designating duties based on crewmember strengths (e.g., knowledge of automated planning systems) rather than crew position (*function allocation*); possessing greater tactical knowledge and more detailed communications exchanges while responding to threats *(tactical* *employment*); and greater time awareness throughout mission planning and execution (*time management*). Other behaviors emerged that did not fit into any predesignated CRM category, as the most successful crews: were focused on the mission with little extraneous socialization during planning or intercom chatter during mission execution; used very aggressive plans; had highly integrated hub-and-spoke communication systems (vs fragmented dyads and triads); and had a strong functional leader (aircraft commander or senior navigator) who weaved the crew together and maximized their crew resources (Spiker et al., 1998).

Lessons learned from the success of the MC-130P study encourage application of the natural observation methodology to other weapon systems, as long as six distinguishing features are present. These include: separate observers rate CRM process and performance; behaviorally anchored five-point rating scales; a "record by exception" philosophy that requires comments to explain other-than-average (3) ratings; a data collection instrument that rates the CRM of each crew position and the overall crew; breakdown each phase of CRM and mission performance into discrete, observable data elements; and explicit measurements of mission planning.

C-5 Mission Operations

The primary mission of the C-5 is to provide global strategic airlift of outsized cargo, such as helicopters, tanks, and mobile bridges. This aircraft carries fully equipped combat ready troops to any area in the world on short notice and provides full field support necessary to maintain a fighting force. The C-5's sophisticated communications and navigation systems make it virtually self-sufficient, and it can operate from unpaved airfields without ground support equipment. As the largest US airlifter, the C-5 can carry 3 times more cargo than a C-141 and 9 times more than a C-130. The crew consists of two pilots, two flight engineers (FEs), and two loadmasters, along with a relief crew for long-route deployments. Its massive airlift and efficient onload/offload capability have allowed the C-5 to play a major role in disaster relief and contingency operations.

Air Force Instruction 11-290

As the research base underlying the concepts and principles of effective CRM has expanded, the USAF CRM program office has taken strong steps to ensure their incorporation into USAF training operations. The program office issued a new Air Force Instruction (AFI) 11-290 (1 July 1998) which stipulates that all operational wings must have a robust program in place to collect, track, and evaluate CRM proficiency for all aircrew positions. The primary goals of Air Force CRM training are to maximize operational effectiveness and combat capability and preserve Air Force personnel and material resources. In turn, each wing's CRM training should develop: aircrew skills in recognizing and responding to the conditions that lead to aircrew error, and aircrew proficiency in CRM skills.

According to the AFI, CRM training should be delivered progressively to match the aircrew's training phases, beginning with introductory training, and encompassing formal training unit and mission-specific continuation training. CRM skills are to be integrated into flight briefings, debriefings, and training syllabi, and evaluated during initial qualification and recurring evaluations. The specific CRM skills and behaviors to be trained/evaluated are specified in AF Form 4031; its use is intended to spot weak CRM trends *before* incidents or accidents occur. Working with the Navy, the program office has identified six categories for any core CRM curriculum (see Table 1).

Table 1. Six CRM Categories Outlined in AFI 11-290

CRM Category	Core Definition
Situational Awareness	Knowledge and skill objectives to prevent the loss of SA, skills for recognizing the loss of SA, and techniques for recovering from the loss of SA
Crew Coordination/ Flight Integrity	Knowledge and skill objectives covering impact on aircrew performance of command authority, leadership, responsibility, assertiveness, conflict resolution, hazardous attitudes, behavioral styles, legitimate avenues of dissent, and team-building
Communication	Knowledge of errors, cultural influences, barriers (rank, age, experience, position). Skills encompass listening, feedback, precision and efficiency of communication with all members and agencies (i.e., crewmembers, wingmen, weather, ATC, intelligence)
Risk Management/ Decision Making	Includes risk assessment, the risk management process, tools, breakdowns in judgment and discipline, problem-solving, evaluation of hazards, and control measures
Task Management	Includes establishing priorities, overload, underload, complacency, management of automation, available resources, checklist discipline, and standard operating procedures
Mission Evaluation	Includes pre-mission analysis and planning; briefings; ongoing mission evaluation, and post mission debrief; specific tools and techniques to be used in operational and training missions

The AFI mandates that each Wing capture proficiency-related data using the above structure, but with components tailored to its own weapon system. The Wing must use data collectors trained in the CRM process and mission performance areas for that weapon system. Data collection instruments should be organized around these categories but be manageable in scope.

Study Objectives

The study's major objectives were to (1) determine if an operational Wing can collect valid CRM proficiency data without daily assistance by a research laboratory, and (2) identify key behaviors to guide future C-5 CRM training. Ancillary objectives included the need to: extend the naturalistic observation methodology to a strategic-oriented weapon system, organize data collection around the AFI's six CRM categories, and structure CRM and performance ratings around weapon-system specific data elements.

