NASA Technical Memorandum 88322

Crew Factors in Flight Operations: III. The Operational Significance of Exposure to Short-Haul Air Transport Operations

H. Clayton Foushee,
John K. Lauber, Ames Research Center, Moffett Field, California
Michael M. Baetge, Informatics General Corporation, Palo Alto, California
Dorothea B. Acomb, San Jose State University, San Jose, California

August 1986

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

TABLE OF CONTENTS

Page

SUMMARY

Excessive flightcrew fatigue as a result of trip exposure has long been cited as a factor with potentially serious safety consequences. Laboratory studies have implicated fatigue as a causal factor associated with varying levels of performance deterioration depending on the amount of fatigue and the type of measure utilized in assessing performance. From an operational standpoint, these studies have been of limited utility because of the difficulty of generalizing laboratory task performance to the demands associated with the operation of a complex aircraft.

This study examined the performance of 20 volunteer twin-jet transport crews in a full-mission simulator scenario that included most aspects of an actual line operation. The scenario included both routine flight operations and an unexpected mechanical abnormality which resulted in a high level of crew workload. Half of the crews flew the simulation within two to three hours after completing a three-day, high-density, short-haul duty cycle (Post-Duty condition). The other half of the crews flew the scenario after a minimum of three days off duty (Pre-Duty condition).

The results of this study revealed that, not surprisingly, Post-Duty crews were significantly more fatigued than Pre-Duty crews. However, a somewhat counter-intuitive pattern of results emerged on the crew performance measures. In general, the performance of Post-Duty crews was significantly better than the performance of Pre-Duty crews. Post-Duty crews were rated as performing better by an expert observer on a number of dimensions relevant to flight safety. Analyses of the flightcrew communication patterns revealed that Post-Duty crews communicated significantly more overall, suggesting, as has previous research, that communication is a good predictor of overall crew performance.

Further analyses suggested that the primary cause of this pattern of results is the fact that crewmembers usually have more operating experience together at the end of a trip, and that this recent operating experience serves to facilitate crew coordination, which can be an effective countermeasure to the fatigue present at or near the end of a duty cycle. These results have important aircrew training and aviation safety implications.

INTRODUCTION

1.1. Background

This report is the third in a series on the physiological and psychological effects of flight operations on flightcrews and the operational significance of these effects. These studies were conducted in response to a Congressional request. The original response to this request was a workshop convened by the National Aeronautics and Space Administration (NASA) and held at the Ames Research Center (ARC) in August, 1980. This workshop included representatives from the scientific community, airline pilots, and management; its expressed purpose was to make recommendations regarding the type of research necessary to understand the extent of the problem.

One of the conclusions reached at this workshop was that little was known about the effects of exposure to duty cycles on actual flight performance. Despite impressive advances in aircraft technology over the past several decades and an overall decline in the accident rate since the introduction of turbine-powered aircraft, flightcrew performance problems continue to dominate the statistics. There are, of course, many hypotheses concerning the persistence of operator performance problems in this environment, and the issue of pilot fatigue has traditionally received considerable attention. This high level of interest has stimulated a large volume of laboratory research, but much of this work is difficult to generalize to the actual operations, and the issues surrounding its applicability have prompted considerable disagreement about the extent and operational significance of fatigue-related performance decrements. As a result, NASA was asked to undertake a comprehensive program of research in order to assess whether or not fatigue-related problems are prevalent in long- and short-haul flight operations. The two major facets of this project are 1) to assess the psychophysiological effect of exposure to various types of flight and duty cycles and 2) to determine the operational significance of this exposure in terms of flight safety and efficiency. This project also sought to identify individual pilot attributes that might be associated with responses to the operational requirements of air transport flying and to identify adaptive strategies that might enable certain individuals to cope more effectively with operational requirements.

This is the second report in the short-haul series. It describes one aspect of this comprehensive program, the results of a study designed to address the operational significance question in short-haul flight operations (see Gander, Graeber, Foushee, & Lauber, 1986; for a discussion of the psychophysiological effects). In addition, a brief overview of research aimed at fatigue-performance relationships is provided.

1.2. Related Research

Aside from research focusing on cockpit design and engineering issues, there have been relatively few pilot performance studies carried out in actual operational environments (see Graeber, Foushee, and Lauber, 1984; for a more comprehensive discussion). The reasons for this state of affairs are varied. Field research is very difficult to do because of an almost complete lack of experimental control over operational events. Moreover, flight safety considerations prohibit the examination of many variables in the in-flight environment. Thus, studies aimed at the effects of fatigue on performance have generally been confined to laboratory or part-task research environments and have focused upon individual instead of flightcrew performance.

Holding (1983) reviewed many of these studies and much of the related literature on fatigue-task performance relationships. One of the major conclusions of this review and others (e.g. Gagne, 1953) is that fatigue-performance relationships are not easy to define. The literature is populated by many studies demonstrating a deleterious effect of fatigue on some type of task performance, and about an equal number demonstrating minimal or no effects. Those tasks that do show deterioration over time (presumably as a response to some type of fatigue) tend to be relatively simple tasks characterized by repetitive stimulation. The pattern of findings for more complex tasks is less clear.

The work of Bartlett (1943) is commonly cited as one of the major contributions to the literature on complex task fatigue. In this body of work (often referred to as the Cambridge Cockpit Studies), subjects were exposed to tasks which consisted of responding on aircraft-type controls to "changes" in a variety of instruments. Fatigue was manipulated by exposure to these tasks over long periods of time. The findings of these studies seemed to indicate that as alertness declined, progressively larger "deviations" were tolerated before any corrective actions were taken by subjects. It appeared that "fatigued" subjects were more prone to distraction and seemed to suffer from a narrowing of perceptual focus such that attention was reserved for items of more central importance, such as heading and speed indicators. It was further observed that performance on these tasks became more variable (as opposed to less accurate). Moreover, subjective observations indicated that subjects became more irritable with increasing fatigue ("violent" language, etc.).

On the other hand, McFarland (1953), in a study of aircraft incident and accident statistics, could find no effect for the fatigue associated with long hours of flying (although accidents are a crude measure of pilot performance). Likewise, Chiles (1955) could find no fatigue effect even though subjects performed continuously in an aircraft simulator for as long as 56 hr without rest, except for intervals where they left the simulator for testing on a tracking task. It is reported that some of Chiles' test subjects had to be carried to the test apparatus, and the absence of performance effects associated with fatigue under these circumstances is indeed striking. However, aircraft flying performance was not directly measured--the primary performance criterion was associated with the tracking task. Despite what would appear to be high levels of fatigue in this study, performance on this task was well within normal limits.

Studies of automobile driving have yielded similar inconsistent results. Brown (1967) found no performance effects associated with fatigue and found that performance on a vigilance task actually improved over time (when fatigue effects should have begun to appear). In a particularly interesting study, Dureman and Boden (1972) found some evidence of performance deterioration on a simulated driving task (as measured by steering errors and braking reactions). However, when subjects were threatened with electric shock, a factor which apparently increased arousal, performance improved. These data suggest that arousal and accompanying increases in motivation may be important mediators of fatigue-performance relationships.

It is difficult to summarize the implications of the fatigue-performance literature for air transport operations because most of the studies have not been conducted in actual flight environments, or the tasks have not been relevant to the task of flying a complex transport aircraft. Even the studies that have utilized tasks with some face-validity to piloting an aircraft (e.g. Bartlett, 1943) were crude simulations by recent standards, and the primary performance criteria were relatively simple monitoring or vigilance tasks that did not comprise the entire performance spectrum. Another problem with previous studies is the manipulation of fatigue. Most studies have artificially manipulated fatigue by keeping subjects up, exposing subjects to long and continuous performance periods, or testing subjects after rigorous physical exercise. Few studies have coupled consistent manipulations of fatigue with comparable performance measures, which makes comparisons across studies hazardous at best. Even fewer studies have utilized *exposure* to normal duty cycles as a means of inducing fatigue, causing further difficulties in generalizing laboratory evaluations of fatigue effects to real-world settings.

Those studies that have examined performance as a result of actual exposure to flight operations (as opposed to more artificial manipulations of fatigue) have generally not utilized actual flight performance measures for assessment of effects (for reviews of this research see Klein and Wegman, 1980; Graeber, 1982) The usual approach has been to measure subjects' performance after exposure to transmeridian flight (usually as passengers) on a battery of laboratory tests of psychomotor and simple cognitive performance. Only one study has measured performance in flight simulators using experienced pilots (Klein, Bruner, Holtmann, Rehme, Stolze, Steinhoff, and Wegmann, 1970). In this study, subjects were tested periodically after exposure to successive trips eastward and westward across eight time zones. Performance was measured by percent deviation from preset flight parameters, and exhibited significant deficits depending on such factors as time of testing, direction of flight, and number of days in the new time zone. While this is one of the few studies to demonstrate a potentially operationally significant effect of circadian dysrhythmia on performance, it left many questions unanswered. The measures utilized represent only an estimate of the effects of fatigue and jet-lag on raw flying skill and do not address other important predictors of performance in actual operations. The simulator tests lasted only 12 min, were limited to manual control skills, and did not tap any of the higher level cognitive skills required on the flight deck. Furthermore, the study examined a single-pilot operation. In multi-pilot operations there are a number of other factors that affect flight safety and efficiency.

Coupled with the problems of understanding fatigue effects on individual performance tasks is the increasing realization that individual pilot proficiency and the manual control skills necessary to operate a complex transport aircraft are clearly important, but not causal factors in the majority of incidents and accidents. An analysis of jet-transport accidents worldwide for the period 1968 to 1976 (Cooper, White, and Lauber, 1979) revealed more than 60 in which breakdowns of the crew performance and decision-making process played a pivotal role. Much of the aviation community seems to have accepted the notion that inadequate crew coordination, team performance, and the myriad variables that contribute to such difficulties are perhaps the most operationally significant problem areas in flight operations. Recommendations by the National Transportation Safety Board and the number of commercial and military training programs being developed to address "cockpit resource management" problems reflect this high level of acceptance. These attempts to address crew performance issues are in large part predicated on the notion that individuals will always make mistakes, and the crew should act as a system of redundancy to prevent more serious errors from occurring. There is a fair amount of research evidence (for a review see Foushee & Helmreich, 1986 in press) that the crew performance process is not working as well as it should be.

Pilots who are highly proficient in manual control skills continue to be involved in incidents and accidents with causes that fall outside the realm of the operators' psychomotor abilities. Yet, nearly all pilot performance research (and training) is centered around the measurement of dimensions that do not appear to be major problem areas in the current environment. No studies have examined the possible effects of fatigue upon higher-level decision-making skills and those related to effective coordination of crew resources. Holding (1983) cites this failure to use measures of higher-order cognitive function, but he also mentions neglect in accounting for relatively stable effects of fatigue such as increased irritability. For example, the increased irritability associated with fatigue (e.g., Bartlett, 1943) could make it more difficult for crewmembers to work together effectively. Thus, fatigue effects may not be apparent on individual performance parameters, but significant with respect to group performance.

In the past, the major barrier to definitive research in all of these areas has been the lack of a research environment that combines experimental control with high generalizability to operations. Fortunately, the rapid advancement of simulator technology has finally provided an ideal laboratory for the study of these aircrew operating problems. It is now feasible to realistically simulate virtually every aspect of line operations to the extent that actual trips can be flown in a simulator that are almost indistinguishable from those in the airplane. Because of this high degree of realism, it is possible to study individual and group parameters with almost complete confidence that results generated in the simulator are representative of the real world. Moreover, the simulator affords a high degree of experimental control and allows the study of operational problems too dangerous to examine in flight. It is clear that arousal levels associated with realistic simulated flight more closely approximate those experienced in actual flight (at least when compared to laboratory performance tasks). In fact, Ruffell Smith (1979) reports arousal levels in high-fidelity simulation that closely approximating levels measured in flight.

With these issues in mind, our objectives were to conduct an investigation of the effects of exposure to high-density, short-haul airline operations that would be highly generalizable to the operational environment. Specifically, we were interested in whether there were any behavioral or performance changes associated with typical short-haul duty cycles, but more importantly, we were interested in whether these behavioral and performance changes (if any) were operationally significant. For the purpose of this investigation, operational significance was defined as performance that has a major bearing on flight safety and efficiency. Unlike the majority of previous research efforts, this study sought to measure performance on a realistic group performance task that induced high motivation and arousal. The task was also oriented toward complex decision-making skills rather than simple, repetitive tasks involving manual control skills.

METHOD

2.1. *Study Design Overview*

Since the primary objective of this study was to assess the operational significance of exposure to various types of flight and duty cycles, subject crews were evaluated either before or after they had completed a three-day trip. The "target trips" in this study consisted of high-density short-haul airline operations averaging eight hr of on-duty time per day and five takeoffs and landings, with at least one day (usually the last) averaging close to eight takeoffs and landings and thirteen hr on duty. There were two experimental conditions. 1) Subjects in the "Post-Duty" condition flew the simulation as if it were the last segment of a three-day trip, whereas 2) subjects in the "Pre-Duty" condition flew the simulation after a minimum of two days off duty (usually three), as if it were the first segment of a trip. Twenty volunteer crews were run in the study (40 pilots). One Post-Duty crew was eliminated from all analyses, when it was ascertained that the captain had been informed about the operational events associated with the scenario. This left 11 crews in the Pre-Duty condition and 9 crews in the Post-Duty condition.

2.2. *Subjects and Recruitment*

All subjects were recruited from the ranks of active line pilots in one domicile of a major U. S. air carrier. The decision to use pilots from one carrier was necessitated by differences in standard operating procedures and aircraft configurations that are common across air carriers. Thus, for experimental control purposes and to maintain the highest possible degree of realism, the scenario was designed to simulate precisely this particular airline's operation. All of the subject pilots were currently flying the transport aircraft used for the simulation exclusively in line operations.

Both airline management and union officials were contacted and briefed about the purpose of this investigation. When approval was received from both, all pilots in this domicile were sent a brochure describing the purpose of the project and providing information about the extent and type of participation requested from volunteer subjects.

Recruitment was based on several factors. Every month, the investigators received a copy of the trip pairings (which crewmembers are flying what trips) from crew scheduling. From these pairings it was possible to determine which crewmembers were flying and when, as well as, those crewmembers with days off during any given period. The pairings were examined to determine which trips fit the criteria for the study (see Gander et al., 1986 for a thorough discussion of trip characteristics). In general, target trip selection was based upon such factors as the number of segments flown in a day, duration of the duty day, and length of nighttime layover. The most difficult duty cycles for this aircraft type were selected each month, and were characterized by a high number of takeoffs and landings, long duty days, and the shortest overnight layovers.

