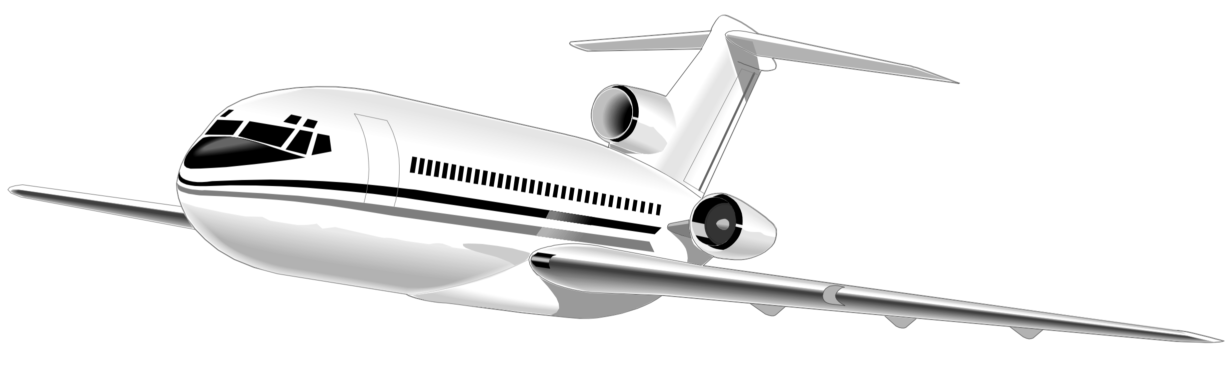
Eleana Edens, Ph. D.
Office of the Chief Scientific and
Technical Advisor for Human Factors, AAR-100
Federal Aviation Administration

Air Carrier Training Research Review

January 30, 1999



The Federal Aviation Administration (FAA) Office of the Chief Scientific and Technical Advisor for Human Factors (AAR-100) directs an air carrier training research program centered on methods of effective pilot training and assessment. Semi-annual research reviews are conducted by this office. The following report summarizes the October 1998 review.

Introduction

Although basic technical and Crew Resource Management (CRM) concepts are widely accepted, much remains to be learned regarding the appropriate methods for effective training and valid and reliable assessment of training programs. The general research philosophy guiding efforts to improve training and assessment is that research must consider distinct segments of aviation training systems. Individuals comprising the crew, instructors who train and evaluate crews in the classroom, the simulator and on the line, as well as the management culture responsible for the safety climate of the carriers should all be considered. Additionally, this research must regard the variables important to Line-Oriented Flight Training (LOFT) development, implementation and evaluation. Thus, this research centers on (1) crew training and assessment, (2) instructor training, (3) LOFT development strategies and (4) organizational and systematic influences on pilot performance.

Address questions or comments to:

Eleana Edens, Ph.D.
AAR-100
(703) 960-1085
Eleana.Edens@faa.gov

Project Title
Identification of Hard-to-Train CRM Skills

Performing Agency
University of New Mexico, Albuquerque

Background
Traditionally, CRM training has focused on the high-level CRM components or elements of pilot performance. The training and assessment of high-level concepts is fraught with many problems. Among them, the validity of course content is questionable, and the reliability and validity of CRM assessment has not been established. Prior research has shown that assessment of observable behaviors, rather than global concepts, especially within simulator scenario event sets, can improve the reliability of CRM assessment. Therefore, observable behaviors have provided a more meaningful and lower level of detail for the CRM assessment process. However, the research and operational communities have not yet developed a comparable level of detail for the CRM training process. It is impractical to train observable behaviors. Thus, at present, it is likely that training would be on the global level but assessment could take place at the observable behavior level. This training paradigm, although improved from all CRM training activities being at the global level, still lacks the training vigor employed, for example, in technical training. An approach to solving this dilemma is to identify the knowledge, skills and abilities underlying CRM concepts and then to train at that level. Observable behaviors could be identified and assessed for skills in situation-specific training events.

In addition to the need for refined skill analysis methods, reliable data collection systems must be developed in order for carriers and the FAA to evaluate the effectiveness of training programs.

Project Description
This research project has been investigating two distinct segments of air carrier training: skill analysis, and reliable performance assessment.

Work has been completed at three major carriers, and data collection is underway at a fourth carrier. This research will identify key CRM skills that need to be trained in transition and recurrent training. The main objective of the research is to provide carriers with a set of steps to identify and link CRM skills to essential elements of crew training so that those skills can be systematically trained and assessed. This research has collected rating and agreement index data to help representatives from different carrier departments form a consensus in identifying CRM skills. It is significant that Advanced Qualification Program (AQP) and Human Factors Subject Matter Experts (SMEs), as well as check airmen participated in the process, providing the CRM skill identification with a substantially broader base than is provided by traditional Instructional Systems Design (ISD) methods. Using a greater number of SMEs for each of the steps allowed for a broader consensus and better understanding of CRM skills across the organization while distributing the workload over a larger number of individuals. This process reduced some of the bottlenecks encountered in traditional carrier skill analyses where one or two SMEs are asked to perform most of the analysis. This has resulted in a CRM skill listing and a set of methods that carriers can use to identify their own critical CRM skills. An effort is under way to develop a CRM skill identification manual that will summarize the process for other air carriers.

The results of this CRM skills identification will be combined with analyses from previous research which focused on LOFT/LOE session debriefings, decision-making and situational awareness skills. The sets of CRM skills, derived from four major carriers, will be used to develop a CRM curriculum outline applicable across air carriers. The CRM curriculum outline will include the basic CRM skill requirements, the advanced CRM skill requirements, and the key knowledge components of those advanced skills. It is expected that data collection will be finished in FY99. The CRM skills document and curriculum outline will be finalized in FY 2000 with substantial air carrier input to the final product. Interim CRM skill analysis results are currently available from AAR-100.

Another component of this research project focuses on pilot assessment. Efforts to reliably assess pilot performance during training have traditionally been less than successful. Since FY96 this research team has been working with several air carriers to improve training assessment by designing rater training and new data collection systems. This work focuses on rater calibration, development of grade sheets, development of videos used in the calibration sessions, presentations to the instructors on the importance of quality data, collection and analysis of the calibration data, and reporting results. Several calibration studies have been completed and additional sessions are planned. Thus far, the results of these calibration studies show (1) how minor changes in the wording of observed behaviors listed on gradesheets can have a significant effect on ratings, (2) large individual differences among raters exits, (3) higher inter-referent than inter-rater reliability and (4) a significant increase in reliability when improvements to the calibration sessions were implemented (wording of the gradesheet, quality of video, better briefing of the event sets). The frequency of the higher inter-referent reliability is important to note because there are differences among carriers in their views of whether a referent should be established and instructors trained to the referent or whether instructor's inter-rater reliability is adequate. AAR-100 is investigating this question further by work being done by George Mason University and NASA Ames at several other carriers employing the inter-rater reliability methodology.

Based on the observations of this process at a major carrier, this team of researchers has developed a software package to be used in group calibration sessions. This software will allow carriers to more effectively train and assess evaluators in their ability to judge Line Oriented Evaluations (LOE) performance. This software facilitates the data collection, statistical data analysis, and report generation for evaluator LOE calibration sessions. A group of evaluators individually rate the performance of a crew shown on a video flying an LOE. After the ratings are entered on a standard gradesheet they are collected and the scores are entered into a PC using Microsoft Access software. After entering the evaluators' ratings a complete set of statistical analyses of the data is provided. The reports summarize each evaluator's performance along with various group performance statistics. The reports also serve as the basis for the calibration feedback, debriefing, and further training of the evaluators.

Although carriers conduct regular and recurrent rater calibration sessions, using the methodologies developed by this project, it is often necessary to recalibrate one instructor. The cost of bringing together a group of instructors in order to conduct a group session to calibrate a single instructor is prohibitive. This research team has developed a prototype PC-based system that effectively presents an individual calibration session.

This software tool is designed to allow individual evaluators to undergo the same type of training and calibration as discussed above in the group sessions but to self-calibrate individually, at their own convenience. The system provides the pre-briefing instructions regarding the specific flight scenarios that would be presented and some general instructions regarding the use of the software. An audio/visual presentation of a crew flying the LOE is presented along with the gradesheet. The evaluator enters ratings by simply clicking the appropriate buttons displayed. After entering all of the ratings for all events, the data are analyzed and the results presented in a form that is meaningful to the evaluator (e.g., how close the ratings were to referent grades, how the ratings compared with other evaluators, which ratings were most deviant from a standard, etc.). This tool can be used either as a training or evaluation tool. Its most obvious strength as a training system is that it allows evaluators to obtain training at their convenience, receive immediate feedback on their performance, and to have their grades linked to fleet qualification standards. Its greatest strength as an evaluation tool is that it allows personalized assessment. Assuming that evaluator grading history is stored in an Evaluator Performance Proficiency Database, it is possible to determine whether an evaluator had specific weaknesses allowing the carrier to address this issue.