METHOD

Participants

Sixteen C-5 aircrews were observed during their quarterly continuation training at Dover AFB. Six aircrews were from the Air Force Reserve Command with the other ten crews drawn from the active duty wing. Because the C-5 WST does not have stations for the loadmasters, there were four training crew positions: aircraft commander (AC), copilot (CP), the flight engineer who operated the panels (FE-P), and the flight engineer responsible for scanning outside the cockpit (FE-SC). One contractor instructor was present throughout the training session, responsible for the two pilots and the two FEs. The instructor had been appraised of the study beforehand, and agreed not to intervene in any way except to perform their normal instructional duties. There was considerable variation in crew experience. Total C-5 flight hours ranged from 100 to 6800, with an average of 2079 hours for pilots and 2908 hours for FEs.

Weapon System Trainer

The C-5 WST reproduces the C-5 pilot, CP, and FE cockpit stations. Crew stations duplicate the aircraft flight station, including any controls, displays, and functional furnishings required for simulation training. The WST incorporates a six-axis, full-motion system and a wrap-around visual system capable of duplicating all phases of flight. Instructor stations are located within the cockpit assembly, each providing control and monitoring capability using touch-screen CRTs.

Scenario

Subject aircrews were observed during the preparation and performance of two mission-oriented simulator training (MOST) scenarios. The first leg entailed a seemingly routine airlift mission from Vandenberg AFB, CA to Travis AFB, CA. Following take-off, the right aft main landing gear fails to retract. The crew is to recycle the gear following their normal corrective procedures. During the re-extension, the nose landing gear (NLG) fails to extend and remains in an unsafe "intermediate" condition. Eventually, the crew is required to accomplish an NLG-up landing. A key consideration during this scenario is whether the crew is able to "mentally shift gears" and recognize they have a landing gear malfunction that should supersede the original problem. In addition, does the crew quickly recognize the gear is not going to extend and start reviewing the appropriate technical procedure? Does the crew attempt to resolve the problem enroute or wait until arriving at the destination? Do they maintain their SA with respect to the terrain? Do they monitor fuel consumption with the increased drag and maintain SA with respect to the terrain? Does the crew follow the correct technical procedures associated with the NLG-up and wheels-up crash landing checklists/procedures? These latter lists are confusing and can lead to errors unless the crew works together to prioritize their actions.

Leg two involves flying from Travis AFB, CA to Fallon Naval Air Station, NV, mostly at night. The challenges include an increase in wind turbulence outside Sacramento, encountering oncoming traffic before reaching cruise altitude, a 120 kt tailwind, a #3 engine overheat condition that requires shutdown of the engine, and the appearance of a #2 Thrust Reverser (TR) "Not Locked" light necessitating the shut down of that engine before slowing below 250 knots. Evaluation considerations include whether the crew: (1) correctly restarts #3 engine before shutting down #2, (2) is aware of the high terrain surrounding Navy Fallon, (3) considers 2-engine altitude capability in their decision, and (4) ensures the strong crosswind reported by Approach is within limits. Again, these decisions are aided to the extent the crew works together to prioritize actions, communicates clearly, and delegates tasks to the appropriate crewmember.

Data Collection Instruments

Separate instruments were used to collect CRM and mission performance data.

CRM Process. The C-5 CRM Process Worksheet is a 12-page instrument that rates CRM behaviors using the six AFI categories defined in Table 1. The instrument was organized around the four phases of the scenario: mission preparation, leg#1, leg#2, and debrief. Within each phase, the instrument designated three observable "elements" per CRM category for which presence/ absence/adequacy were to be assessed by the observer. These elements were tailored to the C-5 strategic airlift mission and had been identified during extensive front-end analyses conducted prior to the study. Figure 1 shows a segment that assesses the SA CRM category during mission preparation. The left column describes the three behavioral elements to be observed for that category (assess mission difficulty, prioritize mission events, identify impact of aircraft configurations). The right-hand column provides space to describe the specific behaviors that occurred, the conditions under which they occurred, and any problems or notable features associated with the behaviors. At the end of the mission phase, the observer assigns a 1-to-5 point rating for each crewmember for each category based on the descriptions of the corresponding three behavioral elements. The observer also assigns an overall CRM rating for the crew as a whole, for each phase and for the entire mission. These ratings are placed in the blocks at the upper portion of Figure 1. Each rating is behaviorally anchored, with half-point ratings permitted. On our rating scale, 1=Needs Improvement, 2=Adequate, 3=Standard, 4=Very Good, and 5=Outstanding.

phase, where each phase is subdivided into a series of ratable behavioral elements, three of which are shown in Figure 2. These elements were also identified during front-end analyses and customized for the C-5 mission. A total of 26 behavioral elements were rated across the four mission phases. Unlike the process instrument, where each crewmember receives a separate CRM rating, mission performance is only rated for the crew as a whole. As shown in Figure 2, the instrument provides space to assign both an element rating and a rating for the phase as a whole. The observer also assigns an overall performance rating for the entire mission. A five-point rating scale is used throughout.