When these trips were identified, the pilots scheduled to fly them were contacted by telephone (if they had not previously participated in the simulator study) and recruited for assignment to the Post-Duty condition. They were scheduled for participation within two to three hr after completion of their duty cycle.

The subjects for the Pre-Duty condition were contacted if the trip pairings indicated that they were scheduled to be off duty for a period of at least three days. Due to the nature of trip pairings in this airline (and in short-haul operations in general), this is a relatively long continuous period off duty, with the exception of vacation time which is at the discretion of the pilot. Although all Pre-Duty subjects were scheduled for the simulation after it was ascertained that they would have three continuous off-duty days, 3 out of the 22 subjects in this condition had only two off-duty days. This occurred because of unscheduled extensions of prior trips that were usually the result of aircraft or weather difficulties forcing alterations of the scheduled pattern. This could not be determined until those involved in schedule alterations had arrived for the simulation experiment.

When contacted by telephone, all potential subjects were briefed on the details of the experiment by the principal investigator. This briefing included the purpose of the simulation study and what was expected from each volunteer subject. All of the subjects were informed that the simulation would be conducted exactly as if it were an actual flight, and crews were aware that the flight was a GSO-RIC segment, but they were not informed of any other details of the simulation scenario.

After obtaining tentative agreement to participate from prospective subjects, a considerable amount of effort was spent explaining the level of experimental realism so that the subject pilots would begin to think of the simulation as "just another flight segment." All pilots were asked to bring their own headsets, flight bags, including all charts and manuals, and anything else they normally take along during line operations.

Although subjects in this study (or any other part of the project) received no compensation for their participation, approximately 60% of pilots contacted agreed to participate in the experiment. The refusal rate was higher than in the

Gander et al. (field) study (15%), and most likely reflected the fact that participation would have caused considerable personal inconvenience. The highest refusal rate was among commuters not living nearby the simulator facility and crew domicile, since participation by many of these individuals would have caused them to remain away from home for an extra night. Field study participation, although requiring more overall time, did not entail any alteration of an individual's normal schedule. Of the pilots contacted, who resided within two hr of the simulation facility, the participation rate was above 90%.

2.3. Confidentiality

Because of the sensitivity of pilot-performance data in general and the focus upon operational significance in this investigation, it was necessary to guarantee all pilots participating in this study complete confidentiality. This was done in several ways. First, all data in this study were identified by a four-digit code number, which identified individuals as captains or first officers. Thus, it was not possible for anyone, including the NASA investigators, to identify any of the participating pilots by name. Second, although the simulation was conducted utilizing the facilities of the participating airline, the company was not involved in any way with the actual data collection. NASA leased the simulator from the company and provided its own operators for the purposes of this investigation. None of these individuals was employed by the subject airline.

2.4. Experimental Equipment

The simulator that was utilized for this study had a six-degree-of-freedom motion platform and four-window visual system. It was manufactured by CAE, Inc. of Canada, was equipped with the special effects and programmed with the aircraft performance data required to meet FAA Phase II certification, and had successfully completed the certification process.

Only minor simulator modifications were necessary for the experiment. These included the provision of input jacks so that background air traffic control (ATC) communications and the Automatic Terminal Information Service (ATIS) recordings for the various airports could be fed into the VHF radios no. 1 and no. 2. All equipment that was not functional in the simulator, such as the weather radar (the scope was present but not functional), the Automated Communications Addressing and Reporting System (ACARS), and the VHF no. 3 radio (present but not functional) was placarded as inoperative just as they would have been in the actual line operation. It was also necessary to develop software so that the simulator computer would output time-coded aircraft performance parameters.

A portable video-data-acquisition-system (PVDAS) was designed and built by NASA for the purpose of this investigation. The advantages of this system are that it is pre-wired, and all components are mounted in a shock-proof case that can be rolled into the simulator cab and set up quickly. All of the simulator sessions were videotaped using PVDAS which included a two-channel, one-half inch VHS videotape recorder, monitor, camera, microphones, time-code generator, and three microcassette recorders. Both crewmembers and the air traffic controller were wired with lapel microphones; the captain on channel one and the first officer and controller on channel two. A single low-light, auto-iris, black-and-white, video camera was located above and behind both crewmembers at the approximate location of the cockpit door. This camera angle allowed a view of the center console, the throttle quadrant, and most of the instrument panel, with the exception of the overhead panel. About one-half of each pilot was visible, which allowed most actions to be monitored, but did not allow individual pilots to be identified. The time-code generator was used to synchronize the videotapes with the aircraft performance data and imprint a permanent record of timing information on each videotape. The microcassette recorders were also prewired and plugged into jacks mounted inside the simulator cab, which fed the VHF no. 1 and no. 2 radios. Each recorder was wired to switches so that any tape in any recorder could be fed into either radio. In this way, it was possible to provide all ATC background communications and ATIS information as appropriate depending on which radio was being utilized by the crewmembers. In this system, transmissions from the "live" controller blocked out the ATC background communications, but it was usually convenient for the controller to time instructions to the flightcrew accordingly.

Background ATC tapes were produced with the cooperation of each ATC facility that the pilots would be in contact with in the area of the simulated flight. These included GSO Ground Control, GSO Tower, GSO Departure, RIC Approach, RIC Tower, RIC Departure, ROA Approach, and ROA Tower. Each facility supervisor taped approximately 1 hr and 30 min of actual transmissions during a busy traffic period. The tapes were then edited for inconsistent information and normal pacing to produce a suitable amount of material for the simulated flight. Washington Air Route Traffic Control Center (ARTCC) tapes were produced by the investigators in the Man-Vehicle Systems Research Facility (MVSRF) air traffic control simulation at ARC using the voice-disguising equipment to simulate different aircraft in contact with Washington Center. These tapes of background communications were quite effective in producing a high level of operational realism.

Tapes for ATIS were produced by the investigators using 60-sec, continuous-loop, microcassette tapes. The tapes were mounted well in advance of the flight's approach to the appropriate facility and could be monitored at any time and for any period because of the continuous-play feature. Tapes were produced for each of the airports used in the experiment, GSO, RIC, and ROA.

2.5. Personnel

The experiment was run with three staff members. The principal investigator dealt with the subject pilots, coordinated the data collection, and made decisions regarding the handling of flightcrew requests. He also acted as the number one flight attendant if requested by crewmembers. A second person delivered all air traffic control communications to the flightcrew and operated the simulator, including the introduction of scenario events. The third person was a retired check captain in the aircraft simulated. His role was primarily performance observation (see section 2.10), but because of his familiarity with company operations, also served as a technical adviser to facilitate operational realism. All three members of the experimental staff were present in the simulator during each run, but were out of the flightcrew's field of vision. Flightcrew members were informed that there would be no intervention by any of the experimental staff and that they should request all information through normal channels. They were advised that if simulator malfunctions caused unplanned events, they would be informed of the appropriate action. Once "airborne", intervention was necessary in only one case, when a simulator malfunction forced an early termination of the experiment approximately 10 min prior to touchdown at ROA. All data were used up until the time of the malfunction. The only exception to the non-intervention rule was that crews usually needed help during taxi operations because of poor visibility conditions and a lack of complete fidelity in the GSO ground scene (taxiway lights were low-intensity).

2.6. Experimental Procedure

When each crew arrived at the simulation facility, they were met by the principal investigator who again briefed the subjects on the purpose of the investigation. The importance of operational realism was again emphasized, and the pilots were urged to treat the simulation as if it were an actual flight. They were informed that they would have access to all resources that they would normally have in-flight, including complete ATC services, dispatch, access to maintenance, ATIS, and so on. However, no details were provided to flightcrews other than the flight's origin and destination. Subjects were asked to treat the flight as a "captain's segment", captain flying the aircraft and first officers performing support duties. This was done for experimental-control purposes and to assure consistent performance evaluation at both crew positions. After any questions were handled, the flightcrew was escorted to a room set up as a flight dispatch facility, where they met with the flight dispatcher (played by the observer). The flight dispatcher provided the crews with the same information and paperwork that they would have normally received prior to a trip. This information included the route of flight, weather information for the vicinity of flight, fuel information, weight and balance information, number of passengers, and so on. All information was provided on standard company forms.

Following receipt of these materials, the crews performed flight planning duties normally reserved for this period. The dispatcher remained available to render any assistance the flightcrew might have requested. Crews were given a standard departure time based on their arrival at the simulator facility and were given adequate time for flight planning activities. After the flight planning phase, crews entered the simulator cab, were wired with microphones, and began preflight checks of the aircraft.

When it appeared that the crew was nearing completion of preflight checks and as departure time approached, the simulator operator extinguished the "aft cargo door light," prompting most flightcrews to realize the aircraft was loaded and ready. Shortly thereafter, the captain was called over the interphone by ground personnel and informed that the aircraft was ready. At this point, most crewmembers radioed for a clearance to pushback from the gate. Push-back was simulated by activating the simulator motion platform at the appropriate moment, which produced a small jolt quite similar to that associated with the tug beginning to push the aircraft backward. The recording of performance data began at push-back (see section 2.10.3). Taxi and ground operations were accomplished in accordance with standard operating procedures, as was the remainder of the simulated flight (the details of the scenario are discussed in section 2.7).

After landing, the crews were escorted back to the flight dispatch area where they filled out a number of rating forms (see sections 2.8-2.10). They were then debriefed by the principal investigator. In many cases, this was a rather extensive process since pilots were often anxious to discuss various events and actions taken during the simulated flight. Subject pilots were then thanked for their participation and received a lengthy explanation from the principal investigator concerning the importance of not discussing the details of the scenario with any other pilots (who might have later been experimental subjects). The initial briefings (when subjects first arrived at the facility) indicated that this procedure was quite successful. Only one captain (see section 2.1) was apparently aware of some of the scenario events. This was confirmed by initial questioning and later behavior during the simulated flight, and this crew was dropped from all analyses. All subjects were promised copies of the results when they became available.

2.7. Simulation Scenario

The experimental scenario was designed primarily to assess factors related to overall crew performance and decision-making. It was not intended to be a test of manual control skills, although data pertinent to these skills were available so that they could be assessed in certain critical flight phases. The scenario design was influenced by a pattern of events often seen in past incidents and accidents--relatively minor mechanical problems complicated by environmental and operational events.

Outlines of potential scenarios were developed by the NASA investigators and taken to the participating airline where specific details were developed in conjunction with subject-matter experts. For this purpose, the experts were two retired company management pilots were hired for assistance in standardizing scenario events to company operating procedures. Typical environmental conditions for the proposed area of flight were considered in great detail, so that scenario events relating to these conditions would be realistic to the experimental flightcrews. The experts also prepared all trip paperwork in standard company format and evaluated all events and procedures for accuracy and realism.

Following this process, nine "pre-test" runs were conducted with qualified flightcrews to refine procedures, train the experimental staff, evaluate the performance assessment techniques, and test scenario events. Feedback was elicited from pre-test subjects, and the videotapes were extensively evaluated. This process allowed continual refinement until the experimental scenario was finalized.

When crews arrived at the simulation facility and picked up the trip paperwork, they found the weather information for the vicinity of flight characterized by a frontal system passing through with low ceilings and visibilities (see Figure 1). This included the departure airport, which was still acceptable for takeoff but nearing the legal limit (RVR 1600, 1/4 mi). The airplane was relatively heavy, and the takeoff was runway-limited because the longer runway at the departure airport (GSO) was closed. Aircraft gross weight was 104,000 lb. Since weight was critical, it was dispatched with a minimum, but legal amount of fuel (13,200 lb). Because of this and the poor weather, which increased the probability of diversion, crews should have been concerned about the amount of fuel (possibility of extra fuel needed for a diversion), and as in the real world, crews had the option of requesting additional fuel from the flight dispatcher. This, of course, had implications for aircraft weight, and if more fuel was requested (prudent under the circumstances), baggage had to be off-loaded. The dispatcher was instructed to explain the implications of extra fuel to flightcrews and otherwise act reluctant to provide the extra fuel (as is sometimes the case). However, if the flightcrew could not be persuaded to go with the legal minimum, the dispatcher agreed to add 5,000 lb extra fuel. Eighteen out of the twenty crews asked for and received this extra amount.

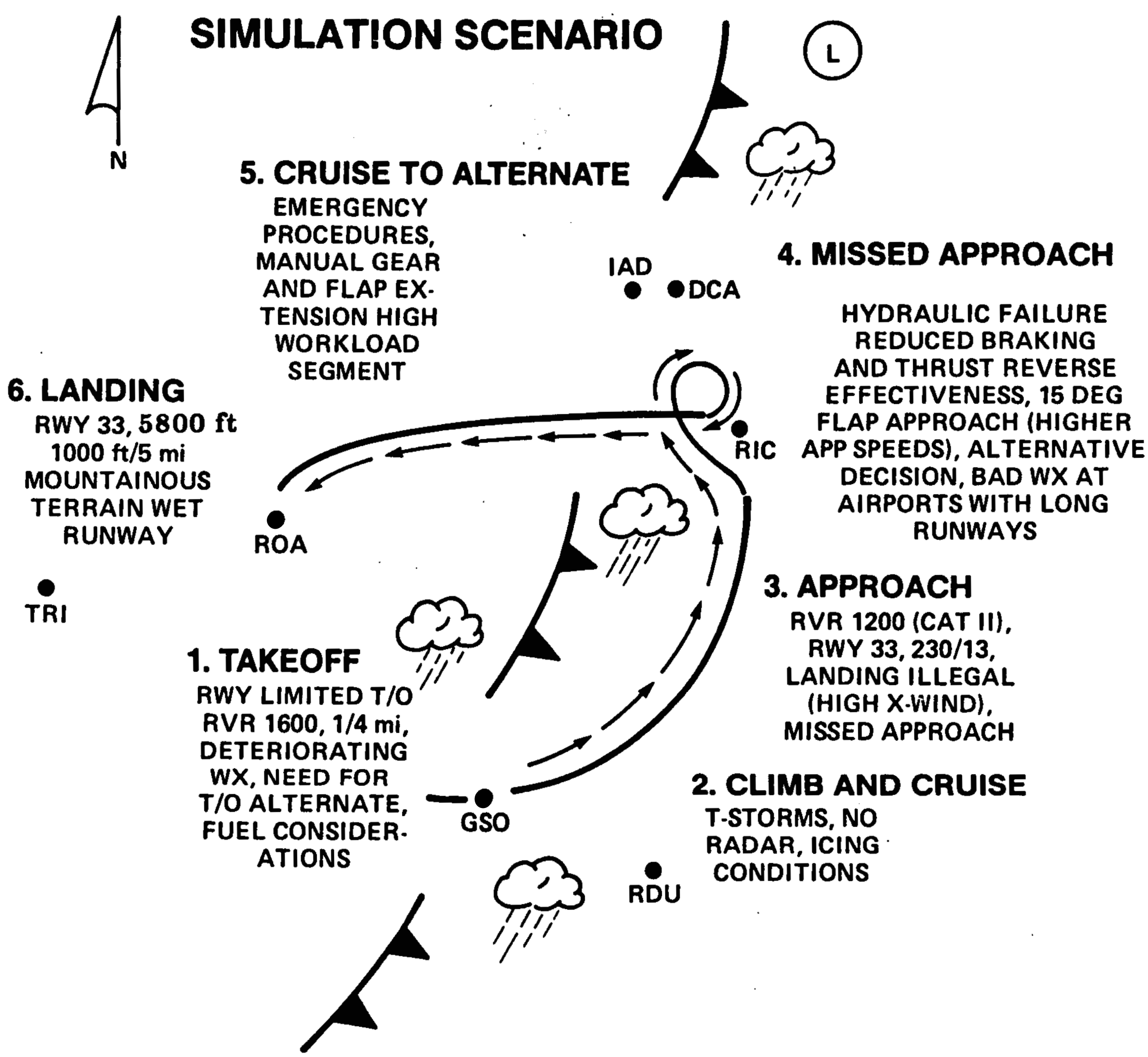

Figure 1. Overview of the simulation scenario (WX=weather, X-wind=crosswind)