This research team has also produced a document describing what it means to have quality data, why quality data are necessary for training under AQP, and the methods for achieving quality in performance assessment. This document, distributed in FY98, is intended for evaluators, and instructor/evaluator supervisors. Another document addressing important issues for carriers just entering AQP has been reviewed by the FAA and will be distributed during the first quarter FY99.

Additional work by this research team focuses on developing a Performance Proficiency Database for Instructors and Evaluators as required in full implementation of AQP. It is necessary to have data that reflect the effectiveness of instructors (i.e., the performance of their former students on various measures extending beyond the specific classroom evaluation conducted by the instructor). It is necessary to track calibration performance over time, as well as any trends in LOE and first-look performance evaluations. This team will develop the capability of linking pilot performance data to the Performance Proficiency Database. It is expected that this segment of this research will be completed in FY99, if air carrier participation is not restricted.

This research team will work with two major carriers to determine if Flight Operational Quality Assurance (FOQA) data is predictive of instructor/evaluator assessments. Part of this work will be to validate methods being developed by other FAA researchers at other carriers. It is essential to determine if work done at one carrier will generalize across carrier environments.

Products

- Skill Analysis methods, instructor/evaluator calibration software, performance data analysis tools.

Title
Pilot Training and Evaluation: Airplane Simulation Human Factors

Performing Agency
Volpe Center, Cambridge, MA

Background
Airplane simulators are critical for the training and evaluation of pilots because they eliminate training accidents and enable scenario-based training. The recently enacted requirement that regional airlines operate under the same rules as major airlines, as well as the Advanced Qualification Program (AQP) training approach, nearly oblige operators to use simulators. However, for regional airlines, cost continues to be an obstacle to flight simulator access, particularly for recurrent training.

Currently, airplane FTD and simulator qualification criteria are contained in Advisory Circulars AC 120-40B and 45B/C and related documents. Changes both in the air carrier training environment and in the field of simulator technology demand a reassessment of these criteria. The shift in training philosophy from a time-based list of specific training events to the Advanced Qualification Program (AQP) training approach based on individual training needs requires that much of the training formerly performed in aircraft will move to the simulator. Considering the initiative that Part 135 carriers be held to the same safety standards as Part 121 carriers, the success of AQP depends on whether regional carriers will be able to afford simulators that are appropriate for training Part 121 proficiency using the AQP training approach.

In the past, the goal in simulator manufacturing was to produce simulators that were engineered as similarly as possible to the airplane, resulting in the highest possible physical fidelity. Today, however, engineering capabilities have progressed in such a manner that not always the highest technical capabilities need be applied to all systems of the simulator, but high fidelity in one system may compensate for lower fidelity in another. The goal, thus, has shifted from highest possible physical fidelity to highest possible perceptual fidelity. That is, a good simulator should feel like the airplane and most important, elicit the same performance and behavior from pilots as the airplane. The purpose of this project is to increase the availability and affordability of airplane simulators by determining exactly what level of perceptual fidelity, applied to the most costly aspects of a simulator, is required to maintain or improve the level of safety in airline pilot recurrent training and evaluation.

Project Description
This project is proceeding in three directions, (1) the collection of Subject Matter Expert (SME) opinion, (2) a literature review and (3) new research.

Two workshops were conducted with experts from the FAA, industry and academia. The topic of the first workshop was the aeromodel used to program simulators. The topic of the second workshop was the motion requirements for simulators. These workshops led to proposed changes to the regulations for Level B simulators and gave direction to subsequent research (AC 120-40B).

The second component of this project is an extensive literature review of simulation research. The information has been organized into a large database, which is updated regularly to incorporate new research results.

The third component is the implementation of new research to address the question of the motion requirement for recurrent training. Both the motion workshop and the literature review showed that although motion is presently required for a level B simulator there is no scientific basis for such a requirement despite much research in this area. Given the advances made in simulating visual scenes subsequent to the earlier research, it may now be possible to achieve sufficient perceptual fidelity of motion with the vision system alone. As counterintuitive as this may seem, one must consider that no amount of technology can erase the fact that the simulator will not take off. Thus, the motion cannot be perfect and in fact may create false cues. In contrast, the visual cueing can be close to perfect. Not requiring physical motion would represent a significant cost saving both in acquisition and maintenance of simulators used in recurrent training.

Research is underway at a regional carrier to determine if a high quality motion platform is necessary for recurrent training and evaluation when using a high quality visual system. Approximately one half of the crews have been tested. This process has been hampered by a variety of problems, which have been resolved as far as possible. Data collection is expected to be completed in FY99.

The analysis of the collected data is also underway. Enormous volumes of data measured from the simulator is being collected for each crew, in addition to subjective data from each crew. Determining how to examine the objective data was a primary goal of this project, as a task of this magnitude has not been undertaken previously. It has now been determined which variables to examine and which specific time periods are important. The subjective data has been summarized and graphed and is now ready for further statistical analyses.

Data collection and analyses should be completed by the end of FY99. The final report has been initiated and will be completed in concert with the data analyses.

Future plans include validation of the current experiment using different equipment and subject populations. Ideally, some transfer studies to and from the airplane would also be performed. Avenues to accomplish this are being investigated.

Product

- Research report.

Title
Realistic Radio Communications Simulations

Performing Agency
NASA-Ames and Volpe Center, Cambridge, MA

Background
Currently air carriers are employing high fidelity simulations in the training and evaluation of flight crews. This approach replicates the aircraft operational environment rather realistically but does not replicate the air traffic component of the system adequately. It is a well-established fact that air traffic communication plays an important role in line operations whether routine or abnormal. In most air carrier training, air traffic control is fabricated by the instructor implementing the entire simulation. Thus, the crews, familiar with the instructor's voice and typical commands, may not view these communications as distracting and therefore the communications may not be threatening to safe decisions in simulation. Air traffic communications are a valuable resource for air crews but can also be a distraction and a contributing factor in performance degradation of flight crews. Therefore, the absence of realistic air traffic simulations in aircrew training is a concern for training departments and the FAA. If crews were presented with realistic air traffic interactions during training, the potential for unexpected air traffic control (ATC) requests, distracting communications, missed information, or misunderstanding of information during line operations would be lessened.

Project Description
This research team will investigate the feasibility of automated air traffic communications during the first phase of this study. Industry and manufacturers will be surveyed for state-of-the-art practices, products and tools. Phase II of the study is dependent on the outcome of phase I. Phase I of the research should be completed in FY99.

Title
Rapidly Reconfigurable Event Sets

Performing Agency
Naval Air Warfare Center Training Systems Division (NAWC-TSD), and the University of Central Florida

Background
Line-Oriented Evaluations (LOEs) are a methodology used in Advanced Qualification Programs (AQPs) to evaluate pilot training performance and establish trainee proficiency. LOEs consist of flight simulation scenarios that are developed by the training organization and approved by the FAA.

In the past, LOEs were developed and apprised individually. That is, each LOE was separately conceived, developed and tested by the training organization and individually reviewed and approved by the FAA. The development and approval of new LOEs was costly and time-consuming. As a result, training organizations usually had only a limited number of LOEs available for evaluation, each of which was only approved for a limited time period. In some cases, this may have increased the risk of LOE scenarios being compromised within a training organization, thus reducing the validity and reliability of the scenarios.

One way to reduce the negative aspects of the previous LOE design process is to develop a set of LOE event sets (modules) which, upon FAA approval, could be assembled by the training organization into a number of unique LOEs, without requiring additional specific approval of each individual LOE by the FAA.

The availability of such a system will not only significantly reduce the time and cost required by training organizations to create valid scenarios for training and evaluation, but will increase the diagnostics of scenarios by directly linking the event sets with CRM and technical training objectives. Further, it will increase the validity of the evaluative process by giving evaluators detailed guidelines and materials for the administration of scenarios. This system will increase the opportunities for air carriers to create valid training and evaluation scenarios for global aviation operations.

This research team is developing the appropriate software to allow the reconfiguration of separate event sets into entire LOEs. The requirements document has been written and approved. Identification of a database structure for the software system has been developed. Databases for aircraft technical data, event sets and aircraft differences have been established. Usability tests at a major carrier are scheduled for the first quarter of FY99 and will be expanded to two other major carriers later in FY99.

A significant element of the development of this software is the determination of the relative difficulty of each event set and therefore each LOE. This research must determine whether component difficulty ratings can be meaningful predictors of overall LOE difficulty and whether raters show reasonable agreement on difficulty ratings. The research problem is to determine how to assess the difficulty of the LOEs that are generated by combining phase-based event sets. Option One is to add task difficulties across an entire LOE. Option Two is to average task difficulties across an entire LOE. Option Three is to average task difficulties per phase, and then per LOE. Option Four is to average phase difficulties for an entire LOE.

This research team tested these different options with evaluators/instructors at a major carrier. Results indicate that overall LOE difficulty can be derived from individual component ratings and the component difficulties should be averaged, not summed. By averaging, the relative differences between LOEs are maintained. The difficulty of a phase-based set of events can be reliably estimated by averaging the difficulties of the events/tasks used to construct the event set. By using component ratings, relative differences between LOEs are preserved. Thus, initial recommendations from this data are that in reconfiguring event sets into LOEs, carriers should use the largest component possible for difficulty ratings, average the component ratings per phase, and average the phase ratings per LOE.