Procedure

Process data were recorded on the CRM Worksheet by an experienced C-5 pilot from the 512 AW. Performance data were recorded on the Mission

Crew	1	2	3	4	5

P1	AC	1	2	3	4	5	F1	PN	1	2	3	4	5
	CP	1	2	3	4	5		SC	1	2	3	4	5
P2	AC	1	2	3	4	5	F2	PN	1	2	3	4	5
	CP	1	2	3	4	5		SC	1	2	3	4	5

C.SITUATION AWARENESS		
7. Do crewmembers ***provide an assessment of mission difficulty***? (e.g., Potential problem areas or limitations in the mission operations plan are identified; Minimum WX requirements determined, etc.)	Y N	
8. Do crewmembers ***prioritize mission execution events***? (e.g., Facts are used to devise solutions to mission planning problems; Options are developed in response to mission challenges, etc.)	Y N	
9. Do crewmembers ***identify the impact of various aircraft configurations*** as they execute the mission tasking? (e.g., discussion of GW, fuel management, minimum equipment requirements, and/or other issues, etc.)	Y N	

Figure 1. Situation Awareness/Mission Preparation Segment from the C-5 CRM Process Worksheet.

Mission Performance. A 10-page instrument was used by a second observer to record mission performance. An example segment is shown in Figure 2. Like the process worksheet, the performance worksheet is organized by Performance Worksheet by an experienced FE from the same wing. Prior to data collection, researchers from AFRL held extensive table-top review sessions with the two data collectors to customize the instruments for the

II. MISSION EXECUTION PHASE				
1 - Poor	**2 - Marginal**	**3 - Standard**	**4 - Very Good**	**5 – Exceptional**
a. Navigation Accuracy -- *Awareness of current location, Adherence to plan w/WX and ATC considered…*				**1 2 3 4 5**
1 - Crew is lost/disoriented; No adherence to planned routing; Unable to meet objective requirements	2 - Often deviates from/Is unsure of routing; Only able to meet objective requirements with difficulty	3 - Generally able to follow the planned routing; Several large off-track deviations performed to meet objectives	4 - Is able to adhere to planned mission routing; Is aware of position with respect to objec-tives at all times	5 - Is continually aware of position; Able to make adjustments to meet objective requirements
Explain:				
b. Detection and Diagnosis of emergencies/malfunctions…				**1 2 3 4 5**
1 - No detection, or very slow and inadequate diagnosis of system malfunctions	2 - Most system malfunctions detected; Not all diagnoses are adequate	3 - All system malfunctions detected and diagnosed correctly, and in adequate time	4 - All system malfunctions detected and diagnosed correctly in a timely fashion	5 - All system malfunctions detected quickly & correctly; diagnosis is quick/timely
Explain				
c. Checklist/Procedural Proficiency -- *Checklists accomplished in a timely, accurate manner…*				**1 2 3 4 5**
1 - Fails to complete many Emergency and/or normal procedures/checklists	2 - Most Emergency and/or normal procedures/checklists complete; Not timely or misses items	3 - All required Emergency and/or normal procedures/ hecklists complete	4 - All required Emergency and/or normal procedures/ checklists are completed on time; Covers all items	5 - All Emergency and/or normal procedures/checklists completed on time or early; Efficiently covers all items
Explain				

Figure 2. Example Segment from the C-5 Mission Performance Worksheet.

C-5 mission and "calibrate" each observer concerning the defining features of behaviors from each of the five scale values (i.e., what makes a "1" vs a "2," etc.). As the instruments were drafted, they were taken into the WST for "shake down" sessions in which the observers used the instruments to collect CRM process and performance data from contractor instructors serving as surrogate students. Based on the preliminary sessions, the content and format of each instrument was modified.

During actual data collection, sessions began in the mission briefing room where the training crews met with the contractor instructor to receive a mission briefing, after which the crews planned their mission. Trainees were told that the two observers were present to "conduct training research for AFRL," with no explicit mention of CRM. This general explanation was satisfactory for all subject crews. After asking each crewmember for his/her overall flight and C-5 experience, the two data collectors sat unobtrusively in the back of the room and observed the crews perform their mission preparation activities.

At the conclusion of mission preparation (one hour later), the crews entered the WST to perform the two legs of the mission. The data collectors made their observations from seats in the WST, with one data collector seated in the back of the WST, and the other seated next to the student FE. Both observers could see everything that went on in the WST and were wearing headsets to hear all internal communications. The data collectors observed on a not-to-interfere basis and completed their respective instruments while in the WST. Following mission execution, the data collectors accompanied the crew and instructor back to the briefing room where the mission debrief was held. They then recorded notable behaviors and rated behavioral elements for that final phase of the training mission. After data collection, the forms were sent back to AFRL where the data were summarized and subjected to statistical analysis.

