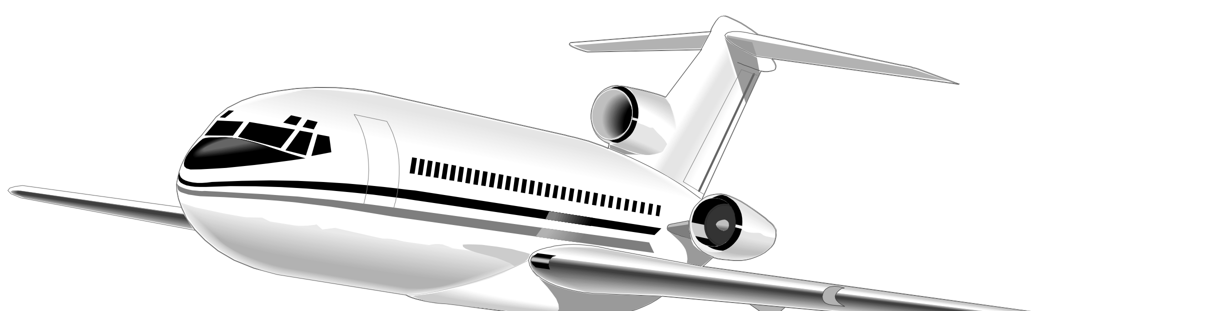
Eleana Edens, Ph. D.
Office of the Chief Scientific and
Technical Advisor for Human Factors, AAR-100
Federal Aviation Administration

Air Carrier Training Research Review

June 2000



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. Research reviews are conducted by this office. The following report summarizes the March 2000 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) and Line-Oriented Evaluation (LOE) development, implementation and evaluation. Thus, this research centers on (1) crew training and assessment, (2) instructor training, (3) LOFT/LOE development strategies and (4) organizational and systematic influences on pilot performance, including automation usage.

Address questions or comments to:

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

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 as well as FOQA data.

Project Description

The Aviation Performance Measurement System (APMS) is a research and development effort whose primary goal is to devise practical systems, methods and techniques for air carriers in the use of digital flight data recorded during line operations. The Flight Crew Training Management and Support emphasis area of this project seeks to develop the tools and techniques to allow flights standards and training organizations to make use of digital flight data. Work in this emphasis area has been closely linked to the other emphasis areas of the overall APMS project.

The primary task of the Flight Crew Training emphasis area is the analysis of flight data from line operations to assess the aggregate proficiency of the flight crew population, and to determine which operational tasks and task components should receive more or less emphasis in continuing qualification training. In order to carry out these analyses, detailed formal standards of proficiency for flight crew operations are required. These standards are most clearly established at carriers who have completed the development of an Advanced Qualification Program (AQP). AQP development is based on a detailed task analysis of each flight crewmember's job, and results in a detailed set of qualification standards that are directly derived from job tasks. Accordingly, a major objective of the Flight Crew Training emphasis area is the identification of flight data parameter profiles which predict levels of proficiency in AQP Qualification Standards.

The approach taken to establish these relationships is to gather flight data during simulator operations and determine the statistical correlation between grades given to trainee performance and variations in various flight data parameters. Once these relationships are established and measured, they can serve as a starting point for researchers to study the relative level of proficiency with which these same qualification standards are performed during line operations.

The development of a training capability for APMS required that FOQA data gathered during line operations be interpreted with respect to an airline's standards of training. In order to do this, the relationship between flight data parameters and specific training standards must be established. Under an AQP program, proficiency evaluations in the simulator take place under line operational conditions, involving each crewmember, and encompassing cockpit resource management (CRM) elements as well as technical tasks. Therefore, the approach taken in this project is to gather digital flight data from a flight simulator during the performance of a graded event, and to statistically determine the relationship of flight data parameter values to variance performance evaluations.

Gathering and analyzing flight data from both simulator and aircraft operations required that a number of preparatory tasks be carried out, involving a range of technical skills. Tasks include tool development, data collection, data analysis, report/product development, and project management. While work is ongoing in all product areas, promising results have been obtained in terms of predictive models, and this year will be spent in adding and extending models, and developing a set of routine procedures within the cooperating airline AQP program to utilize them in planning and implementation of the recurrent training cycles.

Products

- Modification of the MD83 and 737 simulators at Alaska Airlines to support parameter data export (Software products)
- Data collection and conversion software (Software product)
- Data visualization and custom data snapshot archive software (Software product)
- Integration of Flight Training Management System (FTMS) grade data with simulator data collection files (Procedure product)
- Conversion of relevant AQP performance standards into parameter data transforms to provide search patterns for identification of events of interest in simulator and line operations data (Data Table product)
- Predictive models based on simulator data and associated grades for evaluation of deidentified parallel line operations events (Software products, logic table and procedure products)
- Data quality tools (Analysis and procedure products)

Note

The research described in the next section, conducted by the University of New Mexico, is part of this project.

Project Title

Training and Assessing Air Crew Skills: Methods to Achieve Reliable and Valid Performance Data

Performing Agency

University of New Mexico, Albuquerque

Project I: Relating Flight Parameters to Behavioral Grades

Background

AQP requires a detailed task analysis of each flight crewmember's job, resulting in a detailed set of qualification standards that may be used as a metric by which to assess a pilot's proficiency. These qualification standards are presumably used by Instructor/Evaluators (I/Es) when grading Maneuvers Validation and Line-Oriented Evaluations (LOEs). The question that is explored in the first phase of this project is whether specific combinations of flight parameters are able to predict I/Es' grading of critical maneuvers in the simulator during Maneuvers Validation. If I/Es' grading of critical maneuvers is based on qualification standards, then it can be inferred that the predictive flight parameters are also sensitive to qualification standards. This research is important in its own right in that it reveals the degree to which behavioral grades are based on objective flight parameter data and at the same time it explicitly links flight parameters to qualification standards. This information can then be used to facilitate the assessment and training of I/Es. A second, and the ultimate goal of this project, concerns the use of this information as a basis for assessing aggregated FOQA data in the hope that it will identify those FOQA parameters that are sensitive to deviations from qualification standards. This would allow carriers to use FOQA data as an additional means of assessing the proficiency with which certain qualification standards are met during line operations.

Completed Research

Flight parameter data from MD 80 Simulators on Rejected Take-offs, V1-cuts, Normal Landings, and Normal Take-offs have been collected and analyzed. We expect to begin analysis on Approach data during the second quarter of FY 2000.

The flight data that UNM receives from the carrier is in the form of an EXCEL file, typically containing approximately 220 flight parameters recorded over various time intervals at a resolution of 1 Hz. On receiving these data, various transformations based on Qualification Standards and SME input are performed. The transforms typically covert a temporal data stream into a single statistic (e.g., mean air speed). To facilitate the processing of these data, a software tool was created that allows a user to graphically represent a flight parameter over time for various flights and from various levels of grades. The visualization of the data helps to identify what parameters may be predictive and perhaps what transforms on those parameters to compute. In addition, the tool allows a user to carry out many types of transforms on the continuous flight parameter data resulting in statistics that then serve as predictor models for the next step.

These statistics are then entered into a model to predict the behavioral grade for a maneuver. To date, four differing modeling techniques(standard multiple regression, logistic regression, discriminant analysis, and most recently classification tree analysis) have been employed. The results indicate that the classification tree analysis approach provides the best prediction of behavioral grades using the fewest predictor variables. The table below summarizes the percent of time the model correctly predicted the behavioral grade.

Maneuver	Sample size	Predictors	Hit Rate
NTO	96	6	93%
RTO	50	7	94%
V1-cut	94	6	88%
NL	100	8	96%

Planned Research

Normal Approaches. The sample of maneuvers will be extended to include Normal Approaches. FOQA from various carriers indicate that there are a significant number of unstabilized approaches, thus it is important to determine which qualification standards underlie the approach tasks and how these tasks are being graded during training.

Validate Models. All of the findings reported in Table 1 need to be cross-validated with an independent sample of cases. Analyses on 100 additional cases for each of the maneuvers would be ideal. This would reveal whether the models this research has developed will generalize across different samples of crews.

Generalize Models to FOQA Data. The next step in this project is to use the models developed from the simulator data to "grade" FOQA data. This will provide a quantified basis for evaluating FOQA data and a more direct test of the link between Simulator performance and Line performance.

Project II: The Validation of LOE Assessment

Background

The assessment of CRM skills relies primarily on LOEs, where pilots/crews are observed performing standard operations (e.g., take-off, approach, landing, etc.), in a full-flight simulator, often under abnormal conditions. An LOE allows an assessment of the crews' ability to integrate various CRM skills such as workload management, or decision making with the various required technical skills. An LOE often involves several phases of flight, beginning with preflight checks and concluding with the crew pulling up to the gate. LOE performance is typically graded at different levels, ranging from an overall LOE grade, to the grading of each event set, and finally to the more specific grading of several performance indicators or observed behaviors (OBs) that are contained within each event set. These OBs are usually designed to assess specific qualification standards related to both technical and CRM skills. In the case of CRM skills they are assumed to measure specific CRM subskills (e.g., workload management, situational awareness, crew communication, decision-making, etc.).

