Hansberger J. T., Holt, R. W., & Boehm-Davis, D. A. (1999). Instructor/evaluator evaluations of ACRM effectiveness. Proceedings of the Tenth International Symposium on Aviation Psychology. The Ohio State University: Columbus, OH.

INSTRUCTOR/EVALUATOR EVALUATIONS OF ACRM EFFECTIVENESS

Jeffrey T. Hansberger, Robert W. Holt, Ph.D. & Deborah Boehm-Davis, Ph.D.
George Mason University
Fairfax, Virginia

ABSTRACT

A proceduralized form of Crew Resource Management (CRM) or advanced CRM (ACRM) was designed for a regional air carrier. The regional air carrier possessed two similar fleets in which one of the fleets received the ACRM training (experimental fleet) and the other did not (control fleet). A questionnaire was distributed to the instructor/evaluators (I/Es) in both fleets focusing on the comparative evaluation of the performance of both pilot groups. A principal components analysis revealed three principal components: 1) Workload Management, 2) Communication, and 3) Planning. For all three components, ACRM trained pilots were assessed higher than non-ACRM trained pilots.

INTRODUCTION

Resource management can be a critical component of complex job performance. Resource management is particularly important in domains involving complex and dynamic tasks with high costs for errors. Examples of teams in such domains are hospital surgery teams and aviation crews (Lauber 1984, Helmreich and Foushee, 1993, Wiener, Kanki, & Helmreich, 1993). As a result, resource management training programs have been developed for some of these domains.

In the aviation domain, crew resource management (CRM) has focused particularly on issues surrounding crew coordination and communication (Foushee & Helmreich, 1988). This work assumes that better crew coordination and communication will result in improved performance. There is evidence that suggests this is the case (e.g. Foushee & Manos, 1981). This type of research often focuses on measuring and changing basic attitudes toward aspects of CRM.

Salas and his colleagues (Salas, Bowers, & Cannon-Bowers, 1995) have approached this problem from another angle, focusing on developing methods for team training to improve performance. Salas et al. developed more precise measurement methods for identifying important teamwork behaviors and developed training for how and when to perform such skills or behaviors. Their work also demonstrated the relationship of improved team coordination and communication to performance (Stout, Cannon-Bowers, Salas, & Morgan, 1990).

Neither of these approaches has, however, changed the operating context for the task itself. Our approach differed from previous approaches in combining pilot training with a congruent structural change in flight operations. Based on identified carrier needs, the changes in training and operational context were centered around specific procedures designed to facilitate certain aspects of CRM.

The new CRM procedures were designed by identifying the safety concerns of the airline and the weaknesses of traditional CRM training. Key issues were determined using the results from the National Transportation Safety Board (NTSB) Safety Study (NTSB, 1994), data from Aviation Safety Reporting System (ASRS), and information from a survey and personal interviews of the airline's instructor/evaluators (I/Es).

The data from these three sources led to the development of three goals for the CRM training program:

1. reduce distractions to the pilot flying (PF) in both normal and abnormal situations;
2. increase structure in briefings to enhance the crew's performance on the first day together and improve information transfer; and
3. design checklists, the Quick Reference Handbook (QRH), and briefings to reduce workload and enhance decision-making skills, especially when crews would be fatigued, running late, or under high workload.

These goals were translated into actual procedures by a design team consisting of the airline's CRM coordinator (a pilot), a pilot from a major airline acting as a design consultant, an instructional designer, and researchers specializing in aviation, cognitive human factors and team research. The set of newly developed CRM procedures were named Advanced Crew Resource Management (ACRM).

The ACRM procedures were designed to target the three primary goals or areas as well as improve other related areas of crew resource management. Related to these three goals were the three primary topics ACRM was designed to improve: communication, situation assessment, and planning/decision making.

This research was part of a multi-method evaluation of the effectiveness for ACRM training. The method addressed here is focused on the Instructor/Evaluators' (I/E) evaluations of pilot performance for ACRM trained pilots and non-ACRM trained pilots.

Hansberger J. T., Holt, R. W., & Boehm-Davis, D. A. (1999). Instructor/evaluator evaluations of ACRM effectiveness. Proceedings of the Tenth International Symposium on Aviation Psychology. The Ohio State University: Columbus, OH.

METHOD

Design

Several different methods and approaches were implemented to evaluate the effectiveness of ACRM (Holt, Boehm-Davis, & Hansberger, 1999; Incalcaterra & Holt, 1999; Ikomi & Holt, 1999). The method addressed in this research focused on instructor/evaluator (I/E) comparative evaluations of pilot performance in the two fleets. I/Es were told to consider their experience in training pilots from each fleet for LOEs, fleet transitions, or Captain upgrade training. I/Es then made comparative judgments of the relative performance of ACRM trained and untrained pilots.