Statistical Testing Considerations

The analyses reported here are based on Pearson product moment correlations between the ratings of CRM processes and mission performance provided by the two observers. All tests use the crew as the unit of analysis, where the degrees of freedom for the various t-tests are 14 (df = N–2 = 16–2). The 16 crews constitute a sizable portion of the crews receiving continuation training annually at Dover AFB (approximately 100), permitting the use of a finite-population correction coefficient (Winkler & Hays, 1975). The correction coefficient decreases the observed sample variance by the square root of (N-1)/(N-n), where N is the population size and n is the sample size. The critical t-values needed to achieve significance have been reduced by 10%, reflecting a 1.1 finite-population coefficient multiplier.

Conducting a large number of post hoc statistical tests will inflate the experiment-wise alpha level and increase the likelihood of a Type I error unless the nominal alpha level is "adjusted." In such cases, a Bonferroni adjustment is recommended in which the desired experiment-wise alpha level is divided by the number of tests that are performed in a given cycle of testing (Harris, 1994). In the following analyses, we will be performing two cycles of testing consisting of 5 and 18 tests, respectively. Using the Bonferroni adjustment, an experiment-wise alpha level of .05 will correspond to a nominal alpha of .01 (i.e., .05/5) in the first cycle and .002 (.05/18) in the second. Given the population correction described above, our reported correlations must reach .58 in the first cycle of tests and .62 in the second to achieve significance at the .05 level. Our initial tests will be two-tailed, as our hypotheses are bi-directional. However, should any of the initial tests prove significantly positive, the second round will be performed as one-way tests.

RESULTS

Quantitative Findings

Analyses of the principal rating data are depicted in Table 2. The left column lists the rating indices being correlated; the second column presents the correlations for the overall mission. The four right-hand columns depict correlations for the four phases—mission preparation, leg #1, leg #2, and mission debrief. The top row represents the correlations between the ratings of overall CRM and overall mission performance, followed by CRM-performance correlations for the individual mission phases. The remaining six rows follow the same logic, except that the CRM ratings correspond to each of the six CRM categories: mission evaluation (ME), task management (TM), SA, crew coordination (CC), C3, and risk management (RM).

Starting with the top row, we see that the primary correlation of interest, between the ratings of overall CRM process and overall mission performance, was substantial (r=.58) and statistically significant. The link between CRM and performance is fundamental, and while somewhat lower than for the MC-130P, is nonetheless consistent and gives us "permission" to probe the data further for more specific relationships in the data structure (Harris, 1994).

Table 2. Correlations Among Ratings of CRM Processes and Mission Performance.

Process/Performance	Overall Mission	MP	ME1	ME2	MD
Overall Process x Performance	.58**	.86***	.62**	.32	.38
ME Process x Performance	.53	.78*	.50		
TM Process x Performance	.62*	.58	.62*		
SA Process x Performance	.73*	.78*	.74*		
CC Process x Performance	.46	.82*	.64*		
C3 Process x Performance	.46	.74*	.42		
RM Process x Performance	.46	.81*	.58		

Bonferroni adjustment, cycle 1, 5 two-tailed tests: $^{**}p_{EW}<.05$, $p_{NOM}<.01$, critical r = .58; $^{***}p_{EW}<.01$, $p_{NOM}<.002$, critical r = .67
Bonferroni adjustment, cycle 2, 18 one-tailed tests: $^{*}p_{EW}<.05$, $p_{NOM}<.0028$, critical r = .62

The relationship between overall CRM and mission performance is depicted in Figure 3, which shows a scatterplot of the CRM-performance ratings for the 16 subject-crews. Note that the center cross-hairs divide each axis into positive and negative quadrants according to the scale mid-point of 3.0. In line with the positive correlation, we can see that crews who receive an above average rating in CRM were also rated better than average on mission performance. Perhaps the most telling feature of the relationship is that no crew occupied the upper left hand quadrant; i.e., received a substandard rating in CRM and an above standard rating in mission performance.

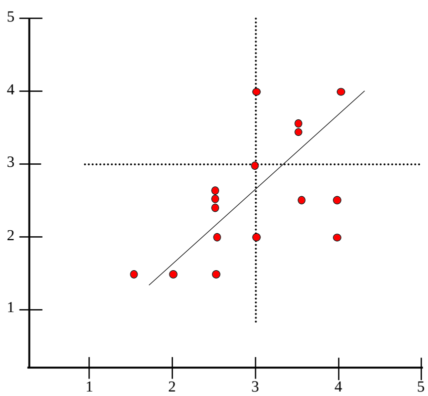

Figure 3. Scatterplot of Overall CRM process and Mission Performance.

Focusing on the upper row of Table 2, we probe the positive CRM-performance relationship further to determine if it holds up across all phases of the mission. As can be seen, CRM was significantly related to mission performance during mission preparation and the first execution leg, but not for the second leg or mission debrief. The degradation of the relationship is partially explained by the lower overall performance scores in the latter two phases (2.4-2.5). As a consequence, we may be seeing a restriction of range effect in which the reduced between-crew variability in these two phases could artificially reduce the correlation (Harris, 1994). However, interviews with participants following the sessions suggest that the lower scores for the second leg reflect the fact that most crews elected to "swap seats" at the conclusion of the first mission leg, i.e., the AC and CP reverse roles, as do the FE-P and FE-SC. Typically, crews started leg #1 with their strongest crewmembers occupying the AC and FE-P seats. Finally, the low score for mission debrief reflects the generally low output of comments from most crews during wrap-ups.