After the crew had preflighted the aircraft and received clearance to push-back and taxi, ATC issued a special weather observation to all aircraft in the vicinity indicating that the weather had deteriorated (RVR 1600, 1/8 mi, rain and fog). The operational implications were that takeoff was still legal, but landing was not, and this meant crews were legally required to obtain a takeoff alternate, in case mechanical problems forced a quick return for landing. If a takeoff alternate was requested, ROA was provided by dispatch since it had the best weather within a relatively short distance from GSO. ROA was also given as the regular alternate landing site for RIC (as will be seen later, ROA was selected intentionally because of certain characteristics and because it was the site where all flightcrews were ultimately forced to land). As in all of the operational events "programmed" into the simulation, some crews realized this necessity and acted appropriately, while some did not. Once airborne, a relatively low-activity, routine segment was planned in order to allow crews to relax so that their behavior would more closely approximate actual flight behavior. This procedure has been found to be effective in both training and research applications of full-mission simulation (e.g. Lauber & Foushee, 1981). For this reason, most of the performance measures were taken after the end of this routine segment.

A few minor events were inserted in the low-workload segment as measures of crew vigilance. These included icing conditions (to make sure crews were paying attention to environmental conditions and utilized the anti-ice system when necessary) and unexpected rain and moderate turbulence along the route of flight. The aircraft was dispatched with the weather radar inoperative, and "vigilant" crews would have checked with ATC for ground-based radar advisories or pilot reports (Pireps) to assure sufficient separation from potential severe weather. Evaluation of crew awareness of such events was included in both observer-rating dimensions and in the error analyses.

The "high-workload" phase of flight was initiated when crews began their approach to the destination airport. When they received standard weather information for the arrival (via RIC ATIS), they discovered that the weather was poor and required a complicated, instrument approach (RVR 1200, Category II approaches to runway 33 in progress). Furthermore, there was a substantial crosswind on the active runway, and which was close to the legal limit (winds from 230 degrees at 9 knots). As the crews continued their approaches and contacted the tower for landing clearance, they were advised by the RIC Tower that the winds had increased to 13 knots from 230 degrees, which was 3 knots over the legal limit of a 10-knot crosswind component for a Category II approach. Some crews realized this and executed the required missed approach, and some did not. In any event, all crews, whether or not they realized that landing was illegal, were forced to execute a missed approach because those who continued the approach to RIC did not have visual contact with the ground at the decision height of 103 ft above the surface. During the missed approach, crews experienced a "System A" hydraulic failure. The implications of this malfunction were complex: 1) landing gear and flaps had to be extended manually; 2) braking effectiveness was reduced, which meant increased stopping distances (the in-board brakes anti-skid system and thrust reversers had accumulator pressure only); 3) a 15-degree flaps approach was dictated when 30 to 40 degrees is the norm (15-degree limit is imposed in case a missed approach is necessary); 4) once the landing gear was extended manually, it could not be raised again (which has substantial implications for fuel consumption and subsequent missed approaches); and 5) nosewheel steering was inoperative (which meant that the aircraft would have to be towed off the runway.

Crews were then faced with a number of complicated decisions to make and procedures to execute. They had to decide where they were going to land, since the original destination did not have legal landing conditions, and in some cases there was only a limited amount of fuel (recall the original dispatch with minimum fuel). They had to diagnose the failure, realize the implications, and secure the failed system. Since higher approach speeds and reduced braking effectiveness were primary problems, it was clear that the most desirable alternate landing site was one with a relatively long runway.

The operational problems induced by the hydraulic failure were more severe than they normally would have been because of poor weather in the general vicinity (this was consistent with the initial weather briefing that crews received in dispatch prior to the flight) and limited fuel. The airports that should have been considered as alternates included TRI, IAD, DCA, RDU, CLT, and GSO, but all had visibilities of 1/4 mi or less, and some were closed due to weather conditions. The only reasonable alternate was an airport (ROA) with acceptable weather, but a with a relatively short runway (runway 33 at 5800 ft) that was wet, sloping downhill, and adjacent to mountainous terrain. ROA had a 1000-ft ceiling, 5 mi visibility, and winds from 360 degrees at 10 knots. The relatively good ROA weather was not surprising to most flight crews, who were familiar with the fact that it is an area frequently characterized by favorable conditions when other airports are marginal. Although the relatively short runway and mountainous terrain posed a dilemma for many flightcrews, it was the best choice compared to the other possible alternates.

Another feature of the scenario was that the manual-gear and flap-extension procedures were time-consuming and required a fair amount of pre-planning. Moreover, the flight time from RIC to ROA was relatively short (approximately 30 min), so time had to be apportioned carefully. In short, this simulation required a high level of crew coordination for effective performance--a good test of high-level decision-making and crew performance. It should again be stressed that some of the features of this scenario are similar to those seen in past incidents and accidents.

2.8. Demographic Data

Each subject pilot was asked to complete an extensive background questionnaire compiled to obtain information on demographic and lifestyle variables such as sleep, nutritional habits, and personality profiles related to pilot performance. (see Gander et al., 1986; for a more complete discussion). While many of these variables were intended for use in the physiological investigation, it was necessary to obtain this information in this study in order to assure that subjects in each condition were matched on variables such as age, general health, and experience. However, the relatively small sample size did not allow extensive analyses of performance as a function of these variables.

2.9. Fatigue Measures

Subject perceptions of fatigue were assessed utilizing the same techniques that were developed for use in the field study (see Gander et al., 1986; for a detailed discussion of these measures). Immediately after the simulation, subjects reported their sleep-wake schedules for each of the four previous nights. They were also asked to rate the quality of their previous night's sleep on four dimensions: difficulty falling asleep, deepness of sleep, difficulty arising, and how restful the sleep was. In addition, subjects completed a 26-item mood adjective checklist (e.g., Moses, Lubin, Naitoh, & Johnson, 1974) and estimated their level of fatigue by placing a mark on a 10 cm. line representing a continuum from most alert to most drowsy (e.g., Wever, 1979). All of these measures were taken from the "Daily Log Book" (see Gander, et al., 1986) used in the field investigation, and are shown in Appendix D.

2.10. Crew Performance Measures

A variety of measures were utilized in an effort to assess the performance of flight crews in this simulation study. These measures were aimed at the assessment of both individual and crew performance parameters. Since the primary objective was to determine whether any observed performance changes were operationally significant, a heavy emphasis was placed upon methods designed to quantify the crew performance process. As previously discussed (section 1.2), the reason for this emphasis was the finding that the vast majority of incidents and accidents in air transport operations appear to be due to breakdowns in crew performance rather than a lack of individual knowledge and skill.

The crew-performance measures included expert observer ratings, subjective assessments of workload, aircraft handling data, error analyses (both real-time and videotape), and crew communication patterns. Each will be discussed in detail.

2.10.1. Observer Ratings. Two types of ratings were obtained. First, a rating form was developed that was organized into categories relevant to the performance dimensions of interest in the test scenario. This instrument can be seen in Appendix A. It was partitioned by phase of flight (i.e., preflight, taxi/takeoff, climb, cruise, approach/missed approach, emergency procedure/hydraulic failure, cruise to alternate, and approach/landing). Within each section were the individual dimensions to be rated. The expert observer was asked to rate both the captain and the first officer on each dimension as it was observed and if it was applicable. Each dimension was scored on a five-point Likert scale. Scale anchors were defined as follows: 1=below-average performance; 2=slightly below-average; 3=average; 4=slightly above-average; and 5=above-average performance. Many of the dimensions were common to each phase of flight (i.e. crew coordination/communications, a/c handling, planning/situation awareness, procedures, overall performance, etc.). However, some were relevant only to specific flight phases or situations (e.g. thunderstorm awareness, stress management, takeoff alternate, etc.). If for some reason the category was not observed (not appropriate or the observer was uncertain), the observer was instructed to circle the "n/a" response provided with each scale. These ratings were made "real-time" as the simulation progressed. This type of rating form was designed as a systematic approach to the types of performance judgements made routinely by supervisory check pilots in training and evaluation. Thus, the observer (a retired check-captain) was highly experienced in making such ratings.

The second type of rating (Appendix B) was more general and completed by the expert observer immediately after the simulation. It consisted of nine categories (overall knowledge of a/c and procedures, technical proficiency, "smoothness," crew coordination and internal communication, external communication, motivation, command ability (for captains), vigilance, and overall performance). The observer completed this overall form for both captains and first officers These ratings were also made on five-point Likert scales (anchors defined same as above) and were intended to assess the expert observer's overall impression of performance throughout the simulation on each dimension.

2.10.2. Subjective Workload. Pilot perceptions of workload were assessed via a technique developed by Hart and coworkers (e.g. Hart, Battiste, & Lester, 1984). This rating form consists of 10 workload-related dimensions: task difficulty, time pressure, performance, mental effort, amount of attention required, complexity, busy-ness, motivation, fatigue, and overall workload (see Appendix C). Each dimension was scaled on 7-point Likert scales with high numbers indicating larger amounts of the dimension in question. These measures have proven useful in assessing subjective levels of pilot workload in a number of experimental settings.

2.10.3. Aircraft Handling Data. These data were recorded directly from the simulator computer. Twelve parameters related to the aircraft configuration and handling were sampled every 15 sec. These included: airspeed, altitude, vertical speed, magnetic heading, engine #1 (engine pressure ratio) EPR setting, glideslope deviation, localizer deviation, gear position, flap position, #1 navigational frequency, and #1 DME reading, and elapsed time (see Appendix F). All measures were time synchronized with the videotaped records so that other performance parameters could be examined along with aircraft configuration. These measures were utilized primarily for analysis of final approach data at ROA and for crosschecking during the error analyses.

2.10.4. Error Analyses. Error analyses were undertaken using two independent sources of data. First, the expert observer kept a record of all errors observed during the course of the simulation. The second source of error data came from the videotape records. Using these records, two independent, "blind," observers reviewed the tapes for operational errors. When an error was recognized by one or both observers, the tape was stopped and the segment containing an alleged error was reviewed at least twice by the observers. After this process, both observers had to agree that the error had occurred or it was not counted in the analysis. Both sources of error data (real-time data collected during the simulation and data from the videotape analyses) were compared for reliability. The video-tape error-coding process captured all errors recorded real-time by the expert observer, however, it also revealed others that were missed during the course of the experimental runs. This is not surprising considering the fact that the videotapes could be endlessly reviewed. Reliability analyses revealed 81% agreement between both sources of error information, yielding a high level of confidence in these data.

Since assessing the operational significance of performance differences was a primary objective, an attempt was made to categorize errors according to their level of severity. This process was accomplished by both of the observers who had undertaken the videotape error-analysis. A three-level classification was utilized--Type I errors were defined as minor, with a low probability of serious flight safety consequences; Type II errors were defined as moderate severity, with a stronger potential for flight safety consequences; and Type III errors were classified as operationally significant errors, those having a direct negative impact upon flight safety.

Examples of Type I errors include missed clearances that were quickly corrected, checklists not run according to standard operating procedures (missing an item, or accomplished from memory), and altitude deviations of less than 200 ft. Type II errors included the failure to notify the company or ATC of an abnormal situation, altitude deviations between 200 and 400 ft., failure to use ice-protection, over or under speed for current configuration, delayed recognition or handling of the hydraulic failure, and failure to run a required checklist. Type III errors included failure even to recognize that the hydraulic failure had taken place, deployment of speed brakes and thrust reversers prior to touchdown, failure to consider alternates other than the company recommended ROA, failure to notice the crosswind or takeoff and landing restrictions, improper handling of the abnormal procedures, and altitude deviations of more than 400 ft.

Each videotape observer rated each error for severity and both observers discussed the rating for each recorded error. In cases where there was disagreement between observers about the correct classification, each observer presented his case in an attempt to reconcile the discrepancy. In most cases, agreement was reached. However, in a few cases when the observers could not agree, the error was assigned the lesser severity rating (there were no cases in which there was substantial disagreement between observers on a single error such as a Type I versus a Type III rating). All data that was used for analytical purposes represented the observers' agreed upon ratings.

2.11. Flightcrew Communication Patterns

Since past research has shown the interaction process of flightcrew members to be a significant predictor of flightcrew performance (e.g., Foushee & Manos, 1981), extensive analyses of communications data were undertaken in this investigation. The procedure used to analyze within-cockpit communication patterns was adapted from the Foushee and Manos (1981) procedure that was in turn derived from the work of Bales (1950). Using this approach, each statement or phrase was coded into one of eighteen categories of communication: command, observation, suggestion, statement of intent, inquiry, agreement, disagreement, acknowledgement, answer supplying information, response uncertainty, tension release, frustration/anger/derisive remark, embarrassment, repeats, checklist, non-task related, non-codable, or ATC communications. Coders worked independently. If there was any doubt about how a given speech act should have been classified, the coders were instructed to place them in the non-codable category. Thus, all speech acts were included in the total communications analyses, even those classified as non-codable. A complete list of communication categories and operational definitions is contained in Appendix E.