It is expected that this project will be completed in FY00.

Product

- Software for reconfiguring event sets.

Project Title
Identifying Leadership and Followership Skills in Crew Resource Management (CRM)

Performing Agency
Battelle Memorial Institute

Background
Historically, leadership has been viewed as a linear, one-way (i.e., downward) action with the primary aim being task accomplishment. This perspective has dominated the cockpit of air carriers. However, as roles in the modern cockpit change the focus from an individual's performance to a crew's performance, the skills required of the crew members change also. It is now thought that leadership is an activity which involves both leaders and followers as they interact to accomplish goals and that followership skills are as important as leadership skills to the safe and efficient performance of flight crews. It has been unclear how to train leadership/followership since the skills underlying these concepts had not been identified.

Project description
This project has analyzed cockpit performance in order to determine the behavioral components of leadership/followership skills.

A model of cockpit leadership was developed earlier with Western Michigan University, which guided the subsequent skill analysis. This model of effective and efficient flight operations points out that the crew must use many skills to deal effectively with operational variables largely out of their control. These variables include the regulatory environment, the corporate environment, the market environment, and the physical environment. Skill analyses were conducted based on the assumptions that (1) a minimum amount of leadership/followership is required to achieve a safe flight, (2) individuals constantly oscillate between leader and follower roles, and (3) weak leader or follower skills will be compensated for by other crew members. Preliminary analysis are complete and determined that in order to deal effectively within the framework described in the model, crews must possess skills in envisioning, modeling, receptiveness, influence, adaptability, and initiative. The analysis also showed that these skills are common to both leadership and followership.

Data from line audits with major carriers reveal that the components described in the model are related to overall crew effectiveness, thus validating the skill analysis data. The line audit data showed that, generally, in the more severely abnormal situations, the crew tends to be less effective in exhibiting leadership/followership skills. Captains tend to be less likely to articulate a vision for the flight, meet company standards, obtain commitment from other crewmembers or be adaptable. Both Captains and First Officers were less likely to initiate actions in response to an operational deficiency when the severity of abnormal situations was high. When handling a complex situation, but not an abnormal one, these crews displayed "outstanding" leadership/followership skills. These ratings were not dependent on position or time together as a crew. This suggests that present leadership training may not be meeting its objectives especially when the crews are faced with an abnormal situation.

The data also showed that Captains do in fact, set the tone in the cockpit. When Captains articulated a vision for the flight, the First Officer would initiate a response to an operational deficiency without further direction from the Captain. When Captains exhibited good conduct and high standards, First Officers exhibited similar behavior. When Captains were receptive, First Officers were likewise receptive. This finding supports the need for appropriate pre-flight briefings.

Maintaining vigilance during the flight seems to be dependent on each crewmember at different phases. During predeparture, takeoff, climb, and cruise, vigilance was related to the Captain's envisioning, modeling and receptiveness. However, during the descent and approach phases, vigilance was related to the First Officer's conduct and standards.

Workload and task distribution were dependent on the Captain's conduct, standards, and receptiveness. Establishing guidelines for automated systems for all phases of flight was related to the Captain's articulated vision, conduct and standards (modeling), receptiveness and use of interpersonal skills to obtain commitment from others.

Based on the skill analysis model and this line data, requirements for leadership/followership training curriculum have been identified. This includes the development of classroom exercises extending beyond role-playing, the creation of event sets that address critical leadership/followership skills and addressing company philosophy and policy issues. A preliminary training package was developed for a major carrier, which includes an instructor manual, student manual, and training videos. This package was distributed to industry in FY98.

Based on this preliminary work, the current work focuses on extending the training curriculum by integrating leadership/followership skills into the second generation Model AQP. This current work greatly expands the leadership/followership skills of the previous model just described, by enhancing the CRM training development methodology. The new work employs task-based methodology in the identification of concrete tasks with their supporting knowledge and skills to foster more focused, skill-oriented training. When complete, this new work will have identified a broader range of leadership/followership skills associated with specific flight tasks, rather than the global model developed in the preliminary work. Rather than awareness training, this new work will result in focused skill training. This methodology reflects the benefit to CRM training that arises from the application of the AQP framework. This effort will incorporate leadership training throughout the entire AQP training program. This will result in products such as events and exercises that air carriers can incorporate into their AQP flight crew training.

Products

- Leadership/followership training guidelines, and training systems.

Project Title
Analysis of Cockpit Management System in Multiple Carrier Environments

Performing Agency
George Mason University

Background
Based on previous research, it is widely believed that Crew Resource Management (CRM) training can result in significant improvements in flightcrew performance and flight safety. Historically, it has been possible to assess the effects of CRM training on pilot attitudes regarding the general management of cockpit duties with self-report inventories administered pre and post-CRM training. It is thought that a change in pilot attitude will result in a change in behavior in the cockpit. It has also been possible to subjectively evaluate the effects of CRM training on pilot performance in Line-Oriented Flight Training (LOFT) and on the Line by noting the differences in CRM performance between crews that have had CRM training and the crews that have not. Further, performance differences between air carriers and between fleets within one air carrier have been investigated. While this approach to assessment has contributed to the advancement of CRM, it is hampered by the practical constraints of self-report and global measurement.

Project Description
This research project will help determine the most efficient method to train and assess CRM and whether conceivable performance differences pre- and post- CRM training can be empirically substantiated. This research project was conducted with one regional and one major carrier. The project designed and implemented a prototype CRM training program based on a task analysis methodology. The research evaluated the effects of this specific CRM training on the Line, LOFT, and Line-Oriented Evaluation (LOE) performance of crews. This study also integrated CRM performance requirements or procedures into the standard operating procedures of the air carrier. CRM procedures are the implementation of specific calls, checks, and/or guidance into one or more of the following: normal checklists, Quick Reference Handbook, Abnormal/Emergency Procedures, Flight Standards Manual, additional Job Aids. This can be viewed as translating critical CRM principles into CRM procedures.

In developing the prototype Advanced CRM (ACRM) course at the regional carrier, data from line operations, instructor comments and ratings, along with findings from the NTSB commuter safety study were used as the basis for the CRM training design. These data were incorporated into a proceduralized management system specifically tailored to the needs of this regional carrier. The advanced CRM course has been given to pilots in the research fleet. Their performance in a LOE was assessed prior to this training and is currently being assessed after the training course was completed following an appropriate lapse of time. The central focus of this research was to determine if a proceduralized system would increase pilot performance, and that using this approach to CRM actually increases pilot performance and thereby increases safety. This research has thus far proceduralized: Team Management; Crew Communication, Decision-making, and Situation Awareness in the operational environment of the regional carrier.

The proceduralized CRM training course is complete, and has been given to flight crews. The Quick Reference Handbook and the normal procedures have been rewritten to reflect proceduralized CRM. Data collection and analyses are complete.

During the entire course of this project, performance data was collected and analyzed. Final results are presented later in this section. Analysis of the base line performance data showed some interesting results regarding instructor/evaluator ratings. In the research process it was necessary to train evaluators to collect data during the LOEs. The evaluators are line pilots working in the training center. One of the critical questions centered on whether the evaluator training was successful and would result in reliable and valid CRM assessment. The analysis confirmed the LOE assessment process taught in the evaluator training, which is based on ratings of observable behaviors (in this program, evaluators are taught to make low-level observations of fundamental behaviors, or clusters of behaviors, which can be used as the basis for higher-level judgements), technical and CRM ratings, overall Pilot In Command (PIC), Second In Command (SIC) and Crew ratings for each event set. Observable behavior ratings significantly contributed to both technical and CRM evaluations of each crew. This confirms that with appropriate training and calibration, evaluators can use specific, well written observable behaviors as the basis for judgements which should increase the objective component of CRM evaluation in an LOE and ensure more fair, unbiased evaluations of each pilot.

CRM ratings across event sets, as well as technical ratings, significantly contributed to the overall evaluation of PIC and SIC. This confirms that evaluators, when properly trained and calibrated, can consistently integrate CRM evaluation with technical performance to obtain overall evaluation of crewmembers on each event set. This is evidence that the LOE evaluation was reliably tapping CRM as well as technical performance. This is important because in the past there were questions as to whether CRM ratings were just a reflection of technical performance. It has been suggested that evaluators could not or would not rate CRM behavior as a separate performance domain, but part of the entire crew performance.

Additionally, the analysis showed that Captains and First Officers were evaluated differently. In evaluations of the entire LOE, CRM performance was more important than technical performance for the PIC, while technical performance was more important than CRM performance in rating the SIC. These differences in evaluations are congruent with holding the PIC more responsible for CRM as part of the Captain's role. These findings are also congruent with previous findings that Captains state that the most important feature of a good First Officer is technical competence and First Officers state the most important feature of a good Captain is CRM competence.