Turning to CRM processes, the lack of significance in the overall correlation obviated the need to explore the CRM relationships in the second leg or debrief phases; hence these cells have been grayed out. Two CRM processes, SA and TM, significantly predicted overall performance whereas the other four processes did not. Not only were these two processes significant predictors, their correlations were actually higher than the overall CRM correlation (.58). This finding contrasts that seen in the MC-130P study, where the overall CRM variable was the strongest predictor. The difference here may reflect the important roles that both processes play in successful completion of the airlift mission. Specifically, two of the most frequent errors in the study involved near misses with terrain and other aircraft (lack of SA), and spending too long attempting to correct the landing gear and engine failures before making an emergency landing (poor TM). Looking at the first two phases more closely, we see that all CRM processes (save TM) are important in effective mission preparation, a finding consistent with the MC-130P study. On the other hand, during the first leg, RM's effects become marginal while ME (is the route going as planned) and C3 (making all required calls) drop out as significant predictors. The latter result is very consistent with the MC-130P data, which once again suggests the dangers of equating CRM with communications (Silverman, et al., 1997).

To further explore the relationship among the individual CRM processes and mission performance, we conducted a multiple regression analysis (MRA) using the six CRM processes as predictors of mission performance. This approach lets us see which processes have the strongest association with mission performance when the other processes have been statistically controlled (Draper & Smith, 1981). In other words, the analysis helps pinpoint which variables uniquely predict performance versus those having high "shared variance" where the high intercorrelations among the processes themselves can be responsible for the association. We employed forward stepwise regression in which each CRM process was "stepped" into the analysis and retained if it achieved a certain degree ($p<.15$) of statistical association. Two CRM processes, SA and CC, were statistically significant predictors of mission performance, jointly accounting for 62% of the variance (i.e., $R^2=.62$), a very respectable level. While we expected SA to be in high, the appearance of CC was surprising, as TM had been a significant predictor in the correlational analyses. We then performed a second MRA, regressing each CRM process on overall CRM, to gain insight into the responsible mechanism. We found that CC was by far the best predictor of overall CRM, producing an R^2 of .87. The next best predictor was RM, which added only 6% to the explained variance in CRM. Thus, it appears that CC serves as a surrogate for overall CRM, which is not surprising since its definition can be taken as virtually equivalent to CRM (Hackman, 1993). Followup analyses showed this to be localized principally in the first mission leg, as that is where the links between SA/CC and performance were the strongest.

Two other analyses are noteworthy. First, comparison of mean overall performance for active duty and reserve crews showed them to be virtually identical (2.55-2.58), with the mean ratings for overall CRM almost as close (2.9-3.1). Despite the equivalence of the means, between-crew variability in mission performance among the reserve crews was more than 3-times that of their active duty counterparts.

A second analysis examined the impact of individual crewmember CRM proficiency on mission performance. In the MC-130P study, we found the CP's CRM score to be the best predictor of mission performance, primarily because the most successful crews found ways to utilize the typically under-tasked CP, particularly during mission planning (Silverman et al., 1997). In the present study, we performed an MRA by regressing the crewmember CRM ratings for each leg on the corresponding mission performance ratings. In the first leg, where the CRM-performance link was far stronger, only the FE-SC's rating achieved significance, producing a multiple R of .635 ($p<.02$). For the second leg, the FE-P's CRM rating best predicted mission performance ($R=.643$, $p<.01$). These findings underscore the importance of the FE in helping solve the multiple malfunctions that afflict the scenarios. The switch between legs in the FE position having greatest impact tracks the seat swapping that mentioned earlier, suggesting that the most experienced FE is the key individual in each crew.

Qualitative Findings – Performance

While analyses of the quantitative rating data substantiate a strong, positive relationship between CRM and performance, it is necessary to probe the data further, to determine *which* performance elements best predicted performance. Though not subjected to inferential testing, the following performance elements showed substantial correlations: quality of the mission briefing ($r=.43$) during mission preparation; completion of checklists during mission execution ($r=.83$ and .59 for legs 1 and 2, respectively), and accomplishment of mission events ($r=.79$ and .59 for legs 1 and 2). Having identified the most important performance elements, we then divided our subject crews into three groups, corresponding to whether they achieved high (4.0 or higher), medium (2.0-3.5), or low (below 2.0) average mission performance. We then performed a content analysis on the comment data from the above performance elements for each crew, and classified the results according to whether they were a high, medium, or low scoring crew. The results of that content analysis are shown in Table 3.

Examination of the table shows that most of the performance elements exhibit a roughly monotonic decline across the three columns, suggesting that the division of the three subgroups was appropriate. Importantly, this breakout provides a fairly concise summary of the key performance elements that most distinguished the 16 subject crews. For the Mission Brief, the upper rows of the table show that the better crews gave more extensive briefings that involved all four crewmembers; the weaker crews gave only a minimal brief with little involvement from one or more crewmembers.