Subjects

All I/Es in a regional airline participated in this study. Only data from 19 I/Es that had direct experience evaluating pilots in both fleets were used for analysis. All these I/Es had observed and evaluated pilots from the ACRM and non-ACRM trained fleets and had received ACRM training themselves.

Materials

The same team that developed the ACRM procedures designed a nineteen-item fleet comparison questionnaire. Each item asked for a comparative evaluation of the ACRM trained pilots to the non-ACRM trained pilots. Nine items targeted the frequency of occurrence for particular behaviors and ten items targeted the quality of the behavior. The pilot groups were compared on a 5-point scale where a "5" response indicated the ACRM-trained pilots displayed the behavior much more frequently or with much better quality. A "1" response indicated the behavior was done much less frequently or with much worse quality. A "3" response indicated that ACRM-trained and untrained pilots performed the same for that item.

The items in the questionnaire were designed to address three content areas, 1) workload management, 2) communication, and 3) situation awareness. We hypothesized that a principal component analysis would confirm these three components as separate response dimensions. We further hypothesized that the majority of evaluations would favor the ACRM pilots.

Procedure

All the I/Es at the carrier completed the fleet comparison questionnaire prior to a recurrent rater calibration training session. I/Es were asked to consider all ACRM trained and non-ACRM trained pilots that they had evaluated in the past 6 months. Researchers were present to answer questions and resolve any ambiguities in the task or response format.

RESULTS

Response Components

A principal component analysis using Varimax rotation of the nineteen frequency and quality items revealed three distinct and orthogonal components. The first component contained eight items. A scale constructed by unit-weighting these items possessed a reliability of $\alpha = .90$ (see Table 1). The items making up this scale matched five items that were labeled workload management items a priori. This scale will be referred to as the workload management scale.

The second component was comprised of five items. A scale constructed by unit-weighting these items had a reliability of $\alpha = .85$ (see Table 2). Four of the five scale items matched the a priori communication item content. Therefore, this scale will be referred to as the communication scale.

The third component was comprised of six items. The unit-weighted scale of these six items had a reliability of $\alpha = .89$ (see table 3). Unlike the other two components, this component resembled a different construct than what was predicted a priori in the construction of the survey items. This scale appeared to be comprised of more planning type items than situation awareness items. Therefore, this scale will be referred to as the planning scale. The high reliabilities for all three scales suggest strong coherence and agreement among the items within each component.

Table 4 shows the intercorrelation between the three scales. All three scales showed a moderate degree of intercorrleation with each other ($r = .528$-$.590$, $p < .05$). This suggests the scales are related but still somewhat distinct from one another.

Tests of Perceived Fleet Differences

For each component, a one-sample t-test was conducted to analyze any potential differences from the baseline score of "3" ("same") and their direction (i.e., ACRM pilots better or non-ACRM pilots better). This test evaluated the I/Es perception of any differences between the two fleets on each scale.

The mean judgements for all three scales were significantly different from the baseline and in the direction of the ACRM trained pilots performing the items better (more frequently or with higher quality) See Figure 1. The workload management scale had a mean score of 3.57, which was significantly greater than 3.0 ($t = 4.85$, $p < .01$). The communication component had a mean score of 3.92, which was also significantly greater than 3.0 ($t = 7.53$, $p < .01$).

Finally, the planning component had a mean score of 3.62, which was significantly greater than 3.0 ($t = 4.62$, $p < .01$).

At an individual item basis, 95% (18 of 19) of the items were significantly higher than the score of 3.0. The single item that was not significantly different at the .05 level was the quality item, "Crews avoid distractions during critical phases of flight".

Post-hoc Tests

A one-way ANOVA for all the items involved showed there were significant differences among items ($F(18, 342) = 2.28$, $p < .01$). To isolate items for which there was noticeably more or less effect of ACRM, post-hoc one-sample t-tests were conducted. These tests compared the mean of each item to the grand mean and revealed that two items were significantly lower than the overall mean for all the items (overall mean = 3.68). The item mentioned above, "Crews avoid distractions during critical phases of flight," was significantly lower than the overall mean ($t = -2.47$, $p < .05$). The other quality item significantly lower than the overall mean was "Crews are organized and prepared in the cockpit" ($t = -3.32$, $p < .01$).

The one item that was significantly greater than the overall mean was the quality item stating "Crews formulate & communicate bottom lines & back-up plans for the abnormal situation" ($t = 2.10$, $p = .05$). These post-hoc items suggest that ACRM is particularly strong in facilitating the establishment of bottom lines and back-up plans, but may not help in reducing distractions and helping the crew be generally organized and composed.