Two coders were trained extensively in the coding procedures by the principal investigator. The pre-test videotapes were used for training purposes so that prior to actual coding, neither coder had seen the experimental videotapes. The training consisted of the selection of particularly demanding 10-min segments of the pre-test tapes on which coders practiced transcribing and coding. A point-by-point agreement method was used to calculate interrater reliability (e.g., Kazdin, 1982). This method is substantially more conservative than commonly used methods which compute interrater agreement based upon total frequencies. The point-by-point method is more conservative because it takes into account both the number of agreements and disagreements (or instances where one observer recorded one category and the other either recorded a different category or nothing at all). Total-frequency-based methods do not account for disagreements or non-events in this manner. Using this method, 71% agreement was established between the two coders.

After the coders' training had established a satisfactory level of reliability, each was randomly assigned half of the tapes to code. Each coder was assigned an equal number of tapes in the Pre- and Post-Duty conditions to rule out potential bias. Moreover, coders were blind to the condition when they were performing the communications analyses.

Although it would have been desirable to have each coder review all of the tapes, this was impractical due to the extraordinary amount of time necessary to code each tape (up to 40 hrs. per experimental run). As a compromise procedure to ensure adequate interrater reliability, coder agreement was checked halfway through the coding process and at the end using selected segments of the pre-test videotapes. On both reliability checks, an adequate level of reliability was evident, 74% on both occasions, which was slightly higher than the training criterion.

RESULTS

3.1. Demographic Data

Captains in the Pre-Duty condition averaged 41.3 yr of age and had been employed by their airline for 14.8 yr Captains in the Post-Duty condition averaged 42 yr of age and had been employed for 15 yr. Neither of these differences was statistically significant. Total average flight time for the two groups was also comparable. There were no significant differences on measures related to height, weight, general health, or personality characteristics.

First officers in the Pre-Duty condition averaged 39.1 yr of age and had been employed by the subject airline for 2.3 yr. Post-Duty first officers had a mean age of 39.7 yr. and had been employed for 3 yr. Average airline experience was greater for both groups of first officers since many had been previously employed by other air carriers. None of these differences were statistically significant. No other differences were significant for height, weight, or general health dimensions.

3.2. Fatigue Data

As expected, captains in the Pre-Duty condition had significantly more sleep the night before the experimental runs than did those in the Post-Duty condition--8.46 hr versus 5.71 hr ($t = 4.00$, $p = .001$). Post-Duty captains also reported marginally less sleep two days prior to the simulated flight. Mean sleep times for the previous night were 7.57 hr for Post-Duty captains and 8.82 hr for Pre-Duty captains ($t = 1.64$, $p < .12$). Differences three and four nights before were not statistically significant, which is indicative of the fact that Pre-Duty crews were often on duty during these time periods. Despite differences in the amount of sleep, there were no significant differences in reported sleep quality between the two conditions by captain subjects.

The differences between conditions on the amount of sleep prior to the experiment for first officers were not as robust, but in the same direction. First officers in the Pre-Duty condition averaged 7.55 hr of sleep the night before, while Post-Duty subjects averaged 6.29 hr ($t = 1.96$, $p < .07$). None of the differences for previous nights were statistically significant. As for captains, no sleep-quality differences between conditions were reported for first officers.

On measures of subjective fatigue, no significant differences were evident for captains or first officers on the 10-cm.-line measure of alertness. However, on the 7-point bipolar scale for fresh versus tired, captains in the Post-Duty condition indicated that they were significantly more tired than captains in the Pre-Duty condition ($t = -2.40$, $p < .03$). The same pattern was evident for first officers--Post-Duty subjects reporting more overall "tiredness" ($t = -4.76$, $p < .001$).

Analyses of the mood data generally confirmed that Post-Duty subjects were experiencing more fatigue at the time of the simulation. As in the field study, where mood changes were strongly correlated with with levels of fatigue, Post-Duty subjects tended to report more negative mood ($t = -1.94$, $p = .06$). However, differences on the positive and activation mood indices were not statistically significant.

Taken together, these data would seem to indicate that Post-Duty crewmembers were experiencing significantly more fatigue than Pre-Duty crewmembers. They reported less sleep, more "tiredness", and more negative mood states than did Pre-Duty crewmembers. Though no attempt was made to control for the off-duty activities of Pre-Duty crewmembers (they may have been engaged in fatigue-inducing activity during this off-duty time), it may safely be assumed that these fatigue differences between conditions are associated with the duty cycle.

3.3. Crew Performance Measures

3.3.1. Observer Ratings. In the preflight segment of the simulation scenario, Post-Duty captains were rated better in crew coordination and marginally better in overall performance ($t = -2.81$, $p < .02$; and $t = -1.81$, $p < .09$, respectively), as can be seen in Tables 1 and 2. For first officers in this segment, differences were in the same direction, but not statistically significant. The lack of significant results for first officers on these measures probably reflects the fact that captains are primarily responsible for coordinating preflight activities.

Table 1. Observer Ratings of Captains on Crew Coordination During the Preflight Segment of Simulation

Crew Coordination for Captains		
	Pre	Post
Mean	3.00	3.44
SD	(0.00)	(0.53)
N	11	9

Table 2. Observer Ratings of Captains on Overall Performance During the Preflight Segment of Simulation

Overall Performance for Captains		
	Pre	Post
Mean	2.89	3.22
SD	(0.33)	(0.44)

No significant differences were evident for either captains or first officers in the taxi/takeoff segment. The same pattern was evident during the relatively uneventful climb segment; a slight trend for better rated performance among Post-Duty crewmembers, although these differences were not statistically significant. The only significant measure was for first officer ATC procedures (Table 3), with Post-Duty first officers rated higher on this measure ($t = -2.10$, $p < .05$).

Table 3. Observer Ratings of First Officers on ATC Communication During the Climb Segment of the Simulation

ATC Communication for First Officers		
	Pre	Post
Mean	2.90	3.89
SD	(1.10)	(0.93)

Both coordination and procedures (Tables 4 and 5) were rated better for captains in the Post-Duty condition during the cruise segment ($t = -1.99$, $p < .07$; and $t = -2.22$, $p < .04$, respectively). These differences appear to reflect better handling of the the vigilance measures programmed into this flight phase (icing conditions and moderate turbulence). Mean differences on overall performance for captains in this segment suggested a slight edge for Post-Duty captains, but this difference was not statistically significant.

Table 4. Observer Ratings of Captains on Crew Coordination During the Cruise Segment

Crew Coordination for Captains		
	Pre	Post
Mean	3.27	3.78
SD	(0.47)	(0.67)

Table 5. Observer Ratings of Captains on Procedures During the Cruise Segment

Procedures for Captains		
	Pre	Post
Mean	3.00	3.33
SD	(0.00)	(0.50)

Post-Duty first officers were also rated better on the coordination measure (Table 6) during the cruise segment ($t = -2.58$, $p < .02$). They were also rated as having performed better on the planning dimension (Table 7) in this segment ($t= -2.02$, $p < .06$).

Table 6. Observer Ratings of First Officers on Crew Coordination During the Cruise Segment

Crew Coordination for First Officers		
	Pre	Post
Mean	3.27	3.89
SD	(0.47)	(0.60)

Table 7. Observer Ratings of First Officers on Planning During the Cruise Segment

Planning for First Officers		
	Pre	Post
Mean	2.91	3.56
SD	(0.70)	(0.73)

For the approach segment into RIC, Post-Duty captains were rated better on the approach-planning measure ($t = -2.01$, $p < .06$), as can be seen in Table 8. The coordination rating (Table 9) was marginally significant, with Post-Duty captains again exhibiting better performance ($t = -1.85$, $p = .08$). First officers in the Post-Duty condition were also rated better on the planning measure ($t = -2.07$, $p < .06$), as portrayed in Table 10.

Table 8. Observer Ratings of Captains on Approach Planning Measure During the Approach to RIC Segment

Approach Planning for Captains		
	Pre	Post
Mean	1.91	3.11
SD	(0.94)	(1.69)

Table 9. Observer Ratings of Captains on Coordination Measure During the Approach to RIC Segment

Coordination for Captains		
	Pre	Post
Mean	3.36	4.00
SD	(0.67)	(0.87)

Table 10. Observer Ratings of First Officers on Planning Measures During the Approach to RIC Segment

Planning Measures for First Officers		
	Pre	Post
Mean	1.82	2.78
SD	(0.87)	(1.20)

For the missed-approach and emergency-procedure segment involving the System A hydraulic failure, Post-Duty captains were again rated better on planning and procedure measures ($t = -2.19$, $p < .05$; and $t = -2.10$, $p < .05$, respectively, Tables 11 and 12). The planning rating (Table 13) was also higher for first officers in the Post-Duty condition ($t = -2.32$, $p < .03$). Differences between groups in this flight phase on the planning and procedures measures are particularly significant because they were designed to tap performance during a critically high-workload period of the simulation scenario (dealing with the implications of the hydraulic failure).

Table 11. Observer Ratings of Captains on Planning Measures During the Missed-Approach and Emergency-Procedure Segment

Planning Measures for Captains		
	Pre	Post
Mean	2.91	3.89
SD	(1.14)	(0.78)

Table 12. Observer Ratings of Captains on Procedure Measures During the Missed-Approach and Emergency-Procedure Segment

Procedure Measures for Captains		
	Pre	Post
Mean	3.00	3.56
SD	(0.45)	(0.73)

Table 13. Observer Ratings of First Officers on Planning Measures During the Missed-Approach and Emergency-Procedure Segment

Planning Measures for First Officers		
	Pre	Post
Mean	2.55	3.56
SD	(0.93)	(1.01)

None of the other ratings for cruise-to-alternate and landing were statistically significant, although in several cases the means were in the same direction (higher ratings for crewmembers in the Post-Duty condition).

Taken as a whole, it is particularly significant that all of the reliable differences on this rating measure were in the same direction. This pattern strongly suggests that Post-Duty condition subjects performed better in several phases of flight. While there were many ratings, upon which no statistically significant differences manifested themselves, there was not a single case in which the pattern was reversed--better rated performance by Pre-Duty crewmembers.

*3.3.2. **Overall Ratings.*** None of the overall ratings assessed at the end of the simulation approached statistical significance.

*3.3.3. **Workload Ratings.*** On subject pilots' own subjective ratings of workload levels experienced in the simulated flights, captains in the Pre-Duty condition felt that they had exerted significantly more mental effort than did captains in the Post-Duty condition ($t = 2.16$, $p < .05$). The workload-rating measure also asked subjects to report how tired they were, and as previously discussed (section 3.2), the Post-Duty captains reported that they were significantly more tired. For first officers, only the fatigue measure was significant.

*3.3.4. **Aircraft Handling Data.*** Since the focus of the investigation was upon operational significance, analyses of aircraft handling data were confined to a particular flight segment in which these parameters were expected to be critically important. This segment involved the last few minutes of final approach to ROA where aircraft stability was of the utmost importance since speed, sink rate, and overall stability were expected to be strong predictors of the task of landing the aircraft at a higher-than-normal speed, with reduced braking effectiveness, on a short, wet runway. This was also the culmination of the scenario, where a number of high workload procedures (e.g. manual gear and flap operation) might have conspired to compromise normal aircraft handling performance. The manual control skills involved in aircraft handling were not considered as important at other flight phases (e.g., climb and cruise), because these segments were characterized by low workload (e.g., autopilot usage).

Four measures were used as indicants of stability during this segment; airspeed, vertical speed, localizer, and glideslope deviation. Absolute values were obtained for each measure during the last two min and thirty sec prior to touchdown at ROA--yielding 10 samples of each of the four parameters for all experimental runs. Because of computer problems, complete data were available for only 15 of the 20 experimental runs (9 in the Pre-Duty condition and 6 in the Post-Duty condition). In order to derive an overall index of aircraft stability and because these parameters are intercorrelated, the average for each of the four parameters was computed for each run, and these values were converted to z-scores. The z-scores for each of the parameters were summed yielding an overall index of aircraft stability during the final-approach segment.

Comparison of the stability index between experimental conditions revealed that Pre-Duty crews were significantly more unstable during this final-approach segment than were Post-Duty crews ($t = 2.35$, $p < .05$). Tables 14 and 15 portray the raw values for airspeed and vertical speed, and while these values were not used in the statistical analyses they help in understanding the nature of the effect.

Table 14. Raw Scores for Airspeed (kts) for Pre-Duty and Post-Duty Crews

Airspeed for Pre-Duty and Post-Duty Crews		
	Pre	Post
Mean	152.09	146.95
SD	(8.53)	(3.74)

Given the aircraft weight and 15 degree flap setting, the correct speed in the landing configuration was approximately 135 kts. Since there was a 10-kt. headwind, and since it is a widespread practice to add the extra speed, 145 kts. was the approximate target speed for final approach and landing. Table 14 shows that Post-Duty crews averaged very close to this value (146.94 kts.), whereas Pre-Duty crews were somewhat faster (152.09 kts.).

Table 15. Raw scores for Vertical Speed (ft/min) for Pre-Duty and Post-Duty Crews

Vertical Speed for Pre-Duty and Post-Duty Crews		
	Pre	Post
Mean	858.90	803.17
SD	(157.78)	(106.73)

Precise tracking of the Instrument Landing System (ILS) approach at ROA converts to a vertical descent rate of approximately 800 ft./min. Again, Post-Duty crews were very close to this value (803.17 ft./min.), while Pre-Duty crews averaged a higher vertical sink rate (858.89 ft./min.).

Pre-Duty crews also averaged higher amounts of localizer and glideslope deviation than Post-Duty crews. This corresponds to more horizontal and vertical deviation from the desired approach path to the runway.

3.3.5. Error Analyses. Table 16 summarizes the mean error frequencies for each condition and error type (Types I, II, III, and total errors). None of the differences between error categories or total errors were statistically significant between Pre- and Post-Duty crews. However, mean differences were in the same direction as previous results, particularly on Type III (operationally significant) and total errors. Mean Type III errors for Pre-Duty crews were 4.3 versus 2.33 errors for Post-Duty crews. Pre-Duty crews averaged 9.2 total errors versus 7.0 for Post-Duty crews.

Table 16. Error Frequencies for Pre-Duty and Post-Duty Crews on Type I, Type II, Type III, and Total Errors.