Further analysis showed stronger relationships of Captain evaluation to Crew performance compared to First Officers. This provided additional evidence that Captains are held more responsible for crew performance (particularly CRM performance). It also empirically confirms that Captains are the focal point for CRM and crew performance, and suggests additional training on establishing and maintaining CRM should be performed in upgrade training when a First Officer transitions to Captain.

LOE data was compared with the data from traditional proficiency checks of specific maneuvers. The proficiency check evaluations could predict about 50% of the overall evaluations of PIC and SIC on the LOE. This confirms that the technical proficiency emphasized by the traditional check is relevant to LOE performance. However, the remaining 50% of the variance unique to the LOE confirms that the LOE is also tapping a distinct type of performance in addition to technical proficiency. More detailed evaluation of proficiency check assessments showed that only evaluations of the SIC do not predict LOE results. Additional analyses are required to better understand the role of rating SIC performance in proficiency checks.

Analysis of different event sets found significant differences in the performance of the targeted observable behaviors. Isolation of performance problems across crews can be used to rationally and efficiently change pilot training. The AQP audit and proficiency databases can be used to find the specific training components for objectives that must be strengthened. This analysis shows that isolating poor performance across crews can be used to directly change relevant components of training.

The final evaluation the Advanced Crew Resource Management (ACRM) training at the regional carrier employed a multi-method protocol. Data were analyzed by a statistical comparison of the two fleets using the Performance Proficiency Data Base, comparative evaluation of ACRM performance by I/Es in both fleets, subjective evaluation of ACRM by pilots in both fleets, and direct jump seat observations of crew performance in both fleets.

Data from these sources revealed that the combination of specific CRM procedures that were both trained and incorporated into fleet SOP was effective in producing specific changes in crew performance.

First, data from LOEs showed that the mean performance of the ACRM-trained fleet was consistently and significantly better than the traditional ACRM fleet. Also, items in the LOE for which the training makes a difference are positively correlated with a variety of other performance items across the event sets at a higher level than would be expected by chance. This data supports the contention that ACRM is effective and that the effects are not just limited to specific items or just one event set.

Second, the line check items, also from the Proficiency Performance database, support a pattern of significant fleet differences in favor of the trained fleet. This demonstrates that the training and SOP implementation affected crew behavior on the line as well as in the simulator, thus demonstrating that the positive effects of ACRM did transfer from the training to the operational environment.

Third, the jump seat observations showed differences in the trained and untrained fleets in all phases of flight. Noticeable differences were reported in items from the Departure, Cruise and Arrival phases of flight. This supports the contention that ACRM effects are not just limited to a specific phase of flight.

Fourth, a survey was developed for pilots to measure attitudes toward CRM and ACRM, knowledge and practice of ACRM procedures, and perceived effects of ACRM. The results from this survey strongly support the ACRM program. ACRM-trained pilots had very positive attitudes toward CRM in general and ACRM in particular. When compared to a baseline, ACRM-trained pilots show significant knowledge of ACRM, frequently perform ACRM procedures, and overwhelmingly endorse ACRM training.

Fifth, the instructor survey asked for a comparison between pilots with ACRM training and those with traditional CRM training on the basis of the frequency and quality of specific behaviors.
The results of this survey show that ACRM-trained pilots were rated higher on workload management, planning, and communication.

These different lines of evidence support the conclusion that the effects of ACRM training and SOP go beyond the specific procedures to improve overall crew performance. Appropriately designed and trained procedures can enhance the crew's ability to communicate effectively, plan, manage workload and solve problems during flight operations.

The ACRM course manual was distributed in FY98. The final report for this segment of the research will be distributed in FY99.

The proceduralized CRM concept has been extended to a major carrier where CRM procedures were integrated into a Quick Reference Checklist (QRC) for emergency procedures. The development and implementation of the QRC represented a fundamental change in how the carrier addressed emergency procedures in the past. The objectives of employing the QRC were to reduce the number of memory items which can lead to operational errors, and to ensure that all critical actions are taken in sequential order, eliminating errors of commission and omission. The study was initiated to validate the usefulness of the QRC and to identify design problems, training issues and overall ability of the QRC to improve system safety. The study assessed 124 crews performing checkrides with the QRC or by the traditional memory immediate action items. The crews were assessed on the following elements:

- timing of the crews from the beginning of the emergency event until the completion or the initiation of reference action items,
- timing of when each action of the checklist was initiated,
- errors in crew performance caused by omission of checklist items,
- errors in crew performance caused by omission of checklist items,
- errors in crew performance caused by addition of items to the checklist,
- errors in crew performance caused by not following the sequential order of the checklist,
- overall crew performance,
- crew survey and feedback regarding QRC philosophy,
- specific timing of events on critical checklist.

The data indicate that the QRC checklist significantly reduced crew error in checklist performance. This was measured in errors of omission, commission and order. The QRC checklist received overwhelming approval by the line crews participating in the study. Significant improvements were made to the QRC during the study, based on line input. The data also show that the QRC does increase the time for completion of the emergency checklist. However, with the increased accuracy, crew performance is actually enhanced. Further, QRC crews received overall higher ratings in CRM, particularly in the areas of communication and workload management.

A paper was written describing the study results and distributed to industry in FY98.

Research has continued with several carriers to identify appropriate questions, data collection strategies and statistical methods to analyze pilot performance data. This data collection system will allow carriers to evaluate their AQP training effectiveness in many dimensions. The overall goal of this part of the research is to determine the appropriate data necessary to extend training intervals for AQP programs. Extensive AQP statistical analysis activities have been performed on data from major and regional carriers.

In order to ensure that reliable and valid performance data is being collected by carriers and reported to the FAA, this research team found it necessary to address rater calibration issues. The overall goal of this segment of the research is to assist carriers with incorporating inter-rater reliability training strategies into existing instructor/evaluator (I/E) indoctrination training. This project involves: selecting and evaluating existing videotapes of crew performance to use as standards for the I/E indoctrination class; developing behavior-based grading sheets for the selected video tapes; and collecting and analyzing data from indoctrination classes on levels of agreement that emerge as a result of training.

This rater training has continued at major and regional carriers. The purpose of this training is to convey the importance and knowledge/skills necessary for instructor/evaluators to provide stable, reliable and accurate assessments of pilot performance. Continuing refinements have been made to the inter-rater reliability software (developed earlier in this research), instruction materials and the process itself. Some of the improvements include better feedback to the instructors, a training guide for conducting inter-rater reliability training and a new summary report for the IRR session. This summary report closes the loop from IRR training to fleet captains and managers and facilitates the needed changes in flight standards operations.

The research team has developed an entire instructor recurrent training package for Aer Lingus in an effort to take the research results to the international community. This training was delivered to Aer Lingus in January 1998.

Future plans for FY99 and beyond focus on extending the expertise in the development and measurement of CRM procedures to the procedures and practices involving crew interaction with automated systems. This research team will be part of an integrated group of scientists that will study automation performance issues under the direction of AAR-100. The George Mason University research team will also continue to focus on quality assurance issues at several carriers.

Products

- A cockpit management training system.
- Research report on performance differences due to proceduralized CRM training.
- Inter-rater reliability training.
- Pilot performance analysis systems.
- Reports on crew interactions and automated systems.

Project Title
Human Factors, Human Error and System Safety.

Performing Agency
The University of Texas, Austin

Background
Prior research by this team consists of several projects that centered on evaluating the impact of Crew Resource Management (CRM) and Line-Oriented Flight Training (LOFT) and developing methods to optimize training and performance evaluation. This research group has maintained an international, longitudinal database of pilot performance since the late 1980s. Prior findings from this work indicate that formal CRM programs combined with LOFT have a significant, measurable, positive effect on crew behavior and attitudes and by inference on system safety. However, the data also indicate that the current CRM and LOFT programs have not solved all human factors problems in aviation. Specifically the following negative outcomes have been documented through this research program:

(1) There is great variability in the impact of programs and in observed crew performance, even after implementation of CRM training. Incomplete acceptance of CRM concepts and practices by instructors and evaluators has been identified as one source of problems.

(2) In the absence of continuing reinforcement, behavior and attitudes regress almost to pre-training levels.

(3) LOFT programs in several organizations show great variability in execution and impact.

(4) Considerable controversy remains regarding the evaluation of individual and team performance, as required of organizations participating in the Advanced Qualification Program (AQP).

(5) There is an absence of reliable data on the human factors of incidents occurring in the aviation system.

(6) Recent investigations of major accidents, including air carrier crashes, have shown that organizational cultures have a profound impact on operations and on the effectiveness of human factors programs.

(7) Human factors aspects of cockpit automation are not being addressed in most CRM/LOFT programs and have been identified as critical in a number of accidents.

This project developed an observational methodology for assessment of crew performance in line operations. This line audit approach using the Line/LOS Checklist (LLC) employs a team of expert observers to collect systematic data on line operations ***without jeopardy.*** The research team collects data in each organization giving common evaluation points that allow comparisons of performance across organizations. Organizations have used the findings to target specific issues during recurrent training.

Project Description
The current research extends investigation of the human factors of flight operations supported by previous work by this research group. The original research focused on Crew Resource Management (CRM) as a means of optimizing crew performance and avoiding human error, which has been implicated as causal in the majority of air crashes and incidents. These investigations have demonstrated that CRM is effective in changing flightcrew attitudes and behavior.