Whether an OB truly measures a specific CRM subskill depends on the validity of the task analysis that served as the basis for the qualification standards, the test designers' ability to translate a qualification standard into a specific OB, and an I/E's ability to use the OB in the intended way. Given the importance of LOE grading in AQP it is necessary to have a clear understanding of what precisely is being measured with these grades. A rationally driven task analysis approach to defining CRM skills is a necessary first step toward understanding these skills. However, eventually development of an empirical basis for defining the taxonomy of CRM subskills and the accompanying evidence that OBs (or whatever is used to measure specific skills) are capable of discriminating among the different subskills must be accomplished. In this project, continued research efforts will demonstrate how an empirically driven approach can provide a unique and important contribution to our understanding of CRM and technical skills.

Methodology

The empirical method developed for analyzing OB data begins with correlations between pairs of OBs, where a single pilot (or crew) provides scores on both OBs. Mean correlations across sets of OBs, where a set of OBs is defined by some common category such as a skill category are examined. For example, to determine whether OBs are measuring CRM subskills, crew performance on a single OB (assumed to measure a specific subskill type, such as decision making) was correlated with crew mean grade across all other OBs assumed to measure the same subskill (e.g., decision making). We did this for each OB within a specific subskill and then averaged the correlations for that subskill. We repeated this process for each subskill and then computed the average correlation between OBs measuring the same subskill. This averaged correlation was then compared with the averaged correlation between OBs that are from different subskills and different event sets. If OBs are indeed measuring subskills the correlation within subskills should be higher than the correlation between subskills. The magnitude of the difference (in terms of squared correlations) reflects the degree of variance that can be allocated to the subskill component of OB performance.

Summary of Completed Work

To date, an analysis of 249 crews is completed. This analysis, comprises three LOEs containing a total of 34 different OBs. In the course of computing the intercorrelations between Obs, it was discovered that the degree to which an OB correlated with all other OBs differed widely across the 34 OBs. These averaged correlations, ranging from .24 to .87, may be thought to reflect an OB's typicality. Those OBs with the higher average correlation are more typical OBs, whereas atypical OBs have a lower correlation with other OBs. OB typicality had an important influence on the comparisons between different groupings of OBs, necessitating the inclusion of the typicality factor in all subsequent analyses.

CRM Subskill Effect. Using both Typical and Atypical Obs, no evidence of a subskill effect was found. However, when the analysis was only conducted on Atypical OBs , 7% of the variance in OB grades was related to a subskill effect. Thus, it appears that a certain class of OBs (those defined to be Atypical) are capable of assessing CRM subskills.

General CRM Skill. The average correlation between OBs from different event sets and different subskills was .60. The square of this correlation (36%) can be viewed as indicating a general skill effect. Thus, while CRM skill is to some degree subskill specific, there appears to be a much larger component of CRM performance that is related to some general skill.

Event Set Effect. Analyses also showed an "event set" effect. The average correlation between OBs within the same event set (.70) was higher than the average correlation between OBs from different event sets (.60). Squaring these correlations (.49 and .36, respectively) and computing the difference indicates that 13% of the variance in OB grades was related to an event set effect. This result was somewhat larger looking only at Atypical OBs (18%). These findings indicate there is some context specificity to OB grades.

OB/Event Set Grade Relations. The correlation between OB and event set grades was much higher for Typical OBs (.71), than for Atypical OBs (.57). This suggests that the Atypical OBs are less dependent on Event Set grades and is consistent with the idea that they may be better measures of specific CRM subskills.

CRM/Technical Relations. Perhaps the most surprising result to emerge from the analyses was that the average correlation between CRM and Technical pairs of OBs (.68) was the same as the correlation between CRM pairs of OBs (.68). This indicates that OBs did not discriminate between CRM and Technical skills. Optimistically, this finding could indicate that CRM and technical skills have become integrated in the sense that crews who perform well on CRM are also likely to perform well on technical OBs. On a less optimistic note it may indicate that OB grading is simply not discriminating. Clearly, this question needs to be more carefully investigated.

Characterization of Discriminating OBs. One of the goals in this research is to provide a method for improving the writing of and/or selection of OBs. Although findings indicated that the atypical OBs were more discriminating of CRM subskills, it was not possible to identify any obvious attributes that distinguished them from the Typical OBs. More importantly, there was only moderate stability in an OB's typicality across different LOEs (i.e., the same OBs were not always the most Atypical across different LOEs). There are a number of factors possibly contributing to this instability in typicality that will be studied in the proposed research.

Summary of Completed Research. It appears that CRM performance is significantly influenced by at least three factors; 1) a general skill; 2) specific subskills; and 3) an event set context effect. The relative influence of general versus specific subskill effects depends very much on the nature of the OBs that are being looked at. The most important implication for assessment and training is that it may be possible to select and/or create OBs that are more effective in discriminating what they are assumed to measure. Ideally, more information with fewer items could be possible.

Extended LOE Research

Pilot Experience Effect. A primary goal is to increase the diversity and size of the sample. To date, all of the results were based on Qualification data. Data has been collected from pilots in Continuing Qualification training. These are more experienced pilots and are expected to show less context specificity and more subskill effects. It is also possible that this additional experience may reduce the Typicality effect.

OB Abstractness Effect. We now have access to similar kinds of LOE data from two other major carriers who use more abstract OBs. By conducting the same types of analyses on these items, it may be possible to determine what effect this change in item abstractness has on the assessment of subskills.

Evaluator Bias Effect. The correlations between OBs were computed across different evaluators (i.e., for a set of 50 crews, six or more different evaluators may have graded different subsets of crews). Evaluators use the 5-point grading scale somewhat differently, some tending to give higher grades overall and some lower grades. This differential in grading creates a confound that very likely influences the magnitudes of the averaged OB correlations we computed. An increased sample size will provide a sufficient number of crews for some evalutors to allow us to compute correlations within evaluators. At that point, analysis could determine what type of an effect the evaluator bias has on our estimations of correlations.

General Skill Effect. There are two important factors that may have influenced the estimate of the General Skill effect. First, the carrier from which data was obtained did not grade the Captain and First Officer independently on OBs. Thus, OB grades always refer to crew performance, and as a consequence our estimation of a General Skill refers to a crew rather than an individual. It is important to be able to estimate the consistency of both individual and crew performance across different types of OBs, however, with the present carrier, this is limited to crew consistency. Access to OB grades where the Captain and First Officer receive separate grades will allow additional analyses.

Subskill Effect. The work will be extended to a larger sample and range of OBs and LOEs. In addition to investigating the Typicality dimension across a larger and more diverse sample of OBs, some standard psychometric methods will be applied. The most obvious is to rank-order OBs in their ability to discriminate subskills for a given sample of data and then cross validate these findings with a new sample. The focus would be on those items that show good discrimination across both samples. Finally, the LOE data from other carriers will provide a way to look at the effect of OB abstractness on subskill discrimination. While it might be assumed that more specific and concrete OBs would optimize the discrimination of subskill, it is quite possible that there is an optimal level of abstractness that provides the most valid assessment of a subskill.

Empirically Derived Skill Categories. Given the relatively small proportion of variance related to the subskill effect (7%) it is quite possible that the taxonomy of subskills that was used is invalid. Access to a sufficiently large data set (30,000 individual grades over 80 different tasks) will allow analysis using clustering techniques. Development of a taxonomy of CRM subskills on a purely empirical basis has not previously been attempted. Regardless how these analyses come out they should have important implications for the grading and interpretation of OB grades.

Project III: Group and Individual Instructor/Evaluator Training and Calibration Tool (IETC)

Background/Completed Work

Over the past fiscal year, modifications to existing software, with the goal of creating a version that would be stable and usable by the airline community, was accomplished. Much of this work involved standard improvements to the code such as generalizing data structures and variables, ensuring the code would work on various platforms and configurations, and changing from the existing avi video files to the more efficient mpeg format. A more substantial change was to create software that would allow a user at a carrier (say an I/E supervisor or computer systems person) to easily create an electronic gradesheet and tie it into the appropriate tables in the Access database. This was accomplished through MS Word Forms and Visual Basic code.

Extended Work

1. **Upgrade incorporating user comments.**

2. **Extention to Maneuvers Validation.** Work has begun with two major carriers to extend the IETC software to calibrate I/Es on grading maneuvers validation. This

requires downloading a set of physical flight parameters from a simulator and then using SimAuthor replay the flight on a desktop computer. Sets of flight parameters could be obtained for various levels of performance on critical maneuvers and then replayed for training and calibrating I/Es to grade the maneuvers.

3. **Evaluation of Calibration Software.** Evaluation of the effects of calibration training by tracking I/Es' performance across repeated calibration sessions and to observe the effects of calibration on actual LOE grading performance will be performed.

Project IV: Skills Maintenance and Reacquisition Training (SMART)

Background

The goal of SMART is to determine the decay function of pilot performance skills and for those skills that show significant decay within a 12-month interval develop cost-effective low fidelity retraining tools. Based on the literature review ,it may be safe to conclude that there are very few generalizations that can be made regarding the skill decay. Because a large range of factors seem to influence decay rate it is dangerous to generalize from previous findings to situations that differ in any significant way. As a consequence, we believe that it will be necessary to design studies and collect the data relevant to specific training situations of concern. This, for the most part, concerns critical maneuvers that are not practiced on a regular basis (e.g., rejected take-off, V1-cut, etc.).