Execution of the checklists turned out to be a very sensitive index of how well the individual crews performed. Once again, we see a progressive decline in performance across the three subgroups, as the best crews executed most or all of their checklists in a timely, accurate, and complete fashion. The weaker crews exhibited problems in completing certain checklists in a timely fashion, with the weakest crews still performing key checklists while descending below 10,000 feet. Across crews, we saw evidence for late, inaccurate, or incomplete checklists. For the best crews, not only were their checklists done correctly, they were also generally "ahead of the game," completing their checklists long before workload became high.

Table 3. Behavior Profiles of Key C-5 Mission Performance Elements

Performance Element	High Performing Crews (N=4)	Medium Performing Crews (N=6)	Low Performing Crews (N=6)
Mission Brief			
Level of Detail	3/4 w/ extensive or good detail, 1 w/ standard detail	1 w/ extensive coverage, 3/6 w/ standard detail, 1 w/ min. detail	3/6 w/ min. or no detail, 1 extensive, 1 standard (no ceiling or brake temp)
CM Involvement	3/4 w/ all CMs involved, 1 w/ Ps&FEs mod involved	3/6 w/ all CMs involved, 3 w/ 1 P min. involved	2/6 w/ no involvement, 2/6 w/. all CMs involved, 2/6 w/ Ps or FEs not involved
Checklists			
Timeliness	2/4 w/ all checklists done on time, 2/4 w/ key checklists (e.g., landing gear pro) done early	3/6 w/ all checklists done on or mostly on time, 3/6 w/some late: 2ENG, 3ENG procs, crash landing procs	2/6 completed on time, 2/6 had many late, 2/6 had some late: NLG down proc
Accuracy	4/4 were all or mostly (slight MWS errors) accurate, even did extra checks	2/6 all or mostly correct, 2/6 w/ extra or wrong steps (emerg ENG shutdown), 2/6 w/ minor errors (TR, NLG procs)	2/6 somewhat accurate; 4/6 w/ many errors (NLG extension, TR locked, slow ground reference speed)
Completeness	4/4 all complete, 2/4 w/ extensive emerg preps	2/6 all complete, 2/6 mostly complete, 1/6 w/ many not completed	2/6 all complete, 4/6 w/ many missing items (MWS, crashland,)
Events			
Not Accomplished	3/4 all accomplished; 1/4 minor error in alt. setting and throttle	3/6 all accomplished, 3/6 w/ some not accomplished (restart ENG#2, report fuel dump, nacelle bleed duct failure, missed climb rates)	0/6 all accomplished, 1/6 w/ minor error (2 ENG approach consider); 3/6 w/ many problems (ENG shutdown check, fuel consid, throttle)
Poorly accomplished	1/4 w/ no error; 3/4 w/ some error (fuel dump over airfield, ENG shutdown improperly)	3/6 w/ minor errors (incorrect trim, slow ground reference speed); 3/6 w/major errors (wrong MADAR proc, jettison fuel incorrectly, long time to declare emergency, wrong flat/slap proc, NLG retraction proc.)	2/6 w/ minor errors (slow to extend TRs, slow below KCAS w/ ENG out, wrong flap setting); 4/6 w/ major errors (dump fuel in holding pattern, dump fuel over airfield, near miss w/ traffic, near miss w/ terrain)

Event accomplishment was also indicative of performance, where we again see a systematic decline across the three subgroups. The four high performing crews accomplished all mission events with few or minor errors. The intermediate group's performance was characterized by failing to complete some key event (e.g., restart engine, report fuel dump) or failing to accomplish a larger number of minor events. For the poorest crews, these problems appear in even greater numbers, as all of the six crews experienced at least one major event problem. In fact, most of these crews experienced a number of problems, reflecting the cascading of time-intensive events built into the scenario.

Qualitative Findings – CRM Behaviors

As in the MC-130P study, the most interesting results involve identifying the notable or signature CRM behaviors exhibited by the most successful crews. To that end, we performed a content analysis on the observer comments for each data element in the CRM Worksheet for the four highest performing crews. Analysis yielded more than 150 notable concrete behaviors across the four mission phases. Space limits prevent a complete listing of these (but see Spiker in press), so the following is a partial summary of the most interesting findings.

Like the MC-130P, there are some very definite, concrete CRM behaviors that characterize the most effective C-5 crews. During mission preparation, the most effective crews aggressively challenged the assumptions in their mission tasking, conscientiously engaged in map study, decomposed the mission in a logical fashion, applied their own experiences, and very importantly, ensured direct and frequent information interchange between pilots and FEs.