Type I Errors		
	Pre	Post
Mean	1.20	1.56
SD	(1.03)	(1.24)

Type II Errors		
	Pre	Post
Mean	3.70	3.11
SD	(3.56)	(1.54)

Type III Errors		
	Pre	Post
Mean	4.30	2.33
SD	(4.00)	(2.83)

Total Errors		
	Pre	Post
Mean	9.20	7.00
SD	(7.27)	(3.87)

In an effort to better understand this somewhat counterintuitive pattern of findings (tired crews apparently performing better than rested crews) internal analyses were performed. These analyses addressed the fact that Post-Duty crews had typically flown the entire trip together, whereas Pre-Duty crews were typically composed of individuals who may not have flown together recently.

This phenomenon is probably representative of actual operational practice. At the end of a trip, a pilot is more aware of the capabilities and tendencies of other crewmembers than at the beginning of a trip. It was felt that this "crew familiarity" factor may have had some impact on the results.

The reanalyses addressed the familiarity factor. One of the Post-Duty crews did not fly their last trip together prior to simulator evaluation, while two of the Pre-Duty crews had flown together the last time on duty. Thus, all of the data were reanalyzed based on who had flown together the last time on duty. The Pre- versus Post-Duty crew assignment was discarded for the purpose of this analysis. It is important to note that no attempt was made to partition subjects according to whether they knew each other or had ever flown together in the past. In fact, several crews in the "Not Flown Together" condition had flown together at some point in the past. However, the analysis only addressed whether the crewmembers had flown together on the last duty cycle.

The results of one of these reanalyses are presented in Table 17. As can be seen, there were several significant differences apparently attributable to this crew-familiarity factor. The difference between Type I (minor) errors was not significant, however for Type II (moderate) errors, crews that had not flown together averaged significantly more errors (4.78) than did crews that had flown together (2.20) on the last duty cycle ($t = 2.20$, $p < .04$). The same pattern was evident for Type III (major) and total errors. Crews that had not flown together averaged 5.67 Type III errors versus only 1.30 for crews that had flown together, and this difference is highly significant ($t = 3.36$, $p < .004$). There was also a strongly significant difference between these groups on total errors ($t = 2.96$, $p < .009$). Crews that had not flown together averaged 11.67 total errors whereas crews that had flown together averaged less than half (5.0) of this error total.

Table 17. Error Frequencies for Crews that had Flown Together and Not Flown Together on Type I, Type II, Type III, and Total Errors.

Type I Errors		
	Flown	Not Flown
Mean	1.50	1.22
SD	(1.27)	(0.97)

Type II Errors		
	Flown	Not Flown
Mean	2.20	4.78
SD	(1.55)	(3.19)

Type III Errors		
	Flown	Not Flown
Mean	1.30	5.67
SD	(1.34)	(3.87)

Total Errors		
	Flown	Not Flown
Mean	5.00	11.67
SD	(2.58)	(6.60)

In summary, significantly better performance by Post-Duty crewmembers was suggested by all of the various types of crew-performance measures. Many individual parameters were non-significant, but there were no reversals of this general pattern. It appears that much of this performance difference is due to the fact that Post-Duty crews, since they were tested at the end of their duty cycle, were more likely to have operated together.

3.4. Crew Communication Analyses

Extensive analyses of the crew communications process were undertaken since these variables represent perhaps the best reflection of how the crew coordinates its activities. Therefore, it was expected that these communications measures would facilitate the understanding of the performance effects, as has been suggested in the past (e.g., Foushee, 1984).

Two types of analyses were performed. First, a 2 x 2 (Pre- vs. Post-Duty x captain vs. first officer) between-subjects analysis-of-variance (ANOVA) was performed for each category as well as for total communication. The second type of analysis was designed to look at communication variables as they were affected at different phases of flight. It involved a 2 x 2 x 3 (Pre- vs. Post-Duty x captain vs. first officer x phase of flight) mixed-design ANOVA, also performed for each category. The three-factor phase-of-flight parameter was a within-subjects variable, and was broken down in the following manner: 1) the 10-min period immediately after rotation that was completely routine; 2) the 10-min period beginning after the decision to execute a missed approach; and 3) the 10-min period immediately prior to touchdown at ROA. Thus, one relatively low-workload period and two relatively high-workload periods were included in these analyses.

Since the familiarity variable appeared to be strongly related to crew performance in this study, the same analyses were conducted incorporating this factor. Both the 2 x 2 between subjects ANOVAs and the 2 x 2 x 3 mixed-design

ANOVAs were identical except that the Flown Together-Not Flown Together variable was substituted for the Pre-Post Duty variable.

3.4.1. Commands. The 2 x 2 x 3 ANOVA for commands revealed a significant main effect for the Pre-Post variable ($F(1.32) = 4.07$, $p = .05$), indicating that in general Post-Duty crewmembers exchanged more commands, and this was true of both captains and first officers in this condition. Not surprisingly, captains utilized this form of communication more often than did first officers ($F(1,32) = 107.34$, $p < .001$). It has been suggested elsewhere (e.g., Foushee & Manos, 1981) that commands appear to have a coordinating effect on crew performance because of their strong influence on subordinate crewmember actions. Commands were much more predominant during high workload phases of flight ($F(2,64) = 37$, $p < .01$). It is also interesting to note that first officers who had recent operating experience with the captain they were flying with were more likely to use command-type statements. Increased familiarity may raise the probability that subordinate crewmembers will be more assertive when the circumstances call for such behavior. These results are summarized in Table 18.

Table 18. Means and Standard Deviations (SDs) of Commands for Captains and First Officers in Pre-Duty and Post-Duty Crews for Three Phases of Flight - 10-min After Rotation (ROT), 10-min After Missed-Approach (MA), and 10-min Prior to Touchdown (TD)

Means and SDs of Commands				
	Pre- Capt	Pre- F/O	Post- Capt	Post- F/O
Mean (ROT)	10.00	0.00	15.56	0.22
SD	(2.78)	(0.00)	(6.56)	(0.44)
Mean (MA)	13.33	0.44	19.44	0.11
SD	(4.74)	(0.73)	(10.31)	(0.33)
Mean (TD)	12.89	0.56	17.44	1.22
SD	(3.82)	(0.88)	(7.70)	(1.56)

The same pattern was evident for commands on the 2 x 2 ANOVAs and on the 2 x 2 x3 ANOVAs on the familiarity factor. In short, performance appeared to be facilitated by the more prevalent usage of commands by captains in the Post-Duty condition, particularly during high workload phases of flight.

3.4.2. Observations. The analyses for the variable, observations about flight status, revealed a significant main effect for crew position ($F(1,32) = 14.84$, $p < .001$). This effect was due to the fact that first officers utilize this category of communication more frequently than captains as can be seen in Table 19. This is logical in light of the support role assigned to first officers in flight duties, since observations about flight status are a primary means of providing information for the captain to act upon. The main effect for phase-of-flight was significant ($F(2,64) = 13.23$, $p < .001$) indicating more observations during high-workload periods. The interaction of crew position and flight segment was also significant ($F(2,64) = 3.23$, $p < .05$) and is due to the more prevalent use of this category of communications by first officers during high-workload phases.

Table 19. Means and SDs of Observations for Captains and First Officers in Pre-Duty and Post-Duty Crews for Three Phases of Flight - 10-min After Rotation (ROT), 10-min After Missed Ppproach (MA), and 10-min Prior to Touchdown (TD)

Means and SDs of Observations				
	Pre- Capt	Pre- F/O	Post- Capt	Post- F/O
Mean (ROT)	12.89	20.33	19.11	22.33
SD	(7.18)	(7.97)	(13.46)	(4.64)
Mean (MA)	12.11	23.33	17.33	23.56
SD	(6.41)	(8.76)	(9.08)	(10.88)
Mean (TD)	19.78	29.67	18.11	33.44
SD	(5.65)	(7.02)	(5.90)	(10.89)

3.4.3. Suggestions. For the category, suggestions, the main effect for crew position was again significant ($F(1,32) = 26.97$, $p < .001$), with captains responsible for more suggestions than first officers (Table 20). This is likely reflective of the captain's role in directing subordinate behavior, as suggestions are probably a "softer" means of providing directions than commands. The crew position by crew familiarity interaction approached statistical significance ($F(1,32)$, $p < .09$). Captains who had flown with the same first officer tended to offer more suggestions than those who had not (Table 21). This may imply a somewhat less "directive" style among captains who are relatively familiar with other crewmembers on the flightdeck, since suggestions tend to be less directive.

Table 20. Means and SDs of Suggestions for Captains and First Officers in Pre-Duty and Post-Duty Crews for Three Phases of Flight - 10-min After Rotation (ROT), 10-min After Missed-Approach (MA), and 10-min Prior to Touchdown (TD)

Means and SDs of Suggestions				
	Pre- Capt	Pre- F/O	Post- Capt	Post- F/O
Mean (ROT)	2.78	0.89	3.56	0.89
SD	(3.07)	(0.93)	(3.24)	(0.78)
Mean (MA)	3.67	0.56	4.33	1.22
SD	(2.65)	(0.73)	(2.12)	(1.99)
Mean (TD)	4.67	1.56	3.89	1.56
SD	(2.35)	(1.74)	(2.26)	(1.42)

Table 21. Means and SDs of Suggestions for Captains and First Officers in Flown Together (Ft) and Not Flown Together (Nf) Crews for Three Phases of Flight - 10-min After Rotation (ROT), 10-min After Missed-Approach (MA), and 10-min Prior to Touchdown (TD)

Means and SDs of Suggestions				
	Ft Capt	Ft F/O	Nf Capt	Nf F/O
Mean (ROT)	4.30	0.90	1.75	0.88
SD	(3.65)	(0.88)	(1.39)	(0.83)
Mean (MA)	4.80	0.90	3.00	0.88
SD	(2.04)	(1.91)	(2.45)	(0.83)
Mean (TD)	4.50	1.40	4.00	1.75
SD	(2.37)	(1.43)	(2.27)	(1.75)

3.4.4. Statements of Intent. This was another category which was assumed to reflect the amount of overall coordination. These communications are generally utilized to inform others of the actions that the speaker is about to undertake, and thus keep other crewmembers informed. Again, the main effect for crew position was significant ($F(1,34) = 9.6$, $p < .004$). First officers exhibited this form of communication more frequently than captains, but the crew-familiarity main effect was also marginally significant ($F(1,34) = 3.58$, $p < .07$). Statements of intent were relatively more prevalent among crewmembers who had flown together, as Table 22 portrays. This suggests one reason for coordination deficiencies that were apparent in the Pre-Duty or Not Flown Together conditions and may be in part responsible for the performance differences seen on previous measures.

Table 22. Means and SDs of Statements of Intent for Captains and First Officers in Flown Together and Not Flown Together Crews

Means and SDs of Statements of Intent		
	Flown	Not Flown
Capt Mean	8.90	3.67
SD	(6.40)	(2.12)
F/O Mean	14.20	11.44
SD	(7.24)	(8.35)

3.4.5. Inquiries. These are information-seeking behaviors designed to elicit assistance from other crewmembers. Mean differences can be seen in Tables 23 and 24. Captains sought more information than first officers ($F(1,34 = 3.87$, $p < .06$), but this type of information-seeking behavior was far more prevalent during high workload phases of flight ($F(1,32) = 9.81$, $p < .001$). Neither the fatigue variable, nor the crew familiarity variable predicted any of the differences on this measure.

Table 23. Means and SDs of Inquiries for Captains and First Officers in Pre-Duty and Post-Duty Crews

Means and SDs of Inquiries		
	Pre-Duty	Post-Duty
Capt Mean	30.80	36.44
SD	(9.30)	(20.18)
F/O Mean	27.20	22.56
SD	(12.03)	(11.37)

Table 24. Means and SDs of Inquiries for Captains and First Officers in Flown Together (Ft) and Not Flown Together (Nf) Crews for Three Phases of Flight - 10-min After Rotation (ROT), 10-min After Missed-Approach (MA), and 10-min Prior to Touchdown (TD)

Means and SDs of Inquiries				
	Ft Capt	Ft F/O	Nf Capt	Nf F/O
Mean (ROT)	9.60	5.80	8.13	6.50
SD	(4.55)	(3.94)	(3.94)	(2.45)
Mean (MA)	13.10	7.40	11.25	11.13
SD	(9.69)	(4.93)	(4.71)	(5.38)
Mean (TD)	10.40	6.60	11.00	9.88
SD	(4.74)	(2.72)	(5.24)	(3.72)

*3.4.6. **Agreement**.* No differences on the agreement variable were statistically reliable, but it should be noted that agreement was an infrequently occurring category.

*3.4.7. **Disagreement**.* On instances of verbal communication reflecting the disagreement of one crewmember with the actions, intended actions, or statements of another, significant two-way interactions were obtained with the crew position variable on both the familiarity and the fatigue variables, and the mean differences for this category can be seen in Tables 25 and 26. In both cases, first officers were largely responsible for this effect. First officers in the Post-Duty condition were far more likely to disagree with the actions of captains ($F(1,34) = 6.20$, $p < .02$). The same was true for first officers who had flown with the same captain previously, only the effect was stronger ($F(1,34) = 11.37$, $p < .002$). It has been suggested that first officers, because of the role structure of the flightdeck, are often hesitant to question or correct the actions of captains and that this reluctance has been a factor in a substantial number of incidents and accidents (e.g. Cooper, White, & Lauber, 1979; Foushee & Manos, 1981; and Foushee, 1984). This result suggests that crewmember familiarity may mediate against this hesitancy and raises the probability that familiar first officers or subordinates will be more assertive when the circumstances call for such behavior.

Table 25. Means and SDs of Disagreements for Captains and First Officers in Pre-Duty and Post-Duty Crews

Means and SDs of Disagreements		
	Pre-Duty	Post-Duty
Capt Mean	1.70	0.78
SD	(1.70)	(0.67)
F/O Mean	0.60	2.00
SD	(0.70)	(2.12)

Table 26. Means and SDs of Disagreements for Captains and First Officers in Flown Together and Not Flown Together Crews

Means and SDs of Disagreements		
	Flown	Not Flown
Capt Mean	0.70	1.89
SD	(0.67)	(1.69)
F/O Mean	2.10	0.33
SD	(1.91)	(0.50)

3.4.8. Acknowledgements. Past research has demonstrated that acknowledgements to other communications are often associated with fewer crew-performance errors, and that these categories of communication tend to reinforce the interaction process (e.g. Foushee & Manos, 1981). The same was true in the present investigation. Acknowledgements were significantly more prevalent in crews that had flown together ($F(1,34) = 8.33$, $p < .007$), as Table 27 suggests. Acknowledgements were also seen more frequently in high workload segments of flight ($F(2,64) = 3.35$, $p < .05$), suggesting that they play an even more important role in the communications process during critical phases of flight (Table 28).