Proposed Research

Three major carriers have agreed to provide performance data. Analyses will begin during the third quarter of FY 2000.

Project Title
Pilot Training and Evaluation: Flight Simulator Fidelity - The Effect of Platform Motion

Performing Agency
Volpe Center, Cambridge, MA

Background

This project is part of the Federal Aviation Administration's (FAA) initiative towards promoting the availability and affordability of effective flight simulators for all U.S. airlines. Simulators provide a safe and effective means for pilot training and evaluation, enabling presentation of scenarios including emergencies requiring both technical and crew resource management skills. Therefore, the FAA is proposing a rule that would mandate the use of simulators for all air carrier flight crew training and qualification, limiting the use of the aircraft itself as a training option even for small regional airlines. However, there is a lack of sound scientific data on the relationship between certain key training device features, such as platform motion cuing, and their effect on the transfer of performance to and from the airplane. The goal of this project is to provide a scientific basis to ensure that FAA requirements are commensurate with safety objectives. Particularly, it addresses the question of the need for simulator motion for commuter airline pilot recurrent training and evaluation in the presence of a state-of-the-art visual system. The resulting data will also help the FAA to evaluate air carrier proposals for the alternative use of full flight simulators, whose availability and affordability may be limited especially for small regional airlines, and other training equipment.

In the past, technical constraints naturally limited the level of fidelity of a simulation. Today, however, technical capabilities have expanded to a point where they may enable a degree of fidelity that may exceed the one required for a particular purpose. This may lead to a situation where the benefit resulting from increased fidelity no longer justifies its cost. The focus thus needs to shift from ever more sophisticated technology to the level of fidelity required to train and evaluate to a specific safety standard. One of the common misconceptions is that the higher the physical fidelity of the simulation the better the training will be. The level of fidelity of the simulator should be determined, however, by the level needed to support the learning or evaluation of the tasks that will be trained or evaluated using the device. Any other approach would unnecessarily preclude the less affluent sectors of the aviation community from the benefits of using simulators.

Project Description

This project consists of four parts: (1) the collection of subject matter expert opinion, (2) a literature review, (3) empirical research, and (4) validation and generalization of the results of the research. For the first part, two workshops were conducted with experts from the FAA, industry, and academia. The first workshop focused on the aeromodel validation standards used in the flight simulators and the second focused on the motion requirements for simulators. These workshops led to proposed changes to the regulations for Level B simulators (the minimum level of simulators required for recurrent training).

The second part of this project is an extensive literature review. The relevant literature has been organized into a large electronic database, which is constantly updated to incorporate new information. Currently, the database contains 450 references, most of them summarized and annotated.

The third part is the execution of research addressing the need for simulator motion. The research empirically examined the effect of FAA qualified Level C six degree-of-freedom synergistic motion in the presence of a wide-angle high quality visual system on pilot training and evaluation. Transfer of skills acquired in the simulator to the airplane was measured by comparing the effect of training received in the simulator, with and without motion, on performance and behavior in the simulator with motion. This “quasi-transfer” to the simulator with motion as a stand-in for the airplane ensured the safety of the participants while allowing full experimental control. The research was conducted using regional airline pilots in recurrent training. Every effort was made to avoid deficiencies in the research design identified in a review of prior studies, by measuring pilot stimulation *and* response, testing both maneuvers and pilots that are *diagnostic* of a need of motion, avoiding pilot and instructor *bias*, and ensuring sufficient statistical *power* to capture operationally relevant effects. Two test maneuvers were chosen as diagnostic for an effect of motion on pilot training and evaluation, namely, engine failures on take-off with either rejected take-off (RTO) or continued take-off (V_1 cut). These maneuvers minimally disrupted the host airline's training program while satisfying the criteria recommended in the literature as diagnostic for detection of a motion requirement. These criteria included 1) closed loop, to allow for motion to be part of the control feedback loop to the pilot; 2) unpredictable and asymmetric disturbance, to highlight an early alerting function of motion; 3) high gain and high thrust, to magnify any motion effects; 4) high workload, to increase the need for redundant cues such as provided by motion, out-the-window view, instruments and sound; and 5) short duration, to prevent pilots from adjusting to a lack of cues. A vast amount of data in the form of instructor grades, instructor and crew opinions, and objective measurements was collected from the experiment and analyzed. This research has been completed and the results are available in two different formats (see below). The results indicate that the motion provided by the test simulator, which may or may not be typical of other FAA qualified Level C flight simulators, does not, in an operationally significant way, affect evaluation, training progress, or transfer of training acquired in the simulator with or without motion to the simulator with motion for the maneuvers and pilots tested. Two *caveats* have to be kept in mind, however. First, the current study used the simulator with motion as a stand-in for the airplane. However, because the crews trained with motion did not show an advantage over the crews trained without motion when transferring to the simulator with the same configuration, it is unlikely that they would have had a greater advantage transferring to the airplane. The second caveat is that the simulator used in this study may not have provided sufficient motion to be effective. The measurements indicate that although the roll and longitudinal accelerations produced by the motion system of the simulator followed the aircraft response fairly well considering the limitations inherent to all simulators, the lateral acceleration seemed to be lacking for the maneuvers tested (RTO and V_1 cut).

The fourth part of the project is the validation and generalization of the results obtained from the above research. This should include a comparison of the objective measures from the motion system used in the experiment with the same measures taken from other FAA qualified Level C simulators to determine whether or not the motion used in the present study is representative. The data collection for this purpose is currently undergoing and should be completed soon. This should be followed by an investigation of whether operationally relevant effects of motion would be found with a simulator where the motion is manipulated to assure that it is representative of the airplane for the maneuvers selected. Additional maneuvers that may be diagnostic and a different pilot population should be tested as well. Ideally, some validation of the quasi-transfer design with a real airplane would also be undertaken.

Products:

Papers:

1. Transcript of the Joint FAA/Industry Symposium on Level B Airplane Simulator Aeromodel Validation Requirements, Washington Dulles Airport Hilton, March 13 - 14, 1996.
2. Transcript of the Joint FAA/Industry Symposium on Level B Airplane Simulator Motion Requirements, Washington Dulles Airport Hilton, June 19 - 20, 1996.
3. Proposed Changes to AC 120-40B submitted to FAA Simulator Program Management.
4. Longridge, T., Ray, P., Boothe, E., Bürki-Cohen, J. (1996). Initiative Towards More Affordable Flight Simulators for U.S. Commuter Airline Training. In *Proceedings of the Royal Aeronautical Society Conference on Training - Lowering the Cost, Maintaining the Fidelity, 15-16 May 1996, London, UK,* 2.1-2.17.

5. Bürki-Cohen, J., Soja, N. N., Longridge, T. (1998). Simulator Fidelity Requirements: The Case of Platform Motion. *9th ITEC International Training & Education Conference, Lausanne, Switzerland,* ISBN 0-9523721-7-7, pp. 216-231.
6. Bürki-Cohen, J., Soja, N. N., Longridge, T. (1998). Simulator Platform Motion-The Need Revisited. *International Journal of Aviation Psychology,* **8 (3)**, 293-317.
7. Bürki-Cohen, J., Soja, N.N., Go, T.H., Boothe, E.M., DiSario, R., Jo, Y.J.: Simulator Fidelity: The Effect of Platform Motion. Report No. DOT/FAA/RD-00/XX, May 2000.
8. Bürki-Cohen, J., Boothe, E.M., Soja N.N., DiSario, R., Go, T., Longridge T.: Simulator Fidelity - The Effect of Platform Motion. In Proceedings of the *International Conference Flight Simulation--The Next Decade, Royal Aeronautical Society*, 10-12 May 2000, London, UK.
9. Go, T.H., Bürki-Cohen, J., DiSario, R.M., Jo, Y.J.: Relationship Between Objective Measures of Pilot Performance/Behavior and Instructor Grades. Report No. DOT/FAA/RD-00/XX, June 2000.
10. Go, T.H., Bürki-Cohen, J., Soja, N.N.: The Effect of Simulator Motion on Pilot Training and Evaluation. Paper accepted for publication in proceedings of AIAA Modeling and Simulation Technologies Conference, Denver, August 2000, in preparation.

Presentations FY 99/00 that are not published in proceedings:

1. Presented preliminary results of investigation of *Simulator Fidelity Requirements: The Case of Platform Motion* to FAA AAR-100 Air Carrier Training Research Review (October 1998).
2. Discussed investigation of flight simulator motion requirements with AIAA Simulation and Modeling Committee at NASA Ames: *Simulator Fidelity: The Effect of Platform Motion* (April 1999).
3. Presented *Relationship Between Grades and Objective Data* to help flight operational quality assessment (FOQA) program with data evaluation at FAA AAR-100 Air Carrier Training Research Review (July 1999).
4. Presented final results and conclusion on *Simulator Fidelity: The Effect of Platform Motion* to FAA sponsors, customers, participating airline, and training center (October 1999).
5. Presented *Simulator Validation: The Case of Platform Motion* at Human Factors in Transportation 33rd Annual Workshop at Transportation Research Board 79th Annual Meeting (January 2000).