Despite demonstrated effectiveness, CRM has come under attack for its failure to eliminate all human error and to prevent accidents involving human error. Some of the attacks correctly reflect the fact that some organizations have provided minimal and outdated CRM training. Other criticisms demonstrate a lack of understanding of the nature of human error and the role of training interventions such as CRM, LOFT, and AQP in system safety. Humans are cognitively limited and further subject to the deleterious effects of stress, fatigue, lapses and cultural biases. The effects of human limitations are exacerbated by complex operating environments that include difficult ATC interactions, increases in cockpit automation and operation into difficult areas such as Latin America and Africa.

Given the physical and psychological limitations of humans, the elimination of error is impossible. What can be achieved is the establishment of defenses against the occurrence and consequences of error and deployment of strategies to manage error and mitigate the consequences of error. In this context, CRM represents a training architecture for developing a set of tools to be used as countermeasures and error management tools.

In this light, it is this research team's belief that CRM training should be refocused to specifically identify error management and the recognition and avoidance of threat. Focusing on error management and threat recognition also has the advantage of basing training on superordinate goals that are globally accepted. However, much remains to be learned about the extent and nature of human error and threat in the current aviation system. The analysis of accidents provides information of only limited utility since an individual accident usually represents the concatenation of rare events.

A focus on system issues and using CRM as a training intervention for error management contrasts with two increasingly common responses to human error accidents and incidents. The first is to increase the level of automation to remove the human from the loop. As behavioral data and other research have shown clearly, automation is not a panacea for human limitations. Rather it adds a new set of errors and problems. The second approach is to proceduralize behavior defining Standard Operating Procedures (SOPs) for pilots to prevent human lapses. This also is a strategy with limitations since data indicate that an over-emphasis on procedures can actually increase the incidence of errors and undesirable behaviors, as crewmembers may ignore highly complex procedural requirements.

There is also growing awareness that safety is a system phenomenon. Factors such as national, professional and organizational cultures play a significant role in safety and in the acceptance of training programs such as human factors/CRM. Cultural factors further influence the use of flightdeck automation, which raises issues regarding the certification of new systems. It follows that attempts to increase the safety of flight operations must address broader system issues as well as training at the individual and crew level. These efforts require accurate and extensive data on the way both individuals and organizations function within the aviation system to guide safety initiatives.

The research program extends prior research in this area. It has eight components that address critical areas of flight operations. These areas are:

(1) Developing and refining a new model of threat and human error in flight operations. The model will be of practical value in developing training programs and in the analysis of incident and operational data.

(2) Extending current data on performance in line operations to include data on environmental threats to safety, error, and responses to error. The line audit methodology developed by this research team is increasingly being used by organizations to determine how crews function under non-jeopardy conditions in line operations. Such data provide validation of the impact of training, including CRM.

(3) Developing a curriculum for organizations to train line audit observers. The curriculum should help organizations develop consistent internal audit methodologies.

(4) Assessing the use of automation in line operations. As part of the line audit methodology, a particular emphasis will be placed on the use of automation, including the number and nature of errors made in automation usage.

(5) Refining incident reporting and analysis methods. This component will develop and refine data collection methodology for incident reporting by flight crews with the goal of generating data that can be used more effectively for analysis and determination of trends.

(6) Determining relationships among safety-related data sources. Different methods (incident reports, line audits, surveys, etc.) produce different data on organizational and system performance. In this phase of the project, relationships among data from these varied sources will be explored with the goal of improving the assessment of system performance and safety.

(7) Extending data on flight attendant CRM. The project has developed a survey instrument to determine human factors attitudes and training needs for flight attendant CRM. The research will extend the database and make the findings available to organizations initiating CRM for this group. To date, survey data from several organizations have been collected showing significant differences from pilot responses.

(8) Developing tools for organizational CRM and safety surveys. The Flight Management Attitudes Questionnaire has been used by most major and many regional airlines to assess CRM curriculum needs and safety concerns. This task will produce a semi-automated package that organizations can use internally for diagnosis.

Products

- Methodologies, data collection and analytic tools, and training products for use by airlines and other agencies.
- Technical reports and scientific papers are also presented on the project's homepage, which is widely used nationally and internationally (www.psy.utexas.edu/psy/helmreich/nasaut.htm).
- Robert Helmreich, the project’s principle investigator, also serves on the ICAO human factors team providing worldwide instruction in human factors issues.
- The project also trains doctoral students who go on to work in aviation positions after receiving their degrees.

Project Title
An Investigation of Training Issues Concerning the Advanced Qualification Program (AQP)

Performing Agency
Battelle Memorial Institute, Columbus, Ohio

Background
In an attempt to encourage the use of innovative training programs, the FAA has proposed the concept of Advanced Qualification Program (AQP) training. Central to AQP is proficiency-based training. Under AQP, carriers who have applied for inclusion in the program can develop their own proficiency objectives, which must address the range of conditions and contingencies that might be faced by pilots working within the carrier's operational domain. These proficiency objectives define the set of skills and tasks a pilot must be able to perform to be proficient on a given aircraft type within the carriers operational domain. The goal is to ensure that the training program meets each carrier's specific requirements and does so in the most efficient way possible. Utilization of a strong analytical frame work for developing a carrier's program helps to ensure that training is systematically oriented towards those objectives of greatest relevance to the individual carrier and also supports meaningful crew and program evaluation. One of the objectives of AQP is to provide seamless integration of CRM and technical skills within the curriculum to ensure that CRM skills are practiced together with all other flight skills and procedures as required by each flight situation. One goal of AQP is that CRM skills should be utilized as a normal and inherent part of aircraft operation, little different from operating automated equipment or performing a proper checklist.

While AQP permits significant departures from the traditional FAR requirements for training and checking airmen, the price of that regulatory flexibility is a detailed front end analysis, the methodology for which is described in Advisory Circular (AC) 120-54. Accomplishing the analyses necessary to create AQP qualification standards has proven to be particularly challenging for participating air carriers. Training developers have had difficulty selecting an appropriate level of analysis detail, efficiently executing their analyses and determining how to incorporate cognitive and crew resource management considerations in a manner that will generate meaningful proficiency objectives, standards, and conditions. These difficulties stem in part from the fact that as a new program, AQP lacks concrete examples for reference purposes. The methodology was developed by the FAA on a priority basis, with the explicit intention that it be refined subsequently as experience with the new programs accumulate. There was a need for a methodology that, if faithfully followed, would produce an effective AQP.

Project Description
The goals of this research project are to (a) assess whether the Instructional Systems Design (ISD) process can be effectively used to develop a prototype AQP, including specification of the process that derives learning objectives from the task analysis; (b) determine whether the ISD methodology needs to be modified to support AQP development; and (c) verify the modified ISD methodology that has been standardized through the development of a database comprised of templates and instructional guidance for developing an AQP.

The approach to integrating CRM into technical training adopted by this research utilizes ideas from several sources: (1) ISD methodology used for AQP curriculum development, (2) the event set approach developed in prior AAR-100 research, and (3) the situation assessment model being developed by this research team for the Model AQP. Each of these sources provides a useful and unique perspective. The ISD methodology, as currently implemented, focuses of the technical skills, knowledge and procedures required to accomplish specified tasks and subtasks. The methodology is especially effective for activities that occur at predictable times and in a standard order. The event set approach, in contrast, focuses on a selected sequence of situations, which attempts to mimic real-world situations with all of the attendant complexity. The objective is to evaluate crew performance in situations that require pilots to utilize both technical and CRM skills. This approach is currently used by many carriers to support line operational simulation. This research project's situation assessment model attempts to provide a cognitive perspective by focusing on those factors that influence a flight crew's assessment of a situation and subsequent management of available resources. In the past there has been little continuity between the task-oriented front-end analysis provided by the ISD methodology and the situation-oriented event set approach. This lack of continuity is exemplified by differences in types of CRM skills addressed by each approach. The ISD methodology is best suited to handle phase-specific skills, that is, those activities which are always performed by the crew for a given task or subtask. Traditional ISD behavioral orientation supports its emphasis on the specific task and subtasks that must be performed to complete a job.

This task orientation does not support those aspects of the pilot's job which fall outside the sequential tasks and subtasks found in a task list. Instead, the unique dynamics of the aviation environment necessitate a change in focus to the situation as a whole, including conditions under which a task or subtask must be performed (e.g. weather, aircraft system, failure) and the requirement to utilize phase-independent flight management skills, either on a need basis or continuously, to ensure that the flight is properly managed. Appropriate utilization of phase-independent skills depends upon crew judgement: accurate assessment of the requirements of the situation, together with effective utilization of those skills and information sources most likely to be useful in that situation. This judgement depends upon an understanding of the situation as a whole, not simply the task in isolation.

This situational focus is the strength of the event set approach. An effective training program will enable flight crews to experience these situations so as to allow them the opportunity to practice the phase-independent skills required to cope with these situations.