DISCUSSION

ACRM procedures were designed to facilitate specific aspects of crew interaction and performance and increase safety in the cockpit. The I/Es observed the ACRM trained and untrained pilots in diverse settings: while pilots underwent upgrade training, during fleet transition training, or while conducting their annual evaluations. The I/E judgements were that the ACRM trained pilots performed significantly better than the non-ACRM trained pilots. All these items were designed to target behaviors in the cockpit that would potentially increase performance and safety for the crew.

It was predicted that the items would factor into three basic categories: 1) workload management, 2) communication, and 3) situation awareness. The workload management and communication components did emerge from the analyses but a different component formed as the third, namely planning. The positive I/E assessment of ACRM trained pilots and the three components found from the questionnaire support the initial design intentions for ACRM.

The emergence of planning as the third component found is not surprising as the new checklists, briefings, and abnormal procedures were all designed to facilitate better preparation and planning. In fact, one of the new additions for every abnormal procedure the ACRM trained pilots implement is a specific "preparation and planning" step.

The absence of the situation awareness component might be reason for concern. However, situation awareness might better be perceived as a more global construct that includes workload management, communication, and planning. Past research on situation awareness has described it broadly and situation awareness is often viewed as dependent on these types of components (e.g., Endsley, 1995). If situation awareness is indeed partially the product of workload management, communication, and planning, and ACRM improved each of these components, then it follows that situation awareness should also be improved by ACRM procedures. However, more research in the area of defining and directly measuring situation awareness in this context is needed to validate this inference.

The perceptions of the I/Es were overwhelmingly positive for ACRM training, especially in the area of communication. One weakness of this research is that the methodology relied on the perceptions of the I/Es. Even though the I/Es are professional evaluators and received regular training to improve their inter-rater reliability at this carrier, their assessments were still susceptible to human biases and judgement error (Hansberger & Holt, 1999).

This weakness, however, is counter-balanced by the numerous other evaluation approaches that were done to assess ACRM (Holt, Boehm-Davis, & Hansberger, 1999; Incalcaterra & Holt, 1999; Ikomi & Holt, 1999). All the methods of ACRM evaluation supported the primary finding here, that ACRM training improved certain aspects of pilot performance above traditional CRM training.

REFERENCES

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 1995, 37(1), 32-64.

Foushee, H. C. & Helmreich, R.L. (1988). Group interaction and flight crew performance. In E.L. Wiener & D. C., Nagel (Eds.), Human Factors in Aviation (pp. 189-227). San Diego, CA: Academic Press.

Foushee, H.C. & Manos, K.L. (1981). Information transfer within the cockpit: Problems in intracockpit communications. In C.E. Billings & E.S. Cheaney (Eds.), Information Transfer Problems in the Aviation System (NASA Technical Paper 1875; pp. 63-71). Moffett Field, CA: NASA-Ames Research Center.

Hansberger, J.T., Holt, R.W. (1999). Comparing evaluator expectations to their evaluations. To appear in Proceedings of the Tenth International Symposium on Aviation Psychology, Columbus, OH.

Helmreich, R. L. & Foushee, H. C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E. L. Wiener, B. G. Kanki, & R. L. Helmriech (Eds.), Cockpit resource management (pp. 3-45). San Francisco: Academic.

Holt, R.W., Boehm-Davis, D.A., & Hansberger, J. T. (1999). Evaluating Effectiveness of ACRM using LOE and Line-Check Data. To appear in Proceedings of the Tenth International Symposium on Aviation Psychology, Columbus, OH.

Incalcaterra, K. A. & Holt, R.W. (1999). Pilot Evaluations of ACRM Programs. To appear in Proceedings of the Tenth International Symposium on Aviation Psychology, Columbus, OH.

Lauber, J. K. (1984). Resource management in the cockpit. Air Line Pilot, 53, 20-23.

Salas, E., Bowers, C.A., & Cannon-Bowers, J.A. (1995). Military team research: Ten years of progress. Military Psychology, 7,55-75.

Stout, R.J., Cannon-Bowers, J.A., Morgan, B.B., Jr., & Salas, E. (1990). Does crew coordination behavior impact performance? Proceedings of the Human Factors Society 34th Annual Meeting, Santa Monica, CA, (pp. 1382-1386).

National Transportation Safety Board (NTSB). (1994). Safety study: A review of flightcrew-involved major accidents of U.S. air carriers, 1978 through 1990 (PB94-917001 NTSB. SS-94/01). Washington, DC: Author.

Wiener, E. L., Kanki, B. G., and Helmreich, R. L. (1993), Cockpit resource management. San Francisco: Academic.

Table 1. Component analysis loadings for the workload management component for both frequency and quality items.