Project Title
Realistic Radiocommunications Simulation (RRS) in Airline Pilot Training and Evaluation

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

Background

Radiocommunications are an integral part of every flight and require not only procedural knowledge, but also increase the need for sound task management, situation assessment, and decision-making skills. In simulator training, however, the incorporation of air traffic control (ATC) and company communications is often left to an already overburdened instructor, who must attend to many other administrative, training, and evaluation tasks. The lack of realistic radiocommunications simulations (RRS) impoverishes the simulation environment and compromises the transfer of skills between simulator and airplane. Full skill transfer to and from the airplane is a critical issue, if simulator use for training and evaluation is to be mandated.

Project Description

In FY99 and part of FY00, Volpe and NASA Ames collaborated on a feasibility study for the development and implementation of realistic radiocommunications (ATC and carrier) in line operational simulations (LOS). The feasibility study consisted of three parts: 1) literature review related to the requirement for RRS, 2) field survey of instructors and evaluators on current radiocommunications simulation practices, and 3) review of emerging technologies supporting the use of RRS. The feasibility study report has been completed and submitted for review.

A summary of results follows. The literature provides both practical and theoretical support for the necessity of realistic radiocommunications to achieve a variety of training objectives; for example, situation awareness, crew resource and task management, as well as pilot/ATC communication and coordination. Realistic radiocommunications are very important to ensure that training in a simulator will generalize and transfer to the complexity of the real world. A realistic radiocommunications environment avoids the negative training acquired in an impoverished environment leading to tunnel vision.

The accident and incident literature shows that radiocommunications are a major contributor to pilot workload and are a causal factor in cockpit confusion, task management and monitoring errors, and flight path deviations. Inadequate ATC service has been identified by the Flight Safety Foundation Approach and Landing Accident Reduction Task Force as a factor in a third of all approach and landing accidents and incidents (Khatwa and Helmreich, Flight Safety Digest, November 1999). Crews need to be trained to recognize and challenge overly demanding ATC instructions such as last minute runway changes which may have contributed to the American Airlines Boeing 757 accident in Cali, Columbia, in December 1995. In general, crews need to practice effective communication with ATC, particularly during emergencies and when dealing with non-native English speaking ATC (or if they are non-native English speakers themselves).

The importance of realistic radiocommunications during full mission simulations has been clearly recognized by the Federal Aviation Administration Advisory Circulars related to Advanced Qualification Program and crew resource management (CRM), which underscore that coordination with air traffic control and company is an integral part of line operations and that frequency monitoring is important for maintaining traffic and weather situation awareness. The CRM process includes not only on-board personnel (flight and cabin crew), but also air traffic controllers and company dispatcher and maintenance personnel.

These opinions were widely shared by the 29 Instructor/Evaluators (I/Es) from 14 airlines participating in the survey of I/E opinion and airline current practices. I/Es consistently rated the presence of air traffic control communications to own aircraft in simulations as important, especially in the terminal environment. Company communications were given above medium importance ratings as well, but slightly lower than air traffic control. The highest importance for company communications was given to communications from cabin personnel and maintenance. I/Es indicated that realistic radiocommunications increase the effectiveness of training such skills as CRM, and dealing with new ATC procedures or non-routine ATC situations. Radiocommunications are also important to train and evaluate distraction management and situational awareness skills. For the latter, simulating communications between ATC and other aircraft, the so called "partyline," was considered especially important. One I/E commented that the "partyline" was his "biggest concern, so that pilots are listening." Another stated that the "partyline enables CRM elements such as workload and distraction to be assessed more effectively."

Several I/Es commented that radiocommunications are a "critical need for LOS realism." "There is a correlation between realism in communications and training effectiveness; better training results from more realism." A final comment from an I/E was that "[t]his is a subject near and dear to my heart. […] I believe the 'simulator mindset syndrome' must be fought with realism. How can we expect crews to 'treat the sim[ulator] like the aircraft' when the audio environment belies the condition so often?"

Despite the agreement between literature, advisory circulars, and I/E opinions, the review of current practices showed that radiocommunications are almost exclusively role-played by the already burdened I/E. This increases I/E workload and detracts from the I/Es main function of instructing and observing, "divid[ing] his attention." I/Es reported that they spend about one quarter of their time and effort role-playing radiocommunications, regardless of whether the simulation is for training or evaluation purposes. This is about the same amount of effort and time as they spend running the simulator.

At the same time, pilot workload is diminished, because the complexity of the radiocommunications provided is "all relative to the creative ability of the instructor," which additionally may lead to uneven training and evaluation across crews. "I/E [radiocommunication] is less than actual, therefore it reduces pilot workload" and "pilots just listen when the instructor keys the mike." One I/E commented that "company comm[unication]s [are] not normally used, too time consuming." "Partyline" communications are "very seldom used due to I/E workload" and because "none of our formal training documentation requires it."

Despite the best intentions of I/Es, then, "when the pilot trainee finally arrives in the 'real world,' he must add another component, which was not learned during training. This new component can really complicate line flying." This last quote of one of the survey participants shows that I/Es need support in providing realistic radiocommunications to remedy this situation. The question is how. One possibility is increased training of I/Es to provide more realistic radiocommunications. Given that the awareness of I/Es regarding the importance of providing realistic radiocommunications appears high, however, it is questionable that additional I/E training alone would be beneficial. A second option would be to use additional instructors to distribute the workload of realistically providing radiocommunications. In fact, three I/Es from one airline indicated that two instructors may share the communications workload occasionally, but only when instructing a three-person crew. In the literature, there has been mention of other airlines using multiple instructors to simulate ATC, company, and cabin personnel (Lyall, Vint, Niemczyk, Wilson and Funk, December 1998, AAR-100). It appears, however, that personnel costs have prevented this from becoming a widespread practice.

Another option is to make use of technology to fully or partially automate the provision of ATC, company, and “partyline” communications. The most complete such technology identified in the review is the United Airlines in-house development called Interactive Real Time Audio System or IRAS. It automates both communications to own aircraft and the “partyline” for all flight phases. In response to a customer requirement that visual terminal environment traffic also be represented by audio, the simulator manufacturer CAE developed its Ground and Air Traffic Environment System (GATES). GATES, however, is limited to the terminal environment and communications between ATC and other aircraft. It doesn’t provide any communications to own aircraft. Commercial off-the-shelf, more advanced as well as supporting technologies such as the Rapidly Reconfigurable Event-Set Based LOE Scenario Generator (RRLOE, see this report) and speech generation and recognition systems were also examined.

Any solution, however, be it the low-tech solution of additional training or I/Es or investing in existing or future technologies, or using a combination, will come at a cost. Before airlines invest in RRS, they will require objective data on the cost/benefit trade-offs achieved by different levels and methods of simulating realistic radiocommunications for different training and evaluation purposes. Phase 1 of the RRS project is now completed. Follow-on activities will focus on the research needs identified above.

Products

Bürki-Cohen, J., Kendra, A.J., Kanki, B.G., Lee, A.T.: Realistic Radiocommunications in Pilot Simulator Training. Report No. DOT/FAA/RD-00/XX, Final Draft March 2000.

Project Title
Rapidly Reconfigurable Line Operational Evaluation (RRLOE) Generator

Performing Agency
FAA/UCF/NAWCTSD Partnership for Aviation Team Training Research at the University of Central Florida, Orlando, Florida

Background

Line-Oriented Evaluation (LOE) is an evaluation methodology used in the Advanced Qualification Program (AQP) to assess trainee performance in scenarios that are designed to challenge the integration of both technical and crew resource management skills. They are employed in AQP for jeopardy grading purposes - i.e. LOE failures are formally reported as a matter of record to the FAA, and pilots who fail an LOE may not be returned to duty until deficiencies are remediated, and performance on a subsequent LOE is determined to be satisfactory. Ratings of pilot performance on all LOEs must be electronically submitted to the FAA AQP office.

LOEs must be approved by the local FAA office responsible for oversight of a given air carrier. The approval procedure preceding the development of the RRLOE methodology has entailed an FAA review of each such gate-to-gate LOE scenario. That has required each LOE to be separately conceived, developed, and tested by the training organization, and then to be individually reviewed and approved by the FAA. Thus, the development of LOEs has been both costly and time consuming. As a result, training organizations typically have had only a limited number of LOEs available for evaluation, each of which was approved for only a limited time period. Repeated use of a small number of fixed LOE scenarios creates the potential for the LOE events to become known in advance by the trainees in the organization, thus reducing the validity of both individual trainee evaluation and fleet proficiency assurance. The FAA is concerned about this potential deficiency as it could impact the overall effectiveness of the program.

One way to improve the validity of the LOE methodology would be to develop a process that avoids the use of a small set of fixed scenarios over an extended period of training within an organization. If events which form the building blocks of a full LOE could be individually developed and approved by the FAA, then these events could be used as a database from which to assemble complete, unique LOEs. The training organization could rapidly build new LOE scenarios with desired events within them, without seeking FAA approval of each complete scenario. The specific content could be varied while controlling the general content and overall difficulty. This would lead to fair evaluations for all trainees on fresh, variable, and valid training scenarios. In addition, to the extent that a methodology could be developed for rapidly generating alternative events that target skills sets on which deficiencies in pilot performance are observed in the initial administration of an LOE, FAA goals for improvements in the remediation strategy for retraining and rechecking pilot LOE failures in AQP could be achieved.