During mission execution, the best crews prioritized events so that flying and troubleshooting duties were conducted sequentially, with the most important characteristic that all planning and troubleshooting was accomplished by the time the aircraft descended to 10,000'. The best crews always had a strong functional leader emerge early. He was usually the AC, who was adept at delegating duties on a task-specific basis, and allowed the senior FE to take functional charge of things while trying to fix the landing gear and engine-out problems, reverting control once the decision had been made to attempt a landing. The communication structure within the aircraft was also notable, as all of the successful crews employed a hub-and-spoke structure where there was continual, direct exchange of information between one of the pilots and one of the FEs. The less successful crews exhibited a more rigid structure in which FEs engaged in their own conversations as did the pilots. The downside of this structure was the excessive amount of "hot mike" chatter that took place between the two FEs, sometimes to the point where critical information from the pilot (e.g., altitude clearances) was missed. Indeed, there were many instances among the less successful crews where the "sterile cockpit" environment dictated by low level flight (i.e., below 10,000') was not maintained.

Mission debriefing revealed few notable behaviors, which was indicative of the overall low emphasis placed on CRM-oriented debriefing. To a large extent, most of the discussions centered on performance rather than CRM process issues. Nevertheless, successful crews were typically most critical of themselves, and were more likely to identify areas where they could improve. There was also a more open atmosphere of information exchange among these crews, reflecting a carryover from the environment that had been produced in mission preparation and maintained during mission execution.

CONCLUSIONS

Research Implications

This study adds a vital piece to a growing research base substantiating the robust ties between CRM proficiency and mission performance. By arming SMEs with behaviorally anchored data collection instruments while they make over-the-shoulder naturalistic observations, we are able to extract a rich array of quantitative and qualitative data that gauge the impact of CRM on performance, as well as pinpoint the specific behaviors which prove particularly effective. Importantly, the study demonstrates that *CRM proficiency can be assessed by operational C-5 personnel without requiring extensive continual involvement by laboratory scientists.* The six CRM categories outlined in the AFI, coupled with aircraft-specific data element descriptions, provide a sufficiently rich content that many important aspects of proficiency are captured. As in the MC-130P study, we see evidence of differential validity, where different aspects of CRM become relatively more important depending on the mission phase. In this study, high levels of SA and time management proficiency were the most powerful discriminators of strong versus weak performance.

At a general level, these results reinforce those from both the MC-130P study and a recent MH-53J helicopter study (Thompson et al., 1999). All three studies underscore the importance of mission planning for subsequent mission success, the necessity of having a strong functional leader, the relative importance of SA across multiple mission phases, as well as the phase- and crewmember-specific role that communications play in affecting mission performance. While previous studies of CRM have given communications a prominent role, the empirical facts in a tactical environment are that the nature and importance of communication depends upon which crewmembers are communicating and when.

The six AFI categories offer a good structure for measuring CRM proficiency, yet certain processes are underplayed. For example, we observed one crew that, while their CRM was rated uniformly high throughout the mission, they made some bad decisions in continuing to try to fix the NLG problem instead of making a timely decision to land NLG-up. They also did a poor job of time management, resulting in extreme task overload for the final several minutes. Because the CRM proficiency ratings reflect an accumulation of the past several hours, whereas mission performance was weighted heavily to the final few minutes of the flight, the relationship was inverted. Deleting the crew from the sample would increase the overall CRM-performance correlation from .58 to .71. Based on such instances, it may be prudent to consider revising the AFI to make time management and decision-making explicit categories rather than embedding them in others.

The data suggest that more work is needed in the area of Mission Debrief. There is only modest support that the better crews are also doing better debriefs. Actually, most of the debriefs were not very satisfying to trainees, as there is a paucity of CRM process discussion, with most of the focus on what happened rather than how or why. Future work should entail more detailed development of the mission debrief evaluation tool, identifying specific data elements as we have done for mission preparation and mission execution, and creating scales that more fully articulate the differential contributions expected of trainees and instructors.

Training Implications - Utilization of Findings by the Wing

At the conclusion of the study, the leaders of both participating wings were briefed on the quantitative and qualitative results. They immediately recognized areas in which CRM proficiency could be improved, and elected to move out smartly by tasking the research team to draft a series of training recommendations based on its findings. *These recommendations have already been formally received by both wings and are now being implemented into both flightline and simulator training.* A synopsis of these recommendations is listed below; a complete listing is provided in Spiker (in press). Collectively, they illustrate the power of a data-driven, research-based approach to training explicitly designed to improve performance.

<u>During Mission Preparation</u>, a review of CRM courseware and training should be conducted to improve the following areas: (1) teambuilding and barriers to communication; (2) quality of mission briefings; (3) direct sharing of information between pilots and FEs; (4) thorough and organized flight planning; (5) and the importance of sharing personal mission experiences.

During Mission Execution, a review of CRM courseware and training should be conducted to ensure (1) the importance of the sterile cockpit environment as successful crews invariably accomplished all non-flying activities while above 10,000'; (2) the use of Stan/Eval to check proper use and times for hot mike communications; (3) crewmembers effectively use all resources, such as jumpseat instructors, ATC, loadmasters, command post, and use appropriate levels and time for automation; (4) crews are given valuable SA information such as periodic location-in-space announcements, crew intentions and announcing priorities when things start to get busy; (5) the presence of a designated or functional central leader operating with a hub and spoke style of communications; and (6) risks are effectively managed by using guides, checklists, and other techniques to enhance terrain awareness and improve checklist responses.