Workload Management Component	Component Loadings
Frequency items	
Crews assign PF & PNF duties as the 1st action done for an abnormal	.74
Crews reduce workload by maintaining division of duties	.68
Crews cross-brief ideas and solutions to one another during abnormals	.81
Quality items	
Crews avoid distractions during critical phases of flight	.75
Crews are organized and prepared in the cockpit	.46
Crews reduce workload by maintaing division of duties	.68
Crews cross-brief ideas and solutions to one another during abnormals.	.85
Crews present critical solutions for abnormal situations to the other pilot before implementing	.69

Table 2. Component analysis loadings for the communication component for both frequency and quality items.

Communication Component	Component Loadings
Frequency items	
Crews give tailored CLEARANCE briefs according to the particular flight conditions	.69
Crew formulate & communicate bottom lines and backup plans, prior to critical phases of flt	.87
Crew formulate & communicate bottom lines & back-up plans for the abnormal	.75
Quality items	
Crews include tone, roles, communication, teamwork & operational issues in PREFLIGHT briefs	.70
Crews formulate & communicate bottom lines & back-up plans for the abnormal situation	.69

Table 3. Component analysis loadings for the planning component for both frequency and quality items.

Planning Component	Component Loadings
Frequency items	
Crews plan in the early phases of flight for possible changes, deviations, or alternate outcomes	.84
Crews give tailored ARRIVAL briefs according to the particular flight conditions	.75
Crews prioritize duties during abnormal/critical phases of flight	.64
Quality items	
Crews give tailored CLEARANCE briefs according to the particular flight conditions	.71
Crews conduct normal briefs during low workload times of flight	.80
Crew assigns and monitors the execution of the plan for an abnormal	.59

Hansberger J. T., Holt, R. W., & Boehm-Davis, D. A. (1999). Instructor/evaluator evaluations of ACRM effectiveness. Proceedings of the Tenth International Symposium on Aviation Psychology. The Ohio State University: Columbus, OH.

Table 4. Scale intercorrelation matrix for workload management, communication, and planning.

	X	SD	1	2	3
1. Workload Management	3.57	.51	(.90)		
2. Communication	3.92	.53	.56*	(.85)	
3. Planning	3.62	.59	.59**	.53*	(.89)

Note: Reliabilities are in ( ); all scores are scale scores.
* $p \leq .05$; ** $p \leq .01$

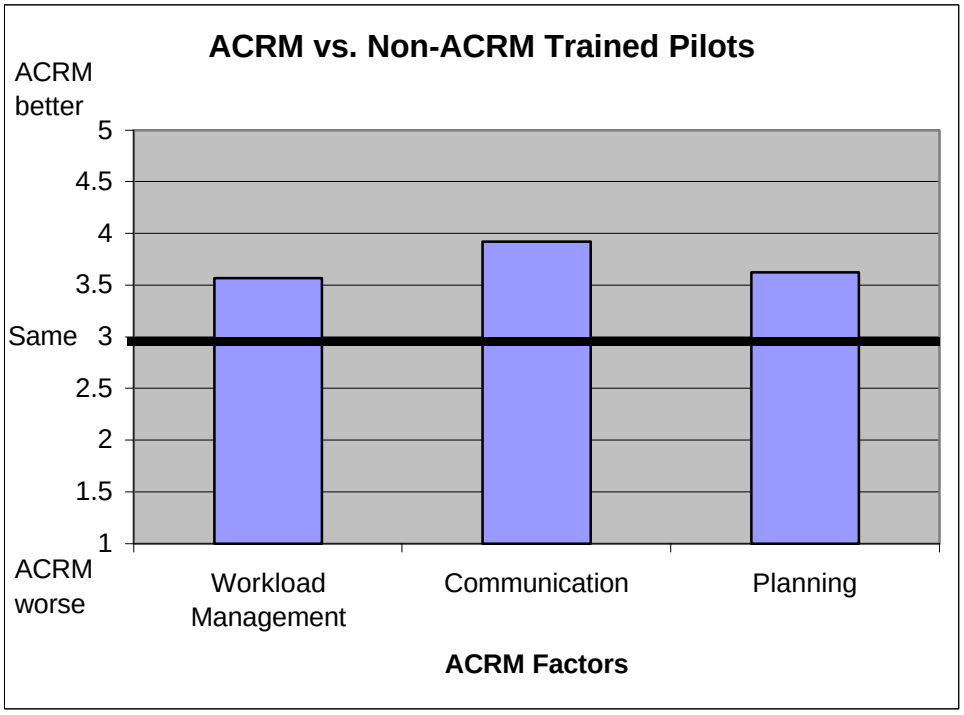

Figure 1. Means of the I/E's comparative assessments for each ACRM component. The bold line represents the comparison value for the one-sample t-tests. The 1-5 scale has been relabeled from ACRM much better (5) and CRM much better (1) to ACRM pilots better (5) or worse (1) than CRM trained pilots.