Project Description

The development of the RRLOE methodology is the focus of the current research project. In the following sections, the progress made between February 1999 and March 2000 is reported.

Realism. A first challenge in developing this type of methodology and associated tools has been to conduct a program of research to identify those aspects of LOE scenarios that are required to make LOEs a valid and realistic assessment situation. The RRLOE research project has approached this problem by integrating the results of modern knowledge elicitation techniques with innovative database development. The result of this process was a theoretical approach (i.e., the "Domino-Method," see Bowers, Jentsch, Baker, Prince, & Salas, 1997) that became the basis for the existing expert system that tracks realism and continuity in the RRLOE program. This novel approach allows the addition of individual event sets to a library of event sets for LOEs without the need to check the compatibility with other event sets manually. In fact, the resulting system can track more parameters relevant to a flight than a human LOE developer could in the past.

A second issue pursuant to realism is the impact of weather in LOEs. Recent research has demonstrated that weather is a prime consideration in aircrew decision making (e.g., Jentsch, Irvin, & Bowers, 1997). However, weather has traditionally been only an afterthought in LOE development. Consequently, weather reports given to pilots in LOEs often were developed on the basis of small changes to existing weather paperwork and did not capture the complexities of the operational environment. To remedy this situation, the RRLOE research team has conducted extensive research into the variables that affect pilot decision making and have generated both a mechanism and a sample library of advanced weather scenarios which systematically vary those critical elements. It should be noted that this is the first time that theoretical research on the effects of weather on aviator decision making has been made available to aviation training organizations in the form of a practical tool.

Difficulty. Another major research thrust of the RRLOE project relates to the estimation and measurement of LOE difficulty. Every LOE, notwithstanding whether it is generated by a human or a computer, must fall within an acceptable range of difficulty. A human LOE developer achieves this task largely through an experiential process that takes the combination of event sets, environmental conditions, and assessment expectations into account. Therefore, the challenge for the RRLOE project was to identify and describe the process used by humans and translate it into a mathematical model that the machine could execute. It was also important to validate the consequent model. The RRLOE research project has accomplished this through several studies at air carriers and has presented the results at international meetings (cf. Jentsch, Abbott, & Bowers, 1999). The research team now feels confident that the proposed methodology leads to a range of equivalent and fair scenarios. Again, this is the first time that a method for LOE difficulty assessment has been developed, validated, and described for the operational community.

Human factors aspects of the software. Finally, the results of the research and development effort described above needed to be translated into a user-friendly tool set, so as to facilitate transfer of the research results to the operational arena. In developing this tool set, the research team has had to study and apply human factors principles related to airline culture, operational environment, and computer sophistication across a wide range of operators participating in AQP. The resulting software tool set includes both tools for operators that are just beginning the AQP process and those which are far advanced within it. Finally, following the skill-based training approach described above, the project had to integrate existing task and skill analyses with the goals and tools of RRLOE. This required the consequent application of HF guidelines in Computer-Human Interface (CHI) design to their program. Further, through research at various airlines, the RRLOE research team was able to establish a standardized way to describe event sets that can be used by all participants in the AQP process.

-

•

Extended Work

The following research needs were identified for follow-on development of the RRLOE software:

Automation. The proliferation of high-technology systems in aviation has important implications impending on instruction. Specifically, future training for high-technology aircraft will not only require the identification of new knowledge and skills, but also the study of how the interaction with automated system impacts the training and maintenance of traditional pilot skills. Furthermore, there is a need to create training opportunities and materials that take the technical capabilities of simulators and training devices into account. Both the research methodologies and the software tools created under the RRLOE effort can now be employed to significantly advance the development of simulation scenarios for automation training. Indeed, it would appear that the only means of assuring that the decision making skills required for the selection of an appropriate level of automation for a given set of flight conditions is through the systematic development of scenarios that specifically target such skills. The RRLOE framework developed to date has a high potential for improvements in the training and evaluation of such skills, but it requires further RE&D refinement if that potential is to be fully realized.

Measurement. As described above, valid measurement of aircrew skills is the cornerstone of the Advanced Qualification Program. The development of tools to accommodate consideration of measurement requirements during the LOE scenario development process itself could significantly enhance the precision of scenario based pilot assessment in the final product. While the groundwork for such an approach has been established by virtue of the RRLOE work accomplished to date, further RE&D to explicitly embed tools validated for that purpose is needed. In addition, although the current RRLOE algorithms can reliably predict perceptions of scenario difficulty among expert pilots and instructors, it is not yet clear whether either of these sources can reliably predict actual performance of trainees in the simulator. It is therefore important to test the estimated difficulty of LOE scenarios against actual performance. It is in this area where the application and extension of the RRLOE tool set could realize important benefits regarding the time and cost needed to create research scenarios.

References

Bowers, C., Jentsch, F., Baker, D., Prince, C., & Salas, E. (1997). Rapidly reconfigurable event-set based line operational evaluation scenarios. Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting, Albuquerque, NM (pp. 912-915). Santa Monica, CA: Human Factors and Ergonomics Society.

Jentsch, F., Abbott, D., & Bowers, C. (1999). Do three easy tasks make one difficult one? Studying the perceived difficulty of simulation scenarios. Proceedings of the Tenth International Symposium on Aviation Psychology. Columbus: The Ohio State University.

Jentsch, F., Irvin, J., & Bowers, C. (1997). Differences in situation assessment between experts and prospective first officers. Proceedings of the Ninth International Symposium on Aviation Psychology (pp. 1228-1232). Columbus: The Ohio State University.

Products

Research Reports Describing Difficulty Research
Methodology for Generating, Cataloguing, and Assembling LOS Event Sets
Event Set Library
RRLOE Software and Documentation
RRLOE Training Course

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 and Proceduralization of Pilot Skills for Automated Flight Decks

Performing Agency
George Mason University

Background

The potential relationship between crew resource management (CRM) and flight deck automation management has been evident for some time. A number of carriers have defined specific automation skills as well as CRM skills, but the research community has yet to establish a common framework that can be used to coordinate research in these two related areas. This research project is divided into two separate but related research efforts. Part 1 extends the use of Advanced Qualification Program (AQP) databases from previous research to include automation topics. Part 2 is a detailed analysis of cognitive skills and procedures related to specific crew performance problems in automated flight decks. These two areas overlap where the information in AQP databases is used to assess crew automation performance. These two efforts will be pursued concurrently.

Project Description

This three-year effort is designed to provide the framework to integrate and extend what was learned in our previous grant work about proceduralized CRM to the detailed procedures and cognitive skills relevant to flight deck automation. Establishing this framework provides potential common ground for research on CRM and team skills to manage automation. More importantly for the air carriers, this framework should also provide a path for applying the relevant research results to the development of better assessment of crew automation use and new operational procedures or crew training for automation use.

Analysis of pilot cognition and teamwork in the context of automated flight deck systems that may have a substantial effect on crew performance are under way. The George Mason University team has established a data collection and analysis foundation at a major carrier, and the team is using that base to develop and evaluate ways of assessing automation performance. A unique combination of cognitive modeling and performance assessment to help carriers more precisely evaluate and change automation procedures and training strategies is being used. In addition, a cognitive model of automation use for several focal automation problems is being developed. This model will be used to develop assessment items and procedures for LOFT/LOE evaluations of automation use that will be evaluated for scientific reliability and validity.

Part 1: Database

To date, some Performance Proficiency Database (PPDB) information for the 1999 LOE has been received. Path analyses have shown that there is significant structural validity for the evaluation process. The observable behaviors and task skills predict the Technical and CLR ratings and these in turn predict the PIC and SIC ratings. Compared to previous data analyses for the proceduralized CRM project, the strength of these relationships are slightly stronger.

In addition to using the PPDB to establish baseline reliability and validity, it has been used to study specific automation performance indicators in the PPDB data; however, that attempt has not been fruitful to date. Qualitative data analysis included assembling and coding automation-related performance comments into a taxonomy of errors. Consensus on coding was difficult, and the results did not seem to illuminate basic automation performance issues.

Quantitative data analysis has focused on the carrier PPDB LOE performance ratings for automation-relevant compared to non-automation event sets. Technical performance was found to be significantly different in the automation-related event sets. However, the CLR, PIC, and SIC ratings as well as specific observed behaviors and task skills were not significantly different in automation vs. non-automation event sets. Therefore, the evidence for the scientific quality of current assessments of automation performance is not compelling. The sensitivity, reliability, and validity of these assessments may be increased by a more detailed analysis of pilot interactions with automation.

Part 2: Cognitive Modeling of Automation Use

The most common problems in automated flight reported in research and by trainers/evaluators are concerned with vertical navigation. Specifically, there seems to be a lack of understanding of the relationships among vertical speed mode (V/S), flight level change mode (FLCH), and the VNAV mode. Thus, we have begun to build a cognitive model focusing on the issues surrounding changes in vertical path during the climb and descent phases of flight.

In collaboration with major carriers, event sets will be designed that emphasize vertical navigation. These event sets are being used to model the interaction of pilots with cockpit automation in climb and descent. A desktop Boeing 747-400 flight simulator was used to gather verbal protocol and eye-tracking data for the cognitive modeling from pilot subjects.