Table 27. Means and SDs of Acknowledgements for Captains and First Officers in Flown Together and Not Flown Together Crews

Means and SDs of Acknowledgements		
	Flown	Not Flown
Capt Mean	46.70	33.89
SD	(12.86)	(17.45)
F/O Mean	59.30	41.33
SD	(20.52)	(13.44)

Table 28. Means and SDs of Acknowledgements for Captains and First Officers in Flown Together (Ft) and Not Flown Together (Nf) Crews for Three Phases of Flight - 10-min After Rotation (ROT), 10-min After Missed- Approach (MA), and 10-min Prior to Touchdown (TD)

Means and SDs of Acknowledgements				
	Ft Capt	Ft F/O	Nf Capt	Nf F/O
Mean (ROT)	11.90	16.10	10.25	10.25
SD	(4.28)	(8.97)	(4.20)	(5.01)
Mean (MA)	15.00	16.60	9.25	10.63
SD	(6.24)	(7.28)	(6.94)	(4.14)
Mean (TD)	15.50	17.60	11.00	13.50
SD	(6.65)	(5.48)	(6.23)	(5.18)

3.4.9. Answer Supplying Information. Responses to requests for information were more prevalent among first officers (F(1,34) = 3.99, p < .06). This is not surprising since by definition, these behaviors are usually responses to commands, inquiries or observations that are more likely to come from the captain. First officers in the Post-Duty condition were more likely to exhibit this type of behavior (F(1,34) = 3.84, p < .06), as were first officers who had recent operating experience with their captains (F(1,34) = 3.38, p < .07). These results are summarized in Tables 29 and 30, and again imply that more overall information exchange occurred in crews with recent operating experience together.

Table 29. Means and SDs of Answers Supplying Information for Captains and First Officers in Pre-Duty and Post-Duty Crews

Means and SDs of Answers Supplying Information		
	Pre-Duty	Post-Duty
Capt Mean	17.80	11.44
SD	(6.43)	(4.16)
F/O Mean	17.90	21.67
SD	(7.75)	(11.72)

Table 30. Means and SDs of Answers Supplying Information for Captains and First Officers in Flown Together and Not Flown Together Crews

Means and SDs of Answers Supplying Information		
	Flown	Not Flown
Capt Mean	12.40	17.33
SD	(4.30)	(7.35)
F/O Mean	21.90	17.22
SD	(10.34)	(8.94)

3.4.10. Response Uncertainty. No differences were evident as a function of any of the independent variables on this measure. There was a marginal tendency for more of this type of behavior in high workload flight phases, although it was not statistically significant ($F(1,32) = 2.43$, $p < .10$). Response uncertainty was infrequently verbalized by crewmembers, but communication coders anecdotally reported non-verbal indications. Such data were not systematically obtained because of their inherent unreliability.

3.4.11. Tension Release. This category was operationalized as a reflection of non-task-related behavior and typically consisted of laughter or humorous remarks. A significant main effect for the crew-familiarity variable was evident on this measure ($F(1,34) = 4.14$, $p < .05$), as can be seen in Table 31. There was significantly more tension release among crewmembers who had not flown together prior to the simulator sessions. This difference may be a reflection of the acquaintance process for crewmembers unfamiliar with each other. However, as can be seen in Table 32, this difference was primarily evident for the low workload segment and diminished significantly during the high workload flight segments ($F(2,64) = 8.0$, $p < .002$).

Table 31. Means and SDs of Tension Releases for Captains and First Officers in Flown Together and Not Flown Together Crews

Means and SDs of Tension Releases		
	Flown	Not Flown
Capt Mean	3.40	10.00
SD	(3.89)	(11.55)
F/O Mean	5.20	8.89
SD	(6.97)	(7.24)

Table 32. Means and SDs of Tension Releases for Captains and First Officers in Flown Together (Ft) and Not Flown Together (Nf) Crews for Three Phases of Flight - 10-min After Rotation (ROT), 10-min After Missed- Approach (MA), and 10-min Prior to Touchdown (TD)

Means and SDs of Tension Releases				
	Ft Capt	Ft F/O	Nf Capt	Nf F/O
Mean (ROT)	1.90	2.60	4.88	5.00
SD	(2.64)	(4.12)	(6.64)	(4.38)
Mean (MA)	0.80	1.10	2.63	1.50
SD	(1.40)	(1.20)	(4.17)	(3.46)
Mean (TD)	0.80	1.30	1.75	2.00
SD	(1.14)	(1.16)	(2.05)	(1.85)

3.4.12. Frustration/Anger. Captains exhibited considerably more of this type of behavior than did first officers (Table 33) regardless of experimental condition ($F(1,34) = 11.58$, $p < .002$). Phase of flight was also a significant predictor as might have been expected ($F(2,64) = 3.76$, $p < .03$). Table 34 suggests that this difference is attributable to the fact that more frustration occurred during the high workload phases of flight. The fact that captains are more prone to this type of behavior is no doubt strongly tied to the captain's authority role.

Table 33. Means and SDs of Frustration for Captains and First Officers in Flown Together and Not Flown Together Crews

Means and SDs of Frustration		
	Flown	Not Flown
Capt Mean	4.50	5.89
SD	(3.87)	(7.20)
F/O Mean	0.10	1.22
SD	(0.31)	(1.64)

Table 34. Means and SDs of Frustration/Anger for Captains and First Officers in Flown Together (Ft) and Not Flown Together (Nf) Crews for Three Phases of Flight - 10-min After Rotation (ROT), 10-min After Missed Approach (MA), and 10-min Prior to Touchdown (TD)

Means and SDs of Frustration/Anger				
	Ft Capt	Ft F/O	Nf Capt	Nf F/O
Mean (ROT)	0.90	0.00	1.13	0.00
SD	(1.85)	(0.00)	(1.13)	(0.00)
Mean (MA)	1.70	0.00	2.25	0.13
SD	(1.89)	(0.00)	(4.10)	(0.35)
Mean (TD)	1.60	0.10	2.63	1.13
SD	(1.51)	(0.32)	(4.14)	(1.73)

3.4.13. Embarrassment. This type of behavior typically consisted of apologetic remarks as a result of mistakes or oversights on the part of one crewmember. None of the main effects for any of the experimental variables were significant, but a marginally significant two-way interaction between the crew familiarity variable and phase of flight was found ($F(2,64) = 3.12$, $p < .06$). This type of behavior was more often seen among crewmembers who had not previously flown together during the high workload phases of flight, as can be seen in Table 35. This finding appears consistent with the performance profiles of crewmembers in this condition. Crewmembers who had not flown together made significantly more serious errors.

Table 35. Means and SDs of Embarrassment for Captains and First Officers in Flown Together (Ft) and Not Flown Together (Nf) Crews for Three Phases of Flight - 10-min After Rotation (ROT), 10-min After Missed- Approach (MA), and 10-min Prior to Touchdown (TD)

Means and SDs of Embarrassment				
	Ft Capt	Ft F/O	Nf Capt	Nf F/O
Mean (ROT)	0.50	0.10	0.13	0.38
SD	(1.27)	(0.32)	(0.35)	(0.52)
Mean (MA)	0.20	0.30	0.13	0.13
SD	(0.42)	(0.48)	(0.35)	(0.35)
Mean (TD)	0.00	0.20	0.63	0.50
SD	(0.00)	(0.42)	(0.74)	(0.53)

3.4.14. Non-Task-Related Communication. This category included crew interaction that was clearly not related to flight tasks. On this measure, the main effect for crew familiarity was significant ($F(1,32) = 7.29$, $p < .02$). Crews that had not flown together engaged in more non-tasked-related interaction than crews that had flown together, which may be related to the fact that they were probably becoming acquainted. This was more apparent during low workload periods of flight since the main effect for flight phase was also strongly significant ($F(2,64) = 7.18$, $p < .002$). The two-way interaction between phase of flight and crew familiarity was also significant ($F(2,64) = 4.26$, $p < .02$), indicating that non-task-related interaction was far more prevalent among crewmembers that had not flown together during low workload periods. The means for these analyses are shown in Table 36.

Table 36. Means and SDs of Non-Task-Related Communication for Captains and First Officers in Flown together (Ft) and Not Flown Together (Nf) Crews for Three Phases of Flight - 10-min After Rotation (ROT), 10-min After Missed-Approach (MA), and 10-min Prior to Touchdown (TD)

Means and SDs of Non-Task-Related Communication				
	Ft Capt	Ft F/O	Nf Capt	Nf F/O
Mean (ROT)	0.40	0.20	2.00	2.13
SD	(0.70)	(0.42)	(3.21)	(3.48)
Mean (MA)	0.20	0.00	0.13	0.13
SD	(0.63)	(0.00)	(0.35)	(0.35)
Mean (TD)	0.00	0.00	0.25	0.25
SD	(0.00)	(0.00)	(0.46)	(0.71)

3.4.15. Repetitions. This category was intended to reflect instructions from one crewmember to another that were repeated in quick succession. Communications of this type are typically used to convey a sense of urgency or to assure that an instruction has been received by the crewmember to whom it was addressed. None of the main effects were significant, but the interaction between the crew position and the fatigue variables was statistically significant ($F(1,34) = 4.59$, $p < .04$), as can be seen in Table 37. Post-Duty captains repeated instructions more often than either captains in the Pre-Duty condition or first officers in either condition. This finding may explain, in part, the apparently better coordination among Post-Duty crewmembers who were previously acquainted, since repetitions may have assured that critical pieces of information were transferred between crewmembers at appropriate times. The 2 x 2 x 3 ANOVAs were not performed on this measure because of insufficient instances of this behavior across all of the flight phases.

Table 37. Means and SDs of Repetitions for Captains and First Officers in Pre-Duty and Post-Duty Crews

Means and SDs of Repetitions		
	Pre-Duty	Post-Duty
Capt Mean	1.00	2.44
SD	(0.67)	(1.51)
F/O Mean	1.80	1.56
SD	(1.23)	(1.33)

3.4.16. Checklist Items. These communications were merely the challenges issued on standard procedural checklists required at various flight phases. No effects for the fatigue or crew familiarity variables were evident, however the main effect for crew position was significant ($F(1,34) = 30.35$, $p < .001$). First officers were responsible for most of these communications (Table 38), which is entirely logical since they were assigned non-flying pilot duties for the simulation.

Table 38. Means and SDs of Checklist Items for Captains and First Officers in Pre-Duty and Post-Duty Crews

Means and SDs of Checklist Items		
	Pre-Duty	Post-Duty
Capt Mean	7.60	7.44
SD	(1.96)	(2.65)
F/O Mean	12.70	13.44
SD	(2.79)	(4.56)

3.4.17. Air Traffic Control Communications. Once again, the main effect for crew position was significant ($F(1,34) = 256.93$, $p < .001$), with first officers almost entirely responsible for ATC communication (Table 39). As with checklist duties, ATC communications are almost entirely the responsibility of the non-flying pilot.

Table 39. Means and SDs of ATC Communications for Captains and First Officers in Pre-Duty and Post-Duty Crews

Means and SDs of ATC Communications		
	Pre-Duty	Post-Duty
Capt Mean	7.50	8.89
SD	(8.48)	(10.91)
F/O Mean	75.60	80.22
SD	(10.63)	(20.85)

3.4.18. Total Communication. It was expected that total communication, or the sum of all types of communication including non-codable verbal behavior throughout the simulated flight, would be related to overall performance (e.g. Foushee & Manos, 1981). This was suggested in the present investigation as the main effect for the crew familiarity variable was marginally significant ($F(1,34) = 3.55$, $p < .07$). Overall, communication was more frequent in crews that had flown together than in crews that had not, as Table 40 reveals. This is particularly interesting in light of the significant performance differences between these groups. First officers exhibited more overall communication than captains ($F(1,34) = 4.72$, $p < .04$), and this is most likely due to the fact that first officers, in their non-flying role, were more involved in supplying task-relevant information for captains' use in decision-making.

Table 40. Means and SDs of Total Communication for Captains and First Officers in Flown Together and Not Flown Together Crews

Means and SDs of Total Communication		
	Flown	Not Flown
Capt Mean	335.40	285.89
SD	(79.03)	(53.22)
F/O Mean	369.90	341.33
SD	(59.42)	(58.60)

The 2 x 2 x 3 ANOVAs were not performed because the three-level flight phase variable did not encompass the entire time period involved in the simulated flight. Thus, this analysis was not an accurate representation of total communications in the simulated flight.

3.4.19. Communications Summary. In general, the communications variables, as measures of group interaction and coordination, reflected the same trends evident for the crew performance measures. Post-Duty crews and crews that had flown together engaged in more task-related communication and less non-task related communication. As expected, instances of various communications behaviors increased with increasing task demands. These analyses appeared to support the conclusion that the performance differences seen in this study were in large part caused by differences in crew coordination. This conclusion is based on the assumption that crew communication patterns are at least partial reflections of the coordination process, since they are the means by which many individual efforts are coordinated. Crews that had flown together seemed better able to coordinate their activities than crews that had not flown together.

DISCUSSION

4.1. Overview of Findings

The issue of excessive flightcrew fatigue, as a result of trip exposure, has been a primary concern of the aviation community for a long time, but there has been little tangible evidence with which to confirm or deny the extent or operational significance of fatigue associated with duty-cycle exposure. We have discussed how laboratory studies have been of little use to those interested in aviation safety because of the difficulty of generalizing laboratory performance measures to the task of operating a complex aircraft. Thus, the operational significance issue was a pivotal part of this investigation.

This study examined the performance of 20 volunteer twin-jet transport crews in a full-mission simulator scenario that included many aspects of an actual line operation. The scenario involved both routine flight operations and an unexpected hydraulic failure complicated by weather problems that resulted in a high level of crew workload. Approximately half of the crews flew the simulation within two to three hr after completing a three-day, high-density, short-haul duty cycle. The other half flew the scenario after an average of three days off duty. The high-density duty cycles that were the focus of this investigation averaged eight hr of on-duty time per day and five takeoffs and landings, with at least one day (usually the last) averaging close to thirteen hr of duty and eight takeoffs and landings. These figures do not include the time associated with flying the simulated flight, and if these numbers are included, the last duty day for Post-Duty crews approached sixteen hr and nine takeoffs and landings.