Both the ISD and event-set methodologies bring important and unique perspectives to an AQP. Merging them into a coherent approach will support the development of a complete training program. One means by which this integration can occur is to place the focus of flight training on situations instead of tasks throughout the program rather than waiting until LOFT. A situational orientation throughout training helps to ensure that phase-specific and phase-independent technical and CRM skills are practiced in an integrated fashion. In addition, the situational orientation gives flight crews the opportunity to practice those skills involved with assessing situations and utilizing available information.

This transition from the task focus of the ISD methodology to situation orientation takes place in the Model AQP by means of the concept of an event. An event includes a specific task (i.e. a maneuver or set of procedures) together with the conditions (weather, malfunctioning aircraft system, etc.) under which the task is to be performed. To handle an event successfully requires that the crew quickly and accurately assess the situation, plan how to manage the event, and utilize the technical and CRM skills appropriate for that event. In addition, the set of events included in curriculum can be selected to ensure that important technical and CRM issues are addressed. Each event has a specific topic or theme that is the point of that event. The flight training curriculum can be designed by strategically selecting and positioning events in accordance with these themes.

One of the strengths of the event concept is its applicability to both ground and flight training. Continuity throughout all parts of an AQP is a critical goal for the Model AQP project. Events can be used as the building blocks for both the ground and flight training curricula. For ground school, one of the goals of the Model AQP is to utilize scenario-based training, where students would be required to not only acquire new information but also learn how to apply that information to solving problems. Events are a natural tool for designing a scenario-based ground school curriculum. Similarly, events can serve as the individual units for Flight Training Device (FTD) and simulator training. Finally, they will continue to serve as the building blocks for LOS scenario development.

The first-generation Model AQP, which has been targeted to regional carrier training, is complete and in use by approximately 20 carriers. Several workshops have been conducted to train air carrier personnel in the use of this model.

During FY 98 development of the Advanced Model AQP was initiated. This second-generation model will incorporate many features requested by the carriers, allowing each carrier to pursue its own level of development and will include a performance proficiency database. Currently the specification documents have been written and approved and development of the second-generation Model AQP is scheduled to begin in the first quarter of FY99. This version of the model will attempt a "modified develop and test" approach, thus completion will depend somewhat on the modifications required by industry and AFS-230.

Products

- Model AQP and research report delineating process, methodology and lessons learned.
- Advanced Model AQP.

Project Title
Line-Oriented Flight Training (LOFT) Debriefing Study

Performing Agency
NASA-Ames

NOTE: This project was completed in FY 97. It is presented here because of the high volume of requests for the training materials.

Background
Training effectiveness of LOFT simulations is heavily influenced by the debriefing that occurs afterwards. The LOFT simulation is a very busy, intense, and sometimes stressful experience. Thoughtful discussion after the experience is necessary in order for the crew to sort out and interpret what happened and to consolidate the lessons learned into long-term memory in a form that can be used later in actual line operations. The debriefing is a window on the entire CRM process. LOFT debriefings can demonstrate how well crews are able to analyze their performance along CRM dimensions. It is thought that in order to implement CRM effectively in day-to-day line operations crews must have the skill and the habit of analyzing their own performance in terms of CRM.

Project Description
The purposes of this project are to determine: which techniques are actually being used by LOFT instructors; the effectiveness of the techniques; the extent to which those techniques are consistent with FAA Advisory circular guidelines; the practicality of the guidelines for real-world training; and what obstacles instructors encounter while teaching to these guidelines.

Data have been collected from five major US air carriers. At each carrier five to eleven instructors from different fleets made audiotapes of the debriefings they conducted. The audiotapes are transcribed and coded, generating a large database. Findings are as follows:

(1) Typically, the debriefing is scheduled after the practice of proficiency maneuvers, rather than immediately after the LOFT. Interposition of the proficiency maneuver practice constitutes an "interference experiment" that probably impedes the ability of the crew to remember what happened in the LOFT.

(2) The similarities among carriers are greater than differences in how they conduct LOFT and LOFT debriefings. There are large differences among individual instructors within carriers in facilitation effectiveness.

(3) Most instructors conscientiously attempt to elicit crew participation, but some unwittingly sabotage their own efforts with behaviors that discourage participation.

(4) Most instructors fail to make clear at the beginning of the debriefing the nature of the participation that they expect of the crew and do not explain why it is important that the crew take an active role.

(5) Although instructors have been told to facilitate crew self-debriefing, they have been given only very general advice on how to do this. Training departments should provide much more explicit hands-on training in facilitation and should mentor new instructors as they start to facilitate.

(6) The content of the debriefing is strongly driven by the instructor's observations and questions, rather than the crew's self-analysis.

(7) The crews are responsive to the instructors, but are more reactive than proactive.

(8) Most crews would probably have difficulty in conducting a deeply analytical debriefing on their own, however they appear to be quite conversant with and accepting of CRM and with practice, could develop debriefing skills and apply them routinely on the line.

Products

- CRM debriefing training guidelines were distributed in FY97.
- A handbook for instructors was developed and distributed in FY97.
- The handbook was also incorporated into the instructors' manual for leadership/followership curriculum development at Alaska Airlines, distributed in FY98.

Project Title
A Methodology for Developing “Gold Standards” for Rater Training Videotapes

Performing Agency
NASA-Ames and American Institute for Research

Background
Central to the Advanced Qualification Program (AQP) is the training of check airmen to provide reliable, consistent, and valid assessments of Crew Resource Management (CRM) skills. It has been suggested that to enhance reliability of CRM assessment, measures should focus on clearly defined, observable behaviors. These behaviors should be explicitly identified so that no ambiguity exists for check airmen conducting CRM assessments. Raters should be calibrated so that check airmen are consistent in their evaluations. Check airmen should be trained by rating and discussing videotapes of aircrews.

Project Description
The purpose of this project was to develop and validate a methodology for constructing a "gold standard" for rater training videotapes. Gold standards are defined as criteria that are assigned by experts to a training videotape. These tapes can be used for training rater reliability and accuracy.

To meet this objective, the five-step methodology, shown below, was developed:

Step 1 Select Videotapes of LOS Scenario

Step 2 Develop Gold Standards Data Collection Forms

Step 3 Conduct Gold Standards Development Meeting

Step 4 Analyze Gold Standards Data

Step 5 Establish Gold Standards

The methodology was tested at an air carrier that was interested in developing gold standard videotapes for check airmen rater calibration training. CRM, technical and event set gold standards were successfully developed for a videotape of a crew performing an LOE scenario. A technical report was published and distributed to industry in FY98.

Although this methodology is hypothesized to increase the level of rater calibration, it is costly to implement. This research team is currently working with a major air carrier to determine if the gold standard methodology for rater calibration is substantially better than inter-rater reliability training. Inter-rater reliability training focuses on raters agreeing with one another regarding the level of crew performance, rather than the gold standard methodology of calibration to a set standard. Inter-rater reliability training is somewhat less costly to implement since it does not require the rigorous development of the training videotapes. Data collection is underway at a major carrier and is expected to be complete in FY99.

Products
- Research reports.

Project Title
Integrating Crew Resource Management (CRM) into Crew Procedures: Checklist Standardization

Performing Agency
NASA-Ames

Background
Research on this project is not on-going but is presented here because it is the foundation for the Operating Documents Human Factors Project, described next.

Non-compliance with checklists has become a potentially serious problem for air carriers and the FAA. In today's aviation system, procedural changes occur frequently and are instituted for a large variety of reasons (company mergers, changes in air traffic procedures, changes in airport environments, introduction of new technologies and new aircraft, etc.). While one obvious goal is to maintain compliance and standardization of crewmember performance, there must not be a sacrifice in the match between procedures and the operational realities in which the procedures are to be used. Not only must procedures reflect a reasonable match to technical task demands, they must reflect consistency with company and system requirements. Within an air carrier for example, philosophy, policies and procedures should be made consistent both within teams (within aircraft type) and across teams (across fleets within the company) to the extent that standardization is feasible. Since there are legitimate aircraft and mission differences across fleets, a means of evaluating the appropriate level of standardization must be developed and reviewed when changes are anticipated.

Project Description
The goal of this research project was to integrate CRM principles into crew roles and procedures in order to: 1) improve the match between current procedures and operational reality, and 2) develop a process for adapting procedures to anticipated changes.

The basic approach was to analyze procedures from a team perspective and to conduct a team task analysis in order to assess whether current procedures match the task flow including team processes. The integration of CRM principles into the task analysis provides the basis for reviewing and re-designing procedures. The development of standard methods and measures for evaluating procedure usage is important for assessing both current and proposed procedures.

The objectives of this research project were to develop a systematic approach to checklist standardization across all fleets, and to develop guidelines for implementing and evaluating procedure changes that optimize within-company standardization, compatibility with aircraft and operational differences, and ensure that the procedures are consistent with CRM principles. This research is complete and has established company-wide philosophy and policy statements regarding operations, automation, training, and checklists. Also, methods for data collection have been identified and developed. Pilot feedback on proposed checklist changes has been completed.

Project Products
Documents internal to the carrier.