The data from this study have been used to collect basic information about how pilots track data in an automated cockpit. Specifically, for example, the data suggest that pilots coming from a major airline do not use a common procedure to execute a descent using automated systems. These data are being used to complete the first version of the cognitive model. The cognitive model will also be coupled with the simulator to test alternative strategies for vertical navigation. The use of the model-simulator package uncovered a strategy that works well in the operational aviation environment. This package should integrate procedural and conceptual elements and guide better approaches to assessment of crew automation performance, development of automation procedures, or crew automation training.

Integrating the Database and Modeling efforts

These efforts will lead to measurement and analysis of pilot performance data associated with flight deck automation in a high-fidelity simulator located at a major airline. This analysis will provide a baseline for current flight deck automation performance with current carrier training and provide information on critical measures or indexes of crew automation performance. The critical behaviors and processes identified by the model will be used to develop more precise and detailed measures of crew automation performance for the airline. The reliability and validity of these revised measures will be evaluated as the final tasks in this grant.

This baseline can also be used by the carrier to compare the performance of pilots with revised or updated flight deck automation training against pilots without this new training. Similarly, the development of new procedures for crew automation use can be based on the problems identified by the cognitive model. These new procedures would be developed in a similar manner with the proceduralized CRM intervention developed and successfully evaluated under a previous FAA grant. Future work may evaluate the effects of either training interventions or automation procedures based on this model.

Products

- ***Training guidelines***

- ***Procedures development guidelines***

Project Title
Automation Skills and Training

Performing Agency

FAA/UCF/NAWCTSD Partnership for Aviation Team Training Research at the University of Central Florida, Orlando, Florida

Background

Recent advances in aircraft instrumentation and equipment have lead to a new generation of advanced automated aircraft. The development of electronic flight management systems (FMSs), in particular, has given flight crews very powerful tools to plan, execute, and control specific flight plans by automating navigation and performance calculations. Indeed, the latest generation of FMSs with four-dimensional (4-D) guidance (i.e., along the three spatial axes, and in time) can conceivably conduct an entire automated flight from engine spool-up at takeoff to deceleration after landing.

While these technical advances have undoubtedly increased the capabilities of modern aircraft, they have also led to a change in the pilot's job: Previously, the pilot was largely "in-the-loop," that is, an active controller who received direct and immediate feedback from the aircraft about the effects of his/her control inputs. With today's modern FMS systems, however, the pilot's role has changed from an active controller to a more passive supervisor/monitor/administrator. System inputs made early in the flight, such as route entries, may have adverse consequences much later in the flight. Also, as with other computerized systems, the capabilities of the system are not always entirely transparent to the operator. In fact, recent studies have shown that pilots frequently are uncertain about a system's status and its future performance (Sarter and Woods, 1998): Questions such as "What is this thing doing now?," and "What is it going to do next?" are, unfortunately, frequently heard in today's modern flight decks.

Sometimes, such uncertainty can have deadly consequences: In the decade since the introduction of the second generation of automated aircraft (Airbus A320, Boeing B757/767), several fatal accidents have occurred in which flight crew uncertainty/error in the operation of advanced autoflight systems has been cited as a causal or contributing factor (e.g., B757 at Cali, Colombia; A320 at Strasbourg) (Billings, 1997).

Project Description

Given the complexity of modern automated aircraft systems and the potential threats that inappropriate flight crew understanding of the systems pose, training for flight crews has become even more important in these aircraft than in traditional aircraft. This has been recognized by the FAA and has been the motivating force behind this research project. A structured analysis of air carrier needs with respect to automation training confirmed earlier FAA (Abbott et al., 1996) and industry reviews which showed that the identification of the skills needed to operate automated aircraft successfully is one of the key issues that need to be studied. The research also responds to the need for better airline-level tools by developing methods for scenario generation and performance measurement that specifically target cognitive skills for automated aircraft in the air carrier environment.

This research is dedicated to the appropriate assessment of automation skills. The eventual goal of this automation research is to create a 'toolbox" for use by training departments in assessing whether trainees truly understand the automated system. The toolbox is comprised of several different measurement products, each designed to tap a critical aspect of automation ability.
Obviously, an important aspect of autoflight skill is knowledge about the system and how it works. Thus, one part of the toolbox will be a set of checklists, quick assessments, and other tools to measure this knowledge.

However, it is also important that operators organize their knowledge of autoflight systems such that it supports accurate inferences. Toward this end, we will provide guidelines on how to use existing knowledge structure assessments for the problem of automated flight. We have also initiated the development of new tools to measure these knowledge structures in one specific flight system in an operational fleet. We will continue our validation of this approach and will include it in the toolbox.

In addition to knowing the facts about automate systems, it is important that the operator be able to execute a complex sequence of activities to bring about desired outcomes. Operators usually acquire this "procedural knowledge" through practicing on low-fidelity trainers. However, it is difficult to capture performance on these trainers so that effective feedback can be provided to facilitate learning. In response to this problem, the UCF team is developing a software tool that will capture trainee inputs. By collecting these data, we will be able to compare the trainee's performance to pre-determined sequences so that we can provide rapid feedback about procedural knowledge deficiencies. We will also provide guidance on how to transfer this technology to other low fidelity trainers.

After acquiring basic knowledge of automated systems and their operation, trainees must be able to use these skills in realistic flight contexts. Consequently, a final aspect of the autoflight skill measurement toolbox includes modifications to the RRLOE that will allow the generation of useful training scenarios. The scenarios will be used to determine whether the trainee can use the automation effectively in specific situations that have been shown to be troublesome to novices.

Overall, the toolbox offers substantial advantages over current assessment methods. It allows the assessment of knowledge through several levels of acquisition, it considers both learning and application, and it promotes assessment of autoflight skills in the broader operational context. Furthermore, it offers a range of data that can be used to diagnose and remediate shortcomings in automation training programs.

- .
-

Products

As of March 2000, the project has resulted in the following products/anticipated products.

Presentations at Professional Meetings

Hitt II, J.M., & Jentsch, F. (1999). Differences in general and commercial aviation automation. Paper presentation at the 18th Digital Avionics Systems Conference, Oct. 23-29, 1999, St. Louis, MO.

Hitt II, J.M., Jentsch, F., Bowers, C., & Oser (under review). An information-processing model approach to identifying training issues for advanced aviation automation. Paper proposal for the Human Performance, Situation Awareness, and Automation: User-centered Design for the New Millenium, October 15-19, 2000, Savannah, GA.

Jentsch, F., Hitt II, J.M., Agliata, D., Milham, L., Evans, B., & Bowers, C. (under review). Theoretical limits of card-sorting as a mental models assessment tool. Paper proposal for the Human Performance, Situation Awareness, and Automation: User-centered Design for the New Millenium, October 15-19, 2000, Savannah, GA.

Jentsch, F., Hitt II, J.M., Bowers, C., & Salas, E. (Accepted, 2000). Do pilots know how they manually control the vertical flight path? Studying conceptual knowledge of vertical flight path guidance. Paper accepted for presentation at the APA Annual Meeting, 2000.

Milham, L., Jentsch, F., & Bowers, C. (Accepted, 2000). Understanding understanding: Comparisons between assessment methods of knowledge structures. Paper accepted for presentation at the APA Annual Meeting, 2000.

Technical Reports and Project Memoranda

Hitt II, J.M. (1999). Reference list/bibliography for automation-related publications. Internal Team Performance Laboratory Technical Report compiled for the FAA Automation Training Team. Can be found at http://www.researchintegrations.com/ (password protected). Orlando: University of Central Florida.

Jentsch, F. (2000). Aviation Safety Reporting System (ASRS) - Air Carrier Report Search: Narratives for the 357 reports mentioning "VNAV". Technical Memorandum for the FAA Air Carrier Automation Skills and Measures Team. Orlando: University of Central Florida.

Jentsch, F., Hitt II, J.M., Bowers, C., & Oser, R. (in preparation). Identifying training areas for advanced automated aircraft: Application of an information-processing model. Interim Technical and Scientific Report to the Naval Air Warfare Center Training Systems Division and the Federal Aviation Administration. Orlando: University of Central Florida.

Jentsch, F., Hitt II, J.M., Bowers, C., Oser, R., & Salas, E. (1999). Do pilots know how they manually control the vertical flight path? Studying conceptual knowledge structures of vertical flight path guidance. Research White Paper for the FAA Automation Research Team. Orlando: University of Central Florida.

Airline-level Products:

- VNAV white paper and assessment vignettes
- Automation knowledge assessment questionnaire
- Automation question catalogue
- Draft of a technical memorandum for the creation of automation vignettes and event sets.
    - Prototype software for data collection in PC-based flight management systems trainers

Project Title
Human Factors, Human Error and System Safety.
99-G-004, Robert L. Helmreich, PI.

Performing Agency
The University of Texas, Austin

Background

Current research extends investigation of the human factors of flight operations previously produced by this research group. The original research focused on Crew Resource Management (CRM) training as a method of optimizing crew performance and avoiding and managing operational threats and human error. CRM training focuses on issues which have been implicated as causal in the majority of air crashes and incidents. These investigations have demonstrated that CRM training is effective in changing flightcrew attitudes and behavior, and by implication, system safety.