The results of the study revealed that, as expected, Post-Duty crews were significantly more "fatigued" than Pre-Duty crews. The former averaged less sleep and reported higher levels of fatigue than the latter. However, results on the crew performance measures indicated that this level of fatigue did not affect the performance of flightcrews in any operationally significant manner. As has been shown, the performance of Post-Duty crews was actually better than the performance of Pre-Duty crews on a number of dimensions relevant to flight safety. Post-Duty crews were rated as performing better by an expert observer on many significant dimensions, and although there were many measures that did not discriminate between the two groups, there were no cases where the performance of Pre-Duty crews was rated as superior. Post-Duty crews flew more stable approaches, and they tended to make fewer significant operational errors than did Pre-Duty crews.

To some, this very consistent pattern of results may seem paradoxical. However, it is important to note when considering how crews are usually assigned to flight duties, that there is a substantive difference between crews at the beginning of the duty cycle and crews at the end of a trip, regardless of the fatigue factor. After three days of flying with another crewmember, one knows a considerable amount about his or her operating characteristics, personality, and communication style. For example, copilots learn when and how an aircraft commander or captain likes to be assisted. Captains become familiar with the tendencies of their subordinates--how they supply information, and how best to elicit their input. Obviously, there is wide variation in human interaction, and the more individuals learn about their coworkers, the better they are able to tailor their behavior to the needs of a particular interaction.

In an effort to control for this familiarity factor, some crews were assigned to conditions differentially. In some cases, Post-Duty or tired crewmembers from *different* trips were assigned as a simulation crew, so they had not necessarily flown together recently. Likewise, Pre-Duty or rested crews were assigned to the simulation from the ranks of individuals who had just finished a trip together, but had been off-duty for three days. All of the data were then reanalyzed based on who had flown together on the most recent duty cycle or not (independent of the fatigue factor), and a very striking pattern of results emerged--the performance differences became stronger. It is readily apparent that crews in which the two pilots had flown together on the preceding duty cycle made significantly fewer errors than crews who had not, particularly the more serious types of errors--the Type II and Type III errors. This same pattern was evident for all of the other measures in these reanalyses, and they too exhibited larger performance differences. Recent operating experience appears to be a strong influence on crew performance and may have served as a countermeasure to the levels of fatigue present in Post-Duty crewmembers.

Examination of the flightcrew communication patterns in this study, as manifestations of the crew coordination process, suggests that this dimension is at least partially responsible for the performance differences. Crews that had flown together communicated significantly more overall, and these differences were in logical directions when compared with the significant performance variations. As in the Foushee and Manos (1981) study which found commands associated with better performance, captains in crews that had flown together issued more commands, but so did copilots (even though the frequency of copilot commands was relatively low). This finding may reflect a better understanding and division of responsibility between familiar crewmembers. There were more suggestions made in crews that had flown together, and more statements of intent by each crewmember, also indicating more willingness to exchange information.

Another replication of the Foushee and Manos findings revealed acknowledgements associated with better performance. There were many more acknowledgements of communications by both captains and first officers who had flown together. Foushee and Manos suggested that acknowledgements serve to reinforce the communications process, and the same phenomenon appears to have played a role in this study. It is particularly interesting that more disagreement was exhibited by first officers who had flown with the same captain during the preceding three days. This suggests that increased familiarity may be a partial cure for the frequently problematic hesitancy of subordinates to question the actions of captains.

There was significantly more non-task-related communication in crews that had not flown together, which may well indicate that they spent more time attempting to get to know each other. There was also significantly more tension release among crews who had not previously flown together. Also interesting was the presence of more frustration among captains who had not flown with the same copilot during the preceding duty cycle.

4.2. Operational Significance

It is the consistency of these results that is particularly striking. Duty-cycle exposure had no apparent effect on any of the parameters associated with flight safety. It is also interesting to note that the positive effects on crew coordination of some unknown amount of recent operating experience can be an effective countermeasure to the levels of fatigue associated with the duty cycles examined in this study. Whereas fatigue tends to be more prevalent during the later stages of a given duty cycle, crew coordination may be better as well because of the increased familiarity of crewmembers.

One of the obvious limitations of this study is that we are unable to closely examine the interaction of fatigue and crew familiarity. For example, it would be enlightening to look at the performance of tired crews who are familiar with each other versus those who are not and to repeat these comparisons for rested crews. Such an analysis could yield important insights into the effectiveness of crew familiarity as a countermeasure. Unfortunately, the sample only included one "fatigued" crews that had not flown together, while only two of the "rested" crews had flown together, and the restricted amount of data precluded these analyses.

Another limitation is related to the fact that we cannot determine from these data the amount or degree of familiarity necessary to produce a desired level of crew coordination. We can say with a fair amount of confidence that the recency of crew familiarity seems to be the key component, rather than the absolute amount. Most of the crews that had not flown together (as operationalized in these analyses) did know each other, and had flown together at some point in the past, but not within the last two or three months. Despite these rather compelling results, it would be a mistake to suggest a policy establishing the creation of relatively permanent crew assignments based on these data. There may be negative aspects associated with flying with the same person over a long period of time, such as complacency, boredom, and so on. It could well be that continued pairing of the same individuals would ultimately lead to a reversal of this pattern--worse performance associated with increased familiarity. Unfortunately, no research presently exists to substantiate this possibility (although the operational community generally believes this to be true), and it is not known how much familiarity might lead to this reversal.

It is interesting to speculate about the differences between this study and other research efforts that have demonstrated performance deficits associated with fatigue. It has been suggested (see Holding, 1983, for a review) that fatigue causes more minor types of errors because of its deleterious effect on the attentional process. This line of reasoning suggests that fatigue lowers attentional capacity, and the cumulative effect of lower attentional capacity coupled with low motivation on "less exciting" tasks tends to produce more error. It is arguable that since these non-engaging types of tasks tend to be of limited significance, any apparent performance deficits might be characterized minor, for the most part. However, this study provided no real support for this notion. There was a slightly larger number of Type I (minor) errors committed by Post-Duty crews, but the difference was not statistically reliable. Nevertheless, this remains a credible hypothesis considering the fact that traditional studies of fatigue effects have utilized performance measures that are not necessarily operationally significant.

The performance environment in this investigation was different from that found in traditional studies, and the high levels of realism and workload associated with segments of the simulation scenario no doubt produced average arousal levels greater than those typically found in lower fidelity studies measuring psychomotor effects. The high levels of crew workload associated with the operational problems faced by crews in this study demanded close attention, produced high motivation, and probably reduced boredom to a minimum. Since the subjects in this investigation were highly skilled, professional pilots performing identical tasks to those they perform in the real world, it is safe to assume that the motivation to perform well was quite high. This is in striking contrast to the boredom and lack of motivation often associated with classic, psychomotor measures such as reaction times. Thus it appears that arousal may be a key moderator of fatigue effects. Optimal levels of arousal appear to be harder to maintain when the operator is fatigued, but task demands may override this difficulty to some unknown extent. However, when task demands are low, the effects of fatigue may be manifested in performance deterioration more often. There is fairly convincing evidence that monitoring and vigilance during boring tasks is substantially degraded when the operator is fatigued (e.g., Holding, 1983).

If arousal does prove to be a key moderator of fatigue effects, it poses something of a complex puzzle for researchers interested in the implications of the fatigue-performance relationship for flight safety. On one hand, during low workload segments of flight, we might expect the effects of fatigue to be apparent, since low task demands produce low arousal. Thus, we can expect reduced performance when fatigue reaches some significant level, but since task demands are low (and the effects of performance decrements likely to be minor) it is reasonable to suggest that the effects of fatigue often may not be operationally significant. On the other hand, when task demands are high and if arousal does effectively counteract the effect of duty cycle exposure, then performance may not be affected (as in the present study). Periods of high task demand are precisely the times when good performance is most important, and performance parameters during these periods are usually the primary concern of aviation safety specialists. Again, one is drawn to the conclusion that fatigue may be present, but that its effects are not necessarily operationally significant. The problem with this line of reasoning is that occasionally minor attentional or performance lapses during periods of low task demand *can* precipitate a sequence of events leading to a serious incident or accident, as we have seen in the past. There was no evidence that fatigue produced such a sequence in this study, but the crew coordination process did appear to play a key role in eliminating the progression of a minor errors into major problems. As we have sought to demonstrate, coordination appeared to be better in short-haul crews at the end of the duty-cycle when they were presumably experiencing the highest cumulative effect of fatigue. These results clearly suggest that the system contains a "built-in" countermeasure, of sorts.

4.3. Implications for Long-Haul Operations

Despite the fact that these results are probably representative of the typical short-haul operation, it is not known whether the same phenomenon will be prevalent in long-haul, transmeridian operations or at higher levels of fatigue. While it is clear that the levels of fatigue associated with these short-haul duty cycles produced numerous psychological and physiological effects (see Gander et al., 1986), it is possible that these levels were not great enough to cause severe performance difficulties, as these results suggest. However, at some point, the level of fatigue or circadian dysrhythmia may well subsume the compensatory advantages of arousal or a well-coordinated crew operation. Long-haul flight operations, with longer duty days and other complications associated with time-zone shifts, may well produce more drastic effects. Another feature that distinguishes long- from short-haul operations is the predominance of extended stretches of low-workload cruise segments, in which arousal levels are no doubt lower, over longer periods of time, than in short-haul operations.

Moreover, too much crew familiarity, as in some long-haul operations where duty cycles can last 10 days or more, may lead to levels of complacency at which adverse effects on performance begin to manifest themselves. Such a phenomenon is particularly characteristic of groups in which too much cohesion or trust has developed, because of a reduced tendency to monitor or criticize the performance of others (Janis, 1972). There are presently no data with which to answer these questions. Plans are underway for further high-fidelity simulation work that may shed light on these issues.

REFERENCES

Bales, R. F.: Interaction process analysis: Theory, research, and application. Reading, Mass: Addison-Wesley, 1950.

Bartlett, F. C.: Fatigue following highly skilled work. Proceedings of the Royal Society 1943; Series B., 131:247-257. 247-257.

Brown, I. D.: Measurement of control skills, vigilance, and performance of a subsidiary task during 12 hrs. of car driving. Ergonomics 1967; 10:665-673. 665-673.

Chiles, W. D.: Experimental studies of prolonged wakefulness. W.D.A.C. Tech. Report No. 55-395 1955, Dayton, Ohio.

Cooper, G. E.; White, M. D.; and Lauber, J. K. (Eds.): Resource management on the flight deck. NASA CP-2120, 1979.

Dureman, E. I.; and Boden, C.: Fatigue in simulated car driving. Ergonomics 1972. 15:299-308.

Foushee, H. C.: Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. American Psychologist 1984. 39:885-893.

Foushee, H. C.; and Helmreich, R. L.: Group interaction and flightcrew performance. In E. L. Wiener and D. C. Nagel (Eds.). Human factors in modern aviation. New York: Academic Press, 1986, in press.

Foushee, H. C.; and Manos, K. L.: Information transfer within the cockpit: Problems in intracockpit communications. In C. E. Billings and E. S. Cheaney (Eds.) Information transfer problems in the aviation system. NASA TP-1875, 1981.

Gagne, R. M.: Task variables in fatigue. In W. F. Floyd and A. T. Welford (Eds.). Symposium on fatigue. London: H. K. Lewis, 1953.

Gander, P. H.; Graeber, R. C.; Foushee, H. C.; and Lauber, J. K.: Crew factors in flight operations II: Psychophysiological responses to short-haul air transport operations. NASA TM-88321, 1986.

Graeber, R. C.: Alterations in performance following rapid transmeridian flight. In F. M. Brown and R. C. Graeber (Eds.). Rhythmic aspects of behavior. Hillsdale, NJ: Laurence Erlbaum, 1982.

Graeber, R. C.; Foushee, H. C.; and Lauber, J. K.: Dimensions of flight crew performance decrements: Methodological implications for field research. In H. M. Wegman (Ed.). Breakdown in human adaptation to stress: Towards a multidisciplinary approach. The Hague: Nijhoff, 1984.

Hart, S. G.; Battiste, V.; and Lester, P. T.: Popcorn: A supervisory control simulation for workload and performance research. Proceedings of the 20 Annual Conference on Manual Control 1984.

Holding, D. H.: Fatigue. In G. R. J. Hockey (Ed.). Stress and fatigue in human performance. New York: John Wiley and Sons, 1983.

Janis, I. L.: Victims of groupthink: A psychological study of foreign policy decisions and fiascos. Boston: Houghton-Mifflin, 1972.

Kazdin, A. E.: Single case research designs. New York: Oxford University Press, 1982.

Klein, K. E.; Bruner, H.; Holtmann, H.; Rehme, H.; Stolze, J.; Steinhoff, W. D.; and Wegman, H. M.: Circadian rhythm of pilots' efficiency and effects of multiple time zone travel. Aerospace Medicine 1970; 41:126-132.

Klein, K. E.; and Wegman, H. M.: Significance of circadian rhythms in aerospace operations. AGARDOGRAPH No. 247. Neurilly-Sur-Seine, France, 1980.

Lauber, J. K.; and Foushee, H. C.: Guidelines for line-oriented flight training, (Vol. 1). NASA CP-2184, 1981.

McFarland, R. A.: Human factors in air transportation. New York: McGraw-Hill, 1953.

Moses, J. M.; Lubin, L.; Naitoh, P.; and Johnson; L. C.: Evaluation of the effects of sleep loss: The NPRU mood scale. Technical Report 74-25, 1974.

Ruffell Smith, H. P.: A simulator study of the interaction of pilot workload with errors, vigilance, and decisions. NASA TM-78482, 1979.

Wever, R. A.: The circadian system of man: Results of experiments under temporal isolation. New York: Springer-Verlag, 1979.