During Mission Debrief, the training environment should be reviewed to ensure: (1) crews have leaders who take individual responsibility; (2) a non-threatening (non-blame) atmosphere is maintained; and (3) all instructors/evaluators become "facilitator qualified."

In summary, the CRM behaviors of the most effective crews are observable, trainable, and, to a certain extent, generalizable across weapon systems. We are presently working on a revised CRM course curriculum for the MC-130P weapon system at Kirtland AFB, to incorporate the CRM behaviors of the most effective crews (Tourville, Thompson, Spiker, & Nullmeyer, 1999). An important element of the course will be to develop techniques that encourage practice and reinforcement of CRM skills, behaviors, and processes in both simulator and flightline mission training sessions. We will be reviewing the qualitative results of the C-5 data to make a similar set of recommendations for the two wings at Dover AFB. The successful application of the CRM over-the-shoulder observation methodology to non-scientist SMEs suggests that it can be used by other USAF units in other weapon systems to assess CRM proficiency to ensure compliance with AFI 11-290 requirements.

REFERENCES

AFI 11-290 (1998, July). Cockpit/crew resource management program. Washington, DC: Headquarters US Air Force.

Draper, N.R.& Smith, H. (1981). *Applied regression analysis, second edition*. New York: John Wiley.

Gregorich, S. E., & Wilhelm, J. A. (1993). Crew resource management training assessment. In E. Wiener, B. Kanki, & R. Helmreich (Eds.), *Cockpit resource management* (pp. 173-198). San Diego, CA: Academic Press, Inc.

Hackman, J.R. (1993). Teams, leaders, and organizations: New directions for crew-oriented flight training. In E. L. Wiener, B.G. Kanki, and R.L. Helmreich, *Cockpit resource management*. (pp 47-69) New York: Academic Press.

Harris, R. J. (1994). *ANOVA: An analysis of variance primer*. Itasca, IL: F.E. Peacock Publishers, Inc.

Povenmire, H. K., Rockway, M. R., Bunecke, J. L., & Patton, M. W. (1989). *Evaluation of measurement techniques for aircrew coordination and resource management skills* (UDR-TR-89-108). Williams AFB, AZ: Air Force Human Resources Laboratory, Operations Training Division.

Ruffell Smith, H. P. (1979). *A simulator study of the interaction of pilot workload with errors, vigilance, and decisions* (NASA TM-78482). Moffett Field, CA: NASA-Ames Research Center.

Silverman, D.R., Spiker, V.A., Tourville, S.J., & Nullmeyer, R.T. (1997). Team coordination and performance during combat mission training. In *Proceedings of the 19th Interservice/Industry Training Systems and Education Conference*. Orlando, FL.

Spiker, V. A. & Campbell, J. L. (1993). *Functionial analysis of Combat Talon I mission planning: Opportunities for integration with advanced mission rehearsal technology* (AL/HRA-TR-1993-0058). Mesa, AZ: Aircrew Training Research Division.

Spiker, V. A. & Nullmeyer, R. T. (1995, June). *Functional analysis of the MH-53J Pave Low mission preparation process*. (AL/HR-TR-1994-0171). Mesa, AZ: Aircrew Training Research Division, Armstrong Laboratory.

Spiker, V. A., Tourville, S. J., Silverman, D. R., & Nullmeyer, R. T. (1996, November). *Team performance during combat mission training: A conceptual model and measurement framework* (AL/HR-TR-1996-0092). Mesa, AZ: Aircrew Training Research Division.

Spiker, V. A., Silverman, D.R., Tourville, S. J., & Nullmeyer, R. T. (1998, July). *Tactical team resource management (T^2RM) effects on combat mission training performance.* (AL/HR-TR-1997-0137). Mesa, AZ: Aircrew Training Research Division.

Spiker, V.A. (in press). *Summary of C-5 CRM Study*. (Technical Report 314-1.) Santa Barbara, CA: Anacapa Sciences, Inc.

Thompson, J.S., Tourville, S.J., Spiker, V.A., & Nullmeyer, R.T. (1999) Crew resource management and mission performance during MH-53J combat mission training. In *Proceedings of the 21st Interservice/Industry Training Systems and Education Conference*. Orlando, FL.

Tourville, S.J., Thompson, J.S., Spiker, V.A., & Nullmeyer, R.T. (1999). Enhanced crew resource management training for the MC-130P mission qualification course. In *Proceedings of the 10th International Symposium on Aviation Psychology*, Columbus, OH.

Tourville, S. J., Spiker, V. A., Silverman, D. R., & Nullmeyer, R. T. (1996). An assessment methodology for team coordination in combat mission training. In *Proceedings of the 18th Interservice/Industry Training Systems and Education Conference*. Orlando, FL.

Winkler, R. L. & Hays, W. L. (1975). *Statistics: Probability, inference, and decision.* New York: Holt, Rinehart, and Winston.