Project Title
Operating Documents Human Factors Project

Performing Agency
NASA-Ames and George Mason University

Background
The original project described earlier was intended to develop a systematic approach to checklist standardization and develop guidelines that would optimize within-company standardization for eight fleets, ensuring compatibility with aircraft and operational differences and consistency with Crew Resource Management (CRM) principles. This project considered the form of the checklist with respect to content, structure, logic, format, terminology, roles and usage. However, this approach was inadequate, as checklists are part of the whole operating document system. It became apparent that operating documents must show internal consistency across fleets (aircraft types, and route structures) and across departments (pilot's handbooks, company policies, etc.). Further, these documents must be externally consistent with regulations and manufacturers. Document systems must be consistent across philosophies, policies, procedures and practices. This concept is not in place in the air carrier industry.

Project Description
This research team developed surveys and held workshops to bring together air carriers (including regionals and cargo), manufacturers, and the FAA to work on a better, more human-centered approach to these document systems. Seventy-eight participants, indicating industry's interest in this topic attended one of the two workshops held in FY97.

Based on the information provided on the surveys and in the workshops, guidelines are being written for the development of operating documents that will include: the Organization of Documents, Standardization of Documents, Usability of Documents, Developing and Maintaining Documents, and Transition to Electronic Media.

Feedback from the air carriers indicates the need for additional workshops. The structure and content has not been determined nor has it been determined whether this process should be turned over to industry to lead. Feedback from internal FAA sponsors indicates that the integration of AQP documentation should be considered in future work by the research team. This focus is presently being considered.

Products
- Reports from the two workshops were written and distributed to industry in FY98.
- "Developing Operating Documents Manual" is currently being developed and expected to be ready for distribution in FY99.

Project Title
Team Situational Awareness

Performing Agency
Naval Air Warfare Center Training Systems Division (NAWC-TSD), Orlando, FL

Background
There is little question that situational awareness is critical for safe flight. Much research has centered on attempts to identify and quantify individual situational awareness. However, since cockpit crew performance depends heavily on team processes, it is necessary to determine exactly what aspects of individuals and the environment enhance or degrade team situational awareness.

Project Description
This research project has identified a model of team situational awareness, developed effective instructional strategies to improve situational awareness training, developed situational awareness assessment tools, and developed guidelines for situational awareness scenario development.

The initial framework for team situational awareness has been completed. This framework will serve as a basis for the team situational awareness model. Preliminary training guidelines based on this framework follow:

(1) Tailor training to the experience level of the crew in terms of time in aircraft, time in position, and familiarity with locality. For example: consider a newly hired first officer or flight engineer that is transitioning to or learning a new aircraft and flying out of a new area. It is important to ensure that the crew member is familiar with cockpit roles, the geographic area, and the aircraft equipment before expecting him/her to have the awareness to handle dynamic problem situations in that environment.

(2) Ensure job/task analysis provides information about the cognitive, perceptual and behavioral demands required for situational awareness. For example: a job analysis of a pilot who flies a highly automated or "glass cockpit" aircraft should solicit information particularly relevant to that environment, emphasizing factors such as seeking information from different sources, collation and interpretation of information from complicated displays, scan through the FMS, mode errors with the FMS, manual data entry task demands, and communications associated with the advanced technology.

(3) Include training which specifically focuses at the team level of situational awareness in addition to training which focuses on individual skill development. For example: individual skill development in situational awareness may include: providing information that increases the individual's knowledge base (so that he/she knows what information to seek and when), helping in the development of planning skills (so that he/she is better prepared for completing tasks within the flight), and providing tips that help free up time from basic tasks in the cockpit (so that more attention may be paid to the situation). Team skill development in situational awareness should include team process skill training such as communication and leadership behaviors to enhance and maintain crew awareness.

(4) Based on the complexity of the situation, provide training which emphasizes adequate information, demonstration, and active practice and feedback, all of which are needed to produce a behavioral change. For example: to enhance the overall training experience provided in a realistic LOFT scenario emphasizing situational awareness, the crew members should receive advanced preparation about the training objectives, including specific information on the skills to be trained and demonstrations or examples of effective and or ineffective behaviors relevant to those skills. Providing general information about situational awareness concepts would be insufficient to expect the crew to obtain the full impact of LOFT. Crewmembers who have a low experience level need

to have specific information on the kind and timing of information transfer that will enhance crew situational awareness. Alerting them to the importance of communication, if they do not know what and when to communicate, would not prepare them to practice good communication skills in the LOFT session.

(5) Design training scenarios to be realistic and to elicit the desired cognitive, perceptual and behavioral task demands encompassing the situational awareness construct. For example: consider that including an unfamiliar event, such as icing, for crews who have only flown in Florida would require them to recognize the relevant cues, comprehend their meaning and project a course of action. A very different type of problem, such as a gradual change in an instrument due to a developing condition, may require perception comprehension and projection as well, but the cognitive and behavioral demands related to this problem are very different from those imposed by icing conditions. Including both types of problems within a scenario would allow the trainee to exercise these different cognitive and behavioral skills, and would help him/her to learn to generalize the application of these skills across various problem situations.

(6) Conduct training which focuses on improving team or process skills related to situational awareness (e.g: planning, communication, leadership) because this may aide the development of team situational awareness. For example: training focusing on communication skills related to situational awareness can include: training that encourages the flow of the information that needs to be transmitted in the planning phase of the flight (e.g., what questions the Captain can ask to solicit information, what questions the First Officer should ask, what information should be covered with Flight Attendants.) It can also include practice in the communications that are most effective (e.g. what information to provide to another crewmember or when to communicate the information to the other).

(7) Include training which emphasizes developing accurate and appropriate knowledge structures for expectations related to situational awareness. For example: training a crew about how traffic flows around a specific airport, prior to arriving there for the first time, helps the pilots understand what can be expected, thereby increasing the readiness with which he/she can recognize that an abnormal situation is occurring.

(8) Use realistic and dynamic scenarios to provide training on: information seeking and processing; detecting anomalous information; prioritizing attention to cues, especially when there is information that may be related to different problems; and taking appropriate actions based on cues. For example: design realistic scenarios with events that require crewmembers to: look for and use information; handle multiple cues; determine when information is signaling a problem or possible problem; and act on the information. Intermittent fluctuations of an instrument during a scenario require crewmembers to recognize the fluctuations and to diagnose the problem. This may include cross-checking instruments, discussing among themselves, and seeking additional information. Putting this event in the scenario at the time of moderately heavy workload or when another troubleshooting event is occurring gives crewmembers practice in dividing their attention and determining priorities.

(9) Design and develop performance measurement tools for the behaviors, the knowledge and perceptions required for situational awareness. For example: a realistic scenario provides information about situational awareness behaviors and an observation scale can be developed for those behaviors. Tapping into the crew member's knowledge and perceptions may require asking questions, either during the scenario or after it is completed to determine if he/she saw the problem and understood its possible impact. For pilots with low experience levels, questions can be either a part of the scenario (e.g. ATC can question the crewmembers) or the scenario can be stopped briefly for questions to be asked.

(10) Measurement tools should measure individual situational awareness skills as well as team interactions related to situational awareness. For example: an observation scale can be designed that measures what individual team members contributed to the team situational awareness and

what actions they performed independently that demonstrated their situational awareness. The same scale should include opportunities to document team interactions (e.g. planning, communications, and leadership) that helped the crewmembers gain and maintain team situational awareness.

(11) Situational awareness measurements of behaviors should include both outcome and process measurements. For example: outcome is the action taken by the crew in response to an event (e.g. the crew diverted to Airport A) and process is how they achieved the outcome (e.g. the Captain gathered information from others about the decision, crewmembers discussed the merits of each airport, the crewmembers noticed that the runway at Airport B was not adequate for their needs). Both the outcome and process are important to understanding performance.

(12) Include instructions and information about training objectives to instructors about each training method or tool. For example: write out the specific objectives to be accomplished at the top or on a cover sheet of scenario instructions. Include clear objective behaviors as part of the scenario goals. The instructor is then informed about how to use the measurement tool to reinforce the training objectives.

(13) Link the measurement criteria to the scenario events and to the debrief plans, for consistency in training. For example: if a generic measurement instrument is used, prepare the raters to recognize and document specific examples observed in the scenarios that relate to the general behaviors on the rating form. It is likely that the raters will need thorough training to map the scenario events and behaviors observed to the rating form. Thus, in the debrief, crewmembers will know what specific actions they took and how those relate to general requirements for situational awareness.

(14) Multiple measurements over time are necessary to yield a good measurement of situational awareness; single measurements at one time are insufficient. For example: in a single LOFT scenario, build a number of opportunities for crew members to demonstrate situational awareness (e.g., NOTAM information, ATC calls to other aircraft, PIREPs, subtle changes in instruments, weather, unexpected circumstances) and make sure that observers know when, where, and how to observe and document situational awareness actions. The same could be accomplished by using a number of small scenarios (each with opportunities to observe situational awareness) and then evaluating the aviator's situational awareness across the scenarios.