APPENDIX A

SHORT-HAUL SIMULATOR STUDY OBSERVER RATING FORM (v.5)

Condition________________________ Pilot Flying____________________(Capt/FO)

Capt. ID____________ F/O ID____________ Observer________________________

Use the following ratings for all categories:

1 - below average
2 - slightly below average
3 - average
4 - slightly above average
5 - above average
n/a - not observed or not applicable

PREFLIGHT

	Captain	First Officer	Notes
Crew Coordination/Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
ATC/Company Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Plan. & Sit. Awareness	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Procedures, Checklists, Callouts	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
PA & PAX Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Overall Performance & Execution	1 2 3 4 5 n/a	1 2 3 4 5 n/a	

TAXI/TAKEOFF

	Captain	First Officer	Notes
Crew Coordination/Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
ATC/Company Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Plan. & Sit. Awareness (mach trim)	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Plan. & Sit. Awareness (T/O alt.)	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Procedures, Checklists, Callouts	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
PA & PAX Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Aircraft Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Overall Performance & Execution	1 2 3 4 5 n/a	1 2 3 4 5 n/a	

CLIMB

	Captain	First Officer	Notes
Crew Coordination/Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
ATC/Company Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Plan. & Sit. Awareness (T-storm)	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Procedures, Checklists, Callouts	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
PA & PAX Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Aircraft Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Overall Performance & Execution	1 2 3 4 5 n/a	1 2 3 4 5 n/a	

CRUISE

	Captain	First Officer	Notes
Crew Coordination/Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	

ATC/Company Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Plan. & Sit. Awareness	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Procedures, Checklists, Callouts	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
PA & PAX Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Aircraft Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Overall Performance & Execution	1 2 3 4 5 n/a	1 2 3 4 5 n/a	

APPROACH & MISSED APPROACH

	Captain	First Officer	Notes
Crew Coordination/Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
ATC Company Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Plan. & Sit. Awareness	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Procedures, Checklists, Callouts	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
PA & PAX Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Stress Management (missed appr.)	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Aircraft Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Overall Performance & Execution	1 2 3 4 5 n/a	1 2 3 4 5 n/a	

EMERGENCY PROCEDURE--SYSTEM A HYDRAULIC FAILURE

	Captain	First Officer	Notes
Crew Coordination/Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
ATC/Company Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Plan. & Sit. Awareness	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Procedures, Checklists, Callouts	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
PA & PAX Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Stress Management	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Aircraft Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Overall Performance & Execution	1 2 3 4 5 n/a	1 2 3 4 5 n/a	

CRUISE TO ALTERNATE

	Captain	First Officer	Notes
Crew Coordination/Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
ATC/Company Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Plan. & Sit. Awareness	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Procedures, Checklists, Callouts	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
PA & PAX Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Stress Management	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Aircraft Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Overall Performance & Execution	1 2 3 4 5 n/a	1 2 3 4 5 n/a	

APPROACH & LANDING

	Captain	First Officer	Notes
Crew Coordination/Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
ATC/Company Communications	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Plan. & Sit. Awareness	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Procedures, Checklists, Callouts	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
PA & PAX Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Stress Management	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Aircraft Handling	1 2 3 4 5 n/a	1 2 3 4 5 n/a	
Overall Performance & Execution	1 2 3 4 5 n/a	1 2 3 4 5 n/a	

GENERAL COMMENTS (briefly summarize your impressions of this crew)

ERRORS (list in detail all errors committed by the crew during this session)

APPENDIX B

CHECK PILOT OBSERVER OVERALL RATING FORM (Captain form*)

Condition ________________ Pilot Flying ____________________

Capt. ID________ FO ID_______ Rater ____________

Use the following ratings:
1 = below average
2 = slightly below average
3 = average
4 = slightly above average
5 = above average

<u>For each item, circle the rating that best describes the captain</u>.

1. Overall knowledge of aircraft and procedures.

   1 2 3 4 5

2. Overall technical proficiency.

   1 2 3 4 5

3. Overall smoothness (flying pilot only)

   1 2 3 4 5

4. Crew coordination and internal communication (intentions are clear to all crew members, proper callouts made, etc.).

   1 2 3 4 5

5. External communication (ATC instructions verified, properly monitored, etc.)

   1 2 3 4 5

6. Overall motivation

   1 2 3 4 5

7. Command ability

   1 2 3 4 5

8. Vigilance

1 2 3 4 5

9. Overall performance

1 2 3 4 5

* First Officer form was identical except that Item #7 was omitted.

APPENDIX C

PILOT WORKLOAD RATING FORM

Pilot Name__________________________ Date ____________

Position_____________ PF______ PNF______ (check one) ----------

---

1. Please circle the number which best corresponds to the overall level of workload.

0	1	2	3	4	5	6
low						high

2. Please rate your own performance on the simulated flight.

0	1	2	3	4	5	6
very poor						very good

3. How much attention did this flight demand?

0	1	2	3	4	5	6
very little						a great deal

4. How complex was the flight?

0	1	2	3	4	5	6
not at all						very

5. How much time pressure did you feel during the flight?

0	1	2	3	4	5	6
none						a great deal

6. How much mental effort did the flight require?

0	1	2	3	4	5	6
none						a great deal

7. How busy were you?

0	1	2	3	4	5	6
not at all						very

8. How <u>difficult</u> was the flight?

0	1	2	3	4	5	6
easy						difficult

9. How <u>motivated</u> were you to perform?

0	1	2	3	4	5	6
not at all						very

10. How do you <u>feel</u> after the simulated flight?

0	1	2	3	4	5	6
fresh						tired

0	1	2	3	4	5	6
relaxed						tense

APPENDIX D

PHYSICAL STATE DATA FORM

ID# ______ Condition __________ Capt. or FO (circle one)

Please fill in the approximate times you went to sleep and when you awoke for the previous four nights as best you can remember.

EST or EDT (circle one)

	Time Asleep	Time Awoke	Total Nap Time	Total Sleep Time
Last night	________	________	________(hrs/min)	________(hrs)
2 nights ago	________	________	________(hrs/min)	________(hrs)
3 nights ago	________	________	________(hrs/min)	________(hrs)
4 nights ago	________	________	________(hrs/min)	________(hrs)

Rate your last night's sleep from least (1) to most (5)

Difficulty falling asleep? 1 2 3 4 5 Difficulty arising? 1 2 3 4 5
How deep was your sleep? 1 2 3 4 5 How rested you feel? 1 2 3 4 5

Please answer the following items about how you feel right now.

	not at all	a little	moderately	quite a bit	extremely
active	0	1	2	3	4
vigilant	0	1	2	3	4
annoyed	0	1	2	3	4
carefree	0	1	2	3	4
cheerful	0	1	2	3	4
considerate	0	1	2	3	4
defiant	0	1	2	3	4
dependable	0	1	2	3	4
sleepy	0	1	2	3	4
dull	0	1	2	3	4
efficient	0	1	2	3	4
friendly	0	1	2	3	4
full of pep	0	1	2	3	4
grouchy	0	1	2	3	4
happy	0	1	2	3	4
jittery	0	1	2	3	4
kind	0	1	2	3	4
lively	0	1	2	3	4
pleasant	0	1	2	3	4
relaxed	0	1	2	3	4
forgetful	0	1	2	3	4
sluggish	0	1	2	3	4
tense	0	1	2	3	4
clear thinking	0	1	2	3	4
tired	0	1	2	3	4
hard working	0	1	2	3	4

Place a mark on the line at the point which best corresponds to your present state of alertness.

MOST ALERT ______________________________ MOST DROWSY

APPENDIX E

COMMUNICATION CATEGORIES

1) COMMAND: a specific assignment of responsibility by one group member to another.

2) OBSERVATION: recognizing and/or noting a fact or occurrence relating to the task.

3) SUGGESTION: recommendation for a specific course of action.

4) STATEMENT OF INTENT: announcement of an intended action by speaker. Includes statements referring to present and future actions, but not to previous actions.

5) INQUIRY: a request for factual information relating to the task. Not a request for action.

6) AGREEMENT: a response in concurrence with a previous speech act; a positive evaluation of a prior speech act.

7) DISAGREEMENT: a response NOT in concurrence with a previous speech act; a negative evaluation of a prior speech act.

8) ACKNOWLEDGEMENT: a) makes known that a prior speech act was heard; b) does not supply additional information; c) does not evaluate a previous speech act.

9) ANSWER SUPPLYING INFORMATION: speech act supplying information beyond mere agreement, disagreement, or acknowledgment.

10) RESPONSE UNCERTAINTY: statement indicating uncertainty or lack of information with which to respond to a speech act.

11) TENSION RELEASE: laughter or humorous remark.

12) FRUSTRATION/ANGER/DERISIVE COMMENT: statement of displeasure with self, other persons, or some aspect of the task; or a ridiculing remark.

13) EMBARRASSMENT: any comment apologizing for an incorrect response, etc.

14) REPEAT: restatement of a previous speech act without prompting.

15) CHECKLIST: prompts and replies to items on a checklist.

16) NON-TASK RELATED: any speech act referring to something other than the present task.

17) NON-CODABLE: speech act which in unintelligible or unclassifiable with respect to the present coding scheme.

18) ATC COMMUNICATION: any communication over the radio with ATC, dispatch, "the company", etc.

ORIGINAL PAGE IS
OF POOR QUALITY

APPENDIX F

SAMPLE AIRCRAFT PERFORMANCE DATA

DATE AND TIME:	17-JUL-1984 20:44:23		
AIRSPEED:	277.03580	ALTITUDE:	19023.11
MAG. HEADING:	53.70029	VERT. SPEED:	979.9
ENG. #1 EPR:	1.812	VOR/LOC DEV:	0.135
G/S DEV	8.0000000	FLAP DEGREES:	0.000
GEAR POS.:	0.00000	TIME SEC.:	68899.70
#1 NAV FREQ.:	110400	#1 DME:	117

DATE AND TIME:	17-JUL-1984 20:44:38		
AIRSPEED:	280.26279	ALTITUDE:	19028.15
MAG. HEADING:	48.10065	VERT. SPEED:	-522.4
ENG. #1 EPR:	1.553	VOR/LOC DEV:	0.141
G/S DEV	8.0000000	FLAP DEGREES:	0.000
GEAR POS.:	0.00000	TIME SEC.:	68914.70
#1 NAV FREQ.:	110400	#1 DME:	102

DATE AND TIME:	17-JUL-1984 20:44:53		
AIRSPEED:	280.84479	ALTITUDE:	18998.43
MAG. HEADING:	51.12605	VERT. SPEED:	81.9
ENG. #1 EPR:	1.555	VOR/LOC DEV:	0.052
G/S DEV	8.0000000	FLAP DEGREES:	0.000
GEAR POS.:	0.00000	TIME SEC.:	68929.70
#1 NAV FREQ.:	110400	#1 DME:	85

DATE AND TIME:	17-JUL-1984 20:45:08		
AIRSPEED:	282.92468	ALTITUDE:	18985.18
MAG. HEADING:	51.75150	VERT. SPEED:	-127.4
ENG. #1 EPR:	1.554	VOR/LOC DEV:	0.053
G/S DEV	8.0000000	FLAP DEGREES:	0.000
GEAR POS.:	0.00000	TIME SEC.:	68944.70
#1 NAV FREQ.:	110400	#1 DME:	69

DATE AND TIME:	17-JUL-1984 20:45:23		
AIRSPEED:	280.86337	ALTITUDE:	19002.85
MAG. HEADING:	51.90620	VERT. SPEED:	107.9
ENG. #1 EPR:	1.554	VOR/LOC DEV:	0.077
G/S DEV	8.0000000	FLAP DEGREES:	0.000
GEAR POS.:	0.00000	TIME SEC.:	68959.70
#1 NAV FREQ.:	110400	#1 DME:	53

DATE AND TIME:	17-JUL-1984 20:45:38		
AIRSPEED:	285.85156	ALTITUDE:	18993.17
MAG. HEADING:	51.97639	VERT. SPEED:	-6.0
ENG. #1 EPR:	1.529	VOR/LOC DEV:	0.113
G/S DEV	8.0000000	FLAP DEGREES:	0.000
GEAR POS.:	0.00000	TIME SEC.:	68974.70
#1 NAV FREQ.:	110400	#1 DME:	37

DATE AND TIME:	17-JUL-1984 20:45:53		
AIRSPEED:	286.68530	ALTITUDE:	18996.42
MAG. HEADING:	51.92190	VERT. SPEED:	4.8
ENG. #1 EPR:	1.530	VOR/LOC DEV:	0.020
G/S DEV	8.0000000	FLAP DEGREES:	0.000
GEAR POS.:	0.00000	TIME SEC.:	68989.70
#1 NAV FREQ.:	114100	#1 DME:	971

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM 88322		

4. Title and Subtitle

CREW FACTORS IN FLIGHT OPERATIONS: III. THE OPERATIONAL SIGNIFICANCE OF EXPOSURE TO SHORT-HAUL AIR TRANSPORT OPERATIONS

5. Report Date

August 1986

6. Performing Organization Code

7. Author(s) H. Clayton Foushee, John K. Lauber, Michael M. Baetge (Informations General Corporation, Palo Alto, CA) and Dorothea B. Acomb (San Jose State University, San Jose, CA 95192)

8. Performing Organization Report No.

A-86338

10. Work Unit No.

9. Performing Organization Name and Address

Ames Research Center
Moffett Field, CA 94035

11. Contract or Grant No.

13. Type of Report and Period Covered

Technical Memorandum

12. Sponsoring Agency Name and Address

National Aeronautics and Space Administration
Washington, DC 20546

14. Sponsoring Agency Code

505-67-41

15. Supplementary Notes

Point of Contact: H. Clayton Foushee, Ames Research Center, MS 239-21, Moffett Field, CA 94035
(415) 694-6114 or FTS 464-6114

16. Abstract

Excessive flightcrew fatigue as a result of trip exposure has long been cited as a factor with potentially serious safety consequences. Laboratory studies have implicated fatigue as a causal factor associated with varying levels of performance deterioration depending on the amount of fatigue and the type of measure utilized in assessing performance. From an operational standpoint, these studies have been of limited utility because of the difficulty of generalizing laboratory task performance to the demands associated with the operation of a complex aircraft.

This study examined the performance of 20 volunteer twin-jet transport crews in a full-mission simulator scenario that included most aspects of an actual line operation. The scenario included both routine flight operations and an unexpected mechanical abnormality which resulted in a high level of crew workload. Half of the crews flew the simulation within two to three hours after completing a three-day, high-density, short-haul duty cycle (Post-Duty condition). The other half of the crews flew the scenario after a minimum of three days off duty (Pre-Duty condition).

The results of this study revealed that, not surprisingly, Post-Duty crews were significantly more fatigued than Pre-Duty crews. However, a somewhat counter-intuitive pattern of results emerged on the crew performance measures. In general, the performance of Post-Duty crews was significantly better than the performance of Pre-Duty crews. Post-Duty crews were rated as performing better by an expert observer on a number of dimensions relevant to flight safety. Analyses of the flightcrew communication patterns revealed that Post-Duty crews communicated significantly more overall, suggesting, as has previous research, that communication is a good predictor of overall crew performance.

Further analyses suggested that the primary cause of this pattern of results is the fact that crewmembers usually have more operating experience together at the end of a trip, and that this recent operating experience serves to facilitate crew coordination, which can be an effective countermeasure to the fatigue present at or near the end of a duty cycle. These results have important aircrew training and aviation safety implications.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Communications Resource management Fatigue	Unlimited Subject category - 51

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	63	A04

*For sale by the National Technical Information Service, Springfield, Virginia 22161