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 regions 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 in managing the threats that occur in daily flying. This is done by cockpit crew deployment of strategies to manage threat and error, and mitigate the consequences of error when it inevitably occurs. In this context, CRM represents a training architecture for developing a set of tools to be used as countermeasures and error management tools. 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.

A focus on system issues (latent system threats) and using CRM as a training intervention for threat and 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. In other words, proceduralization may have the unintended consequence of increasing violations.

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 flight operations. These areas are:

(1) Developing and refining a new conceptual model of threat and human error in flight operations. The model will be of practical value in understanding crew performance on the line, in developing training programs, and in the analysis of incidents and accidents. The model is driven by data collected from pilots flying regularly scheduled airline flights.

(2) Extending current data on performance in line operations to include data on environmental threats to safety, error, and responses to error. This approach (called LOSA – Line Operations Safety Audit) employs a team of expert observers from the jumpseats of scheduled flights to collect systematic data on line operations ***without jeopardy.*** The research team collects common data in each organization giving evaluation points that allow comparisons of performance across organizations. Organizations have used the findings to target specific issues to address in flight operations and during training. The specific data collected are guided by the conceptual model. LOSA has been endorsed by the International Civil Aviation Organization (ICAO) as a preferred means of assessing system safety.

(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 and documentation of the automation anomalies and threats that occur in normal line flying.

(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 determination of trends and latent systemic threats. With the recent approval of the ASAP program, a unique opportunity for studying incidents has presented itself. The conceptual threat and error management model will aid in understanding the dynamics of causes cockpit incidents – and of their effective and ineffective management.

(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 U.S. major and many regional airlines to assess CRM curriculum needs and safety concerns. It has also been used by airlines in more than 20 foreign countries. This task will produce a semi-automated FMAQ self-administration package that airlines can use for data entry and reporting. This package contains norms from both US and international airlines.

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. This role helps disseminate research findings.
- 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: and (d) demonstrate how the Model AQP methodology facilitates the development of integrated CRM and technical training by creating leadership/followership training using the methodology.

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.

Development of the Advanced Model AQP has been 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 AQPhas started. 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. The final product will be delivered to industry at the end of 2000. In addition, work has begun on the development of leadership/followership training within the Model AQP framework. This work involves a number of activities, including:

- Developing an approach to defining and conceptualizing leadership/followership that shifts the focus from individuals working within a team environment to teams organizing themselves to effectively manage flight situations.
- Translating leadership/followership concepts into the AQP format of tasks, knowledge and skills, training objectives, and events.
- Preparing a leadership/followership curriculum and footprint that demonstrates the use of technical issues to create situations that afford opportunities to acquire and strengthen leadership/followership skills within a crew environment.

Products

- Model AQP and research report delineating process, methodology and lessons learned.
- Advanced Model AQP.
- Leadership/followership training guidance provided in a form that carriers can introduce into existing AQPs (tasks, knowledge/skills, training objectives, events, and syllabus).

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

Performing Agency
Key Dismukes, Ph.D., NASA-Ames Research Center

Background

Interruptions, distractions, and preoccupation with one task to the neglect of another task are among the most common causes of pilot error incidents and have contributed to many accidents. 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 is a collaborative project, co-sponsored with the NASA Aviation Safety Program. It 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 Procedures (SOP) to reduce this form of crew error.

To date this project has analyzed ASRS incident data and NTSB reports of accidents attributed primarily to crew error. In roughly half of the accident reports lapses of attention by the crew occurred. The review of ASRS data revealed that the most commonly neglected tasks involved lapses in monitoring (2/3 of reports) or lapses in memory-especially failure to remember to complete deferred actions (1/4 of reports). The competing tasks that distracted or preoccupied the crews were: communication (50%), "head-down" tasks (16%), abnormals (14%), searching for traffic (8%), and miscellaneous (12%).

Two questionnaire studies were developed and administered to senior captains, instructors, and check pilots. The first questionnaire elicited information of what techniques experienced pilots currently use to reduce their vulnerability to lapses in attention and memory. The second questionnaire probed the extent to which experienced pilots agree on the efficacy of these techniques and the practicality of using them in line operations. Preliminary data analysis suggests that some techniques are currently available, however, experienced pilots differ substantially in their opinions of the efficacy and practicality of these techniques.

Analysis of cognitive issues suggests that these lapses of attention and memory are not primarily the result of overload but of difficulty in switching attention back in forth between concurrent tasks reliably in a timely manner. Laboratory studies are being conducted to determine the nature of this difficulty.

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.

A summary of progress to date was published in ASRS Directline. This article was quickly reprinted in its entirety by several major flight safety magazines, including: Airline Pilot (ALPA), USAirways Safety on Line, Flight Safety (Canadian Airlines), Independence (UPS Pilots Association), and USAF Flying Safety. Several airlines and the USAF Reserve Command have requested presentations on this project.

Products

Training guidelines and scenario development methodology.

▪

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.

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. This research is being integrated with the Instructor/Evaluator Calibration Tool developed by the University of New Mexico.

Products

- Research reports.

Project Title
Operating Documents Human Factors Project

Performing Agency
NASA-Ames and University of New Mexico

Background

The original project, Integrating CRM into Crew Procedures, developed a systematic approach to integrating Crew Resource Management (CRM) procedures with normal and non-normal checklists. This emphasis on checklists is only part of the solution, because checklists are just one element 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 approach has not been used systematically in the design, development, and implementation of operating documents.

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 the development and implementation of document systems. There was strong industry interest and participation in the two NASA/FAA Operating Documents Workshops conducted in FY97.

Based on the information provided on the surveys and in the workshops, 120 guidelines have been identified and modified for carrier personnel involved in the development of operating documents. Guidelines are organized into the major steps in the development and implementation process: Organization of Documents, Design of Documents, Developing and Maintaining Documents, and Transition to Electronic Media. The Manual, Developing Operating Documents, is in its final revision awaiting final publication in hard copy as well as in an interactive on-line version.

Feedback from participating carriers indicated the need for ongoing operating documents workshops. To address that need, the research team sponsored the NASA/FAA Operating Documents Workshop III, held in Orlando, Florida, in October of 1999.

This research effort has identified the need for a database to help manage the large number of documents required to support flight operations. A Documents Database (DDB) is a database of the information topics, requirements, sources, users, review, distribution, and related data essential to the efficient management of an operating documents system. To further specify the content of a DDB, a subgroup of the NASA/FAA Operating Document Group has developed a prototype user interface. This DDB user interface provides a concrete example and a usability testing platform for how such a database would be structured and used. A DDB should ultimately be tailored to meet the carrier’s individual needs while adhering to standards that would allow it to be linked to other operational and training databases. Remaining research challenges in this area include the identification of information topics as well as essential database fields to manage flight document systems.

Work on the prototype DDB user interface has lead to a collaborative effort between the NASA/FAA Operating Documents Group and the Air Transport Association's Flight Operations Working Group (FOWG). These two groups have addressed similar operating document issues from different perspectives. The FOWG has been setting the standards for digital data exchange between manufacturers and carriers. FOWG members were able to provide review comments on the DDB while learning about the user interface approach to requirements identification and refinement. In addition, the International Civil Aviation Organization (ICAO) intends to use the Manual, Developing Operating Documents, to harmonize flight document development at the international level.

Products

- Operating Documents Workshop III was held early in FY2000.
- The Manual, Developing Operating Documents, will be ready for distribution in FY2000 in a hard copy as well as an interactive on-line format.
- Additional coordination and harmonization with ICAO and the ATA FOWG has been proposed and is currently under review at the FAA.

Project Title
Aviation Team Decision Making

Performing Agency
NASA Ames

Background

Detecting and Correcting Error in Flight

Maintaining safety in high-risk engineered environments like aviation is a team effort that depends crucially on the team members' efficiency in monitoring each other's performance and on their effectiveness in intervening if they consider a decision or action to be unsafe. Unfortunately, pilots frequently have difficulty with this important crew function, especially when their interventions pose a direct challenge to the other crew member's judgment or decision-making skill. This type of problem has been identified as a "monitoring/challenging error" by the NTSB and was found to occur in over 75% of the accidents reviewed (NTSB, 1994). The problem may be manifest either in failure to say anything to correct an error, or saying something that is inappropriate or ineffective. We conducted a set of studies that sought to understand common communication strategies used to correct crew errors as a function of crew position, risk and face threat, and to determine what strategies are most effective for changing the decision or action.

Three studies were conducted to determine (1) the verbal strategies captains and first officers would use to correct errors in hypothetical situations, (2) the strategies captains and first officers judge to be most effective in getting them to change their own behavior, and (3) the kinds of communication strategies pilots actually use when confronted with errors in simulated flight. The first two studies were paper and pencil tasks and the third was conducted in a B747-400 simulator. The goal of these studies was to establish a set of requirements for training pilots to respond most effectively to reduce monitoring/challenging errors.

We hypothesized that captains would be more direct in correcting first officers than first officers would be in correcting captains' errors. For both crew positions, communications were expected to be more direct during high-risk or emergency situations than during low-risk incidents. In addition, pilots' communications were expected to be sensitive to the level of "face threat" imposed by their correction of the other pilot's errors. If the other pilot has made a serious error, calling attention to it involves a direct challenge to the pilot's status, judgment, or skill. In situations like these, politeness dictates the use of indirect speech compared to situations that are less face threatening.