(15) Use specific feedback tools along with realistic and dynamic scenarios to provide feedback that; is specific to the trainees on their situational awareness skill performance; reinforces the development and maintenance of relevant knowledge; and stresses adapting knowledge to meet situation demands. For example: give observers feedback forms with the specific situational awareness elements that were built into the scenario. Observers should use these forms to document behavior they see in the scenario. This is then used to give specific feedback on what the crewmembers did, how they did it, and the context of the behaviors. Explanation of the rationale for the feedback helps the crewmembers develop and add to their existing knowledge.

(16) Train instructors on making ratings and giving feedback. For example: This training can be done by providing information to the raters on situational awareness, showing them examples of the behaviors associated with situational awareness (so raters can be knowledgeable about the behaviors), giving them practice in rating those behaviors, and giving raters practice in providing feedback on specific situational awareness behaviors.

The above guidelines were expanded, published and distributed in February 1998. The guidelines are too lengthy to be contained in this report. Although these guidelines have received positive comments from the air carrier training community, there remains a fundamental problem with situational awareness training and assessment. Several carriers believe that new instructors, and line pilots, in general, are not able to dissect pilot performance and to identify and articulate when and what exactly led to poor situational awareness. Thus, crews and instructors are not adequately debriefing flight scenarios, either in training or on the line. In response to this critical requirement, this research team is examining the idea of designing a Computer-Based Training (CBT) system to provide the fundamental principals in situational awareness cue recognition. If it is considered appropriate, the system will be designed at one carrier for pilot training and at another carrier for instructor/evaluator training. If the research is initiated in FY99 it is expected to be complete in FY00.

Products

- Situational Awareness guidelines.
- Situational Awareness training systems.

Project Title
Decision-making and Judgement on the Flight Deck

Performing Agency
NASA-Ames and the Naval Air Warfare Center Training Systems Division (NAWC-TSD), Orlando, FL

Background
The National Transportation Safety Board (NTSB) reports that in a four year period, 1983-1987, crew judgement and decision-making was implicated in 47% of fatal accidents. NTSB and ASRS cites many more reports where poor decision-making was evident in many potentially serious incidents. Decision-making is a component of most CRM courses and is encouraged by the FAA Advisory Circular 120-51, 1993. However, scientific research has not been available to support this training. Decision-making in the dynamic, time-constrained cockpit environment does not mirror decision-making in a static laboratory setting. Thus, findings from generic decision-making research do not necessarily generalize to aviation settings.

Project Description
Decision-making is an essential component of a Captain's expertise. The Captain is responsible for making the hard decisions (i.e., choosing where to divert after a system malfunction when fuel is short and weather is deteriorating; determining how to cope with a passenger's medical emergency; evaluating whether to take-off with a placarded system given past experience with the projected weather and traffic at the destination). The Captain's judgement is most critical when conditions are ambiguous and no clear guidance is provided in manuals, checklists, or company policy.

This research project has been in progress for several years and has developed a model of factors that make flight decision efforts difficult, along with a taxonomy of decision problems. This model incorporated the nature of the problem, the time available to solve the problem and the risks associated with the options. This work was based on the analyses of crews facing abnormal and emergency events in full-motion simulators and further supported by NTSB accident analyses and ASRS reports. For example, analyses revealed differences in the importance of risk and time pressure between Captains and First Officers. For Captains, risk was the most important variable in the decision strategy, while First Officers viewed time pressure as most important. As might be expected, experience does affect decision strategies. Experienced crewmembers are more sensitive to available decision time, better at handling uncertainty and have better prioritization of information needs. Based on these results, it is suggested that to encourage effective decision strategies, crews are trained to: (1) understand and verify the problem, (2) assess time and risk, (3) not rush to judgement, (4) consider constraints and consequences of options so as not to oversimplify the problem, (5) use "worst case" instead of "best case" reasoning, (6) plan for contingencies, and (7) manage the situation by setting priorities, assigning tasks and using all available resources.

This research team developed decision-making, situational awareness and planning training scenarios and assessment instruments for a major carrier. They also developed debrief instruments in concert with the training scenarios. Currently, data is being collected on the effectiveness of these debrief materials. This part of the project is expected to be complete FY99. This research team is also developing training guidelines to manage risk in ambiguous situations. It is planned that the error taxonomy database will be incorporated into the Rapid Reconfigurable LOE project described earlier. This research team is working with four major carriers to develop "challenging" training guidelines. The training guidelines are expected to be complete in FY00.

Products

- Decision-making, situational awareness, planning and challenging training and assessment guidelines for air carriers to incorporate into their CRM training or into AQP.

Project Title
Interruptions, Distractions and Lapses of Attention in the Cockpit

Performing Agency
NASA-Ames

Background
Interruptions and distractions are two of the most common causes of pilot error incidents and have contributed to many accidents. One of the most striking of these accidents was the Eastern L1011 accident in the Everglades. Rather than attenuating this problem, glass cockpits have, if anything amplified it. Closely intertwined with the issue of interruptions and distractions are problems with habit capture, tunneling of attention and failing to remember to perform deferred actions.

Project Description
This project will: characterize the interruptions that most frequently occur, the types of situations conducive to distraction, and factors that impede recovery from distraction; identify specific techniques crews can use to control interruptions, recover from distraction, avoid habit capture and prevent tunneling of attention; identify ways to systematically design interruptions and distractions into LOFT/LOE scenarios to realistically challenge crews' task management skills; and explore ways to modify cockpit Standard Operating Procedure (SOP) to reduce this form of crew error.

This project has analyzed ASRS incident data and NTSB reports that were the foundation for the NTSB study of crew-caused accidents. This analysis identified several plausible cognitive vulnerabilities conjectured to underlie pilot errors and identified situational factors that contributed to those vulnerabilities. Attention lapses were involved in over half of all these accidents. Interruptions and distractions were involved in 13 accidents; in eight of the accidents the interruptions or distractions were self-induced by the crew, in the other five they were externally-induced. This project has analyzed 107 ASRS reports that involved lapses of attention in air carrier operations (Part 121). These 107 reports involved 21 different types of routine tasks crews neglected at a critical moment while attending to another task. Sixty-nine percent of the neglected tasks involved either failure to monitor the current status of position of the aircraft or failure to monitor the actions of the pilot who was flying or taxiing. Thirty-four different types of competing activities distracted or preoccupied the pilots. Ninety percent of these activities fell into one of four broad categories: communication (e.g., discussion among crew or radio communication), head-down work (e.g. programming the FMC or reviewing approach plates), searching for VMC traffic or responding to abnormal situations.

Based on this information, this research team will conduct a field study and a laboratory study to address questions arising from the analyses. For the field study, a questionnaire will be designed to determine whether highly experienced pilots and instructors have developed techniques with which to reduce their vulnerability to various situations when they must manage more than one task concurrently. The questionnaire will present prototypical cockpit situations and ask if the respondent uses any particular techniques in these situations to avoid neglecting either of the concurrent tasks. Analyses of this data will be used to refine the questionnaire for further data collection.

Prior research suggests that cockpit errors in managing concurrent tasks are not simply a matter of overload. Rather, in many situations it seems that a central problem for pilots is to switch attention back and forth to monitor the status of both tasks in a timely manner. A laboratory study will be designed to investigate the cognitive processes that drive the switching of attention between two complex tasks and to determine the factors that disrupt switching.

From this data, specific training guidelines will be developed to aid crews in controlling interruptions, recovering from distraction, avoiding habit capture and preventing tunneling of attention. Also, this research will provide methods to design interruptions and distractions into simulator training to realistically challenge crews' task management skills and ways to modify cockpit SOP to reduce this form of crew error.

Products

- Training guidelines and scenario development methodology.

Project Title
Establishing Relationships Between Flight Data Parameter Values and AQP Qualification Standards Using APMS Methodology.

Performing Agency
JIL Information Systems

Background
The overall goal of the Automated Performance Measurement System (APMS) program is the development of tools and techniques to improve the efficiency with which air carriers collect and analyze Flight Operational Quality Assurance (FOQA) data. Part of this program is being conducted by NASA-Ames and is funded by NASA and AAR-100. That particular project is not described in this report, as the work does not focus on training data but rather on operational data. Details can be obtained from FAA Flight Standards, AFS-230 or AAR-100. The present project is closely related to the overall APMS project in that it will use some APMS-developed tools to analyze simulator data for flight parameters.

Project Description
The goal of this project is to develop methods and techniques that will allow air carriers to determine the extent to which daily flight operations are consistent with training qualification standards established in an AQP program.

This project has developed a system to acquire and record simulator flight data at the same level of fidelity and resolution as a FOQA-equipped aircraft and has developed a baseline, flight data parameter profile of each AQP Qualification Standard. This is a preliminary “string” of flight data parameters and conditions that together would identify each AQP Qualification Standard.

The software to collect the simulator data has been developed and is being used by a major carrier. The research team is presently collecting data on the flight parameters. The prototype capability to compare FOQA data and training data has been demonstrated. In FY99 software development will continue and the transformations of qualification standards to aircraft parameters will be refined.

Products
Software tools and methodology to link FOQA data and AQP qualification standards.