In the first study, participating pilots received descriptions of aviation incidents and were asked to state how they would correct errors committed by the other pilot. As predicted, captains were more direct than first officers were: They predominantly used commands, while first officers preferred to use hints. Risk levels also influenced interventions in the predicted direction: Both captains and first officers were more direct in high-risk situations than in low-risk ones. In contrast, pilots' responses to varying levels of face threat were not consistent with predictions made by politeness theory.

The second paper and pencil study examined which types of communication were judged to be most effective in correcting pilot errors. Pilots were asked to rate how effective various communication strategies would be in getting them to carry out the speaker's intent. Both captains and first officers favored communications that appealed to a "crew concept," (e.g., "*Let's deviate to the right*"), rather than to any particular status-based model, and consistently rated commands, the most direct communication strategy (e.g., *"Turn right 10 degrees"*), as less effective than crew suggestions or preferences.

The third study examined pilots' error challenging strategies during a full-mission flight in NASA's B-747-400 simulator. Errors similar to those used in the paper and pencil tasks were scripted into flight scenarios, which participating pilots flew with a research confederate pilot trained to perform the scripted errors. Results of this study showed that both captains and first officers used error-correcting strategies that support a positive crew climate, such as suggestions and strong hints. As expected, both captains and first officers were sensitive to risk, and communicated more directly when risk was high. The influence of face threat was somewhat different for captains and first officers, as expected from their differences in rank: First officers were most direct when risk was high and face threat was low.

Results of this study provide a basis for developing training guidelines to help pilots develop more effective error correction strategies. Findings are being used in a training module by one of the participating carriers. Work is continuing with them in developing training materials based on incidents that have occurred within their carrier.

II. Contextual Factors in Aviation Decision Errors

A second project seeks to reduce the frequency of decision errors on the flight deck by (a) identifying factors that contribute to those errors, and (b) developing strategies to aid crews in avoiding or mitigating errors. A reanalysis by our research group of 37 accidents previously analyzed by the NTSB (1994) indicated the prevalence of one type of decision error that we have called a "plan continuation" error (PCE). This is a decision to continue with the original plan of action in the face of cues that suggest it might be wise to reconsider the plan. This type of error occurred most often in the approach and landing phases of flight, and most often consisted of omissions. A more recent set of 14 accidents occurring between 1990 and the present was analyzed to determine whether any particular factors were associated with accidents involving certain types of errors (PCEs, tactical decision errors, or errors other than decision errors).

Consistent with the earlier data set (NTSB, 1994), tactical decision errors occurred in 10 of the 14 accidents (or 71%). Plan continuation errors are actually a subset of tactical decision errors, and occurred in six of the 10 tactical decision accidents. The remaining four accidents involved no identified decision errors.

Four major categories of contextual factors were examined, using the causes identified by the NTSB for each accident. There were *cue ambiguity* (missing or unclear information), *dynamic risk* (changing conditions), *organizational factors* (training, corporate culture/goals), and *individual stressors* (e.g., fatigue).

First, the relative frequency of the four contextual factors was examined. Their frequency of occurrence varied significantly across all accident types, with organizational factors and cue ambiguity most common (mean frequencies of 3.6 and 3.0 occurrences per accident, respectively); dynamic risk conditions were intermediate (2.5), and individual stressors least often played a role (1.0). The number of influences associated with each type of accident also differed significantly, with a greater number of influenced associated with the two categories of accidents involving decision errors. A mean of 3.1 contextual factors was present in accidents involving PCEs and a mean of 2.9 factors for those involving other types of tactical decision errors. Non-decision error accidents involved only 1.4 contextual factors per accident. This pattern suggests that as the number of potentially disruptive contextual factors increases, decision errors are more likely to increase.

A significant interaction also was found between the contextual factors and types of accidents. Organizational factors had a significant effect on accidents involving decision errors, especially those involving non-PCE decision errors. This pattern may reflect inadequate training by the company, or goal conflicts inadvertently induced by corporate culture and values. Ambiguity and dynamic risk had similar effects on all three types of accidents, perhaps leading to poor situation awareness due in part to difficulty in updating one's situation model in changing conditions. In contrast, individual stressors had a significant effect only on plan continuation errors. This pattern of findings suggests that when conditions and their associated risk levels were rapidly changing, and when pilots were fatigued or otherwise stressed, they had difficulty updating their situation models and "pressed on," when perhaps they should not have.

A caveat regarding these analyses is that no baseline data are available that indicate how often these contextual factors are present in flights that are conducted safely, nor is it confirmed that the set of four major contextual factors are the only, or the most important, ones.

While further analyses remain to be conducted, these initial analyses suggest that training should focus on certain classes of contextual factors within particular training modules. Current work with a major carrier to develop a course on Risk Assessment that emphasizes some of the lessons learned from our analyses is in progress. Once fully developed and operational, validation studies will take place to determine whether pilots are better able to reduce their risks, update their models, and make better decisions in dynamic conditions after participation in the course.

Title
Airline Pilot Training Survey

Performing Agency
American Institutes for Research, Washington, DC

Background

The Federal Aviation Administration (FAA) is presently in the process of rewriting existing regulations that pertain to pilot training. The proposed revisions will change the requirements for traditional Part 121 training, and they will codify the requirements of training conducted under SFAR 58, the Advanced Qualification Program (AQP).

A valuable source of information regarding the strengths and weaknesses of existing training, both Part 121 and AQP, are the pilots who have been the recipients of that training. The FAA believes that the quality of its decision making with regard to future changes in those regulations would benefit significantly from a thorough unbiased assessment of pilot opinion regarding these training programs. Accordingly, the FAA has issued a grant to the American Institutes for Research (AIR) to conduct a survey of airline pilots' perceptions of the effectiveness of their training. AIR is an independent, not-for-profit organization with over a fifty-year history of conducting research in the social and behavioral sciences.

To ensure the success of this effort, the AIR is working closely with the Air Line Pilots Association (ALPA), Independent Association of Continental Pilots (IACP), the Allied Pilots Association (APA), and the Air Transport Association (ATA) to develop the survey's content. The survey is truly a collaborative effort among the federal government, unions, and the air carrier industry. Each group is represented on a technical advisory board that provides oversight to the project.

The survey will be sent to a representative sample of pilots during summer 2000. The results of the survey will be published in a report prepared by AIR in consultation with the members of the project's technical advisory board. It is anticipated that this report will be of great interest to industry and the air carrier pilot community as a whole.

Project Description

A three-phased approach was proposed to meet the project's primary objective. During Phase I, a detailed research plan will be developed. The research plan will be derived from a review of important airline pilot training documents, the wider body of training effectiveness research, and consultation with industry and union experts. The research plan will specify the pilot population to be studied, key subgroups that comprise the population, strategies for sampling the population, and important variables related to pilot experiences in and perceptions of their training. A survey instrument for collecting data on these variables will be developed and administered during Phase II. Survey questions will be developed based on the results of several focus groups conducted with airline pilots who have experience in AQP, Part 121, and Single-Visit training (i.e., an interim step in AQP development). The survey instrument will be pilot-tested prior to its implementation in the main study. Once the survey is finalized, a large representative sample of airline pilots will be identified and administered the survey. The airline pilot sample will be drawn at random from a larger population of airline pilots using a stratified random sampling procedure. During Phase III, data collected from this sample will be analyzed to identify pilot experiences in and perceptions of Part 121, AQP, and Single-Visit training. The procedures employed in the investigation, the analyses conducted, and the resulting findings will be presented in a detailed FAA technical report.

To date, AIR has completed most of the project planning activities proposed for Phase I. Based on these activities, a research plan was prepared describing the pilot population to be investigated, subgroups within the population, strategies for selecting a representative sample from this population, and the key variables to be measured. The research plan was presented to the project's Technical Advisory Board (TAB) on 05 January 2000 in Washington, DC. The TAB consists of representatives from the FAA (AFS-230 and AAR-100), ATA, ALPA, APA, and IACP. The research plan was briefed by AIR staff and discussed and commented upon by the board members during this meeting. Minutes from the meeting were recorded, summarized, and distributed to all participants. Currently, revisions to the research plan are underway. AIR is still working with ALPA, APA, and IACP to identify the appropriate and available databases for selecting survey participants. Access to and review of these databases is a necessary first step in establishing the size of the survey sample and finalizing our sampling strategy.

In addition to completing most of the Phase I activities, several tasks important in the development of the survey under Phase II have been initiated and/or completed. For example, five focus group meetings with pilots from APA, IACP, and ALPA have recently been completed. Data from the focus groups will be used to refine potential variables to be measured by the survey and generate specific items to measure each variable. In addition to the focus groups, AIR staff met with key airline personnel to discuss AQP, Part 121, and Single-Visit training and toured the American Airline's and Continental Airline's pilot training facilities.

Survey development and testing will continue during spring 2000. AIR will work closely with the project's TAB to develop the survey instrument and associated cover letter. The survey is scheduled for distribution during summer 2000. Analysis, results, and reporting are planned for fall.

Products

FAA Technical Report