NASA Technical Memorandum 102259

Personality Factors in Flight Operations: Volume I. Leader Characteristics and Crew Performance in a Full-Mission Air Transport Simulation

Thomas R. Chidester and Barbara G. Kanki
Ames Research Center, Moffett Field, California

H. Clayton Foushee
Federal Aviation Administration, Washington, D.C.

Cortlandt L. Dickinson
Menlo Park, California

Stephen V. Bowles
California Professional School for Psychology, Berkeley, California

April 1990

NASA
National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035-1000

TABLE OF CONTENTS

Page

SUMMARY

INTRODUCTION 2

METHOD 5
- 2.1 Study Design Overview 5
- 2.2 Subjects and Recruitment 5
  - 2.2.1 Pretesting 6
  - 2.2.2 Profile Classification 6
- 2.3 Confidentiality 7
- 2.4 Experimental Equipment 7
- 2.5 Personnel 7
- 2.6 Experimental Procedure 8
- 2.7 Simulation Scenarios 8
  - 2.7.1 Day One Scenarios 9
  - 2.7.2 Day Two Scenarios 10
- 2.8 Measurement 11
  - 2.8.1 Flight Experience and Personality Data 12
  - 2.8.2 Subjective Workload 12
  - 2.8.3 Crew Performance 12
    - 2.8.3.1 Observer Ratings 12
    - 2.8.3.2 Crew Errors 13
    - 2.8.3.3 Aircraft Handling 14

RESULTS 14
- 3.1 Pretest Data 14
  - 3.1.1 Flight Experience 14
  - 3.1.2 Personality Data 14
- 3.2 Subjective Workload 16
- 3.3 Observer Ratings 17
- 3.4 Error Analyses 19
- 3.5 Aircraft Handling 20

DISCUSSION 20
- 4.1 Crew Familiarity 22
- 4.2 Implications for Designing Operationally-valid Research 24

REFERENCES 25

PRECEDING PAGE BLANK NOT FILMED

Page

APPENDICES

A	Definitions of Personality Scales	28
B	Flight Experience Items	29
C	Personality Battery	30
D	Task Load Index Items and Definitions	39
E	Items Comprising the Observer's Rating of Crew Performance	40
F	Observer Ratings of Phase Performance	41

SUMMARY

Crew effectiveness is a joint product of the piloting skills, attitudes, and personality characteristics of team members. As obvious as this point might seem, both traditional approaches to optimizing crew performance and more recent training development highlighting crew coordination have emphasized only the skill and attitudinal dimensions. This volume is the first in a series of papers on this simulation. A subsequent volume will focus on patterns of communication within crews. This paper reports the results of a full-mission simulation research study assessing the impact of individual personality on crew performance. Using a selection algorithm described in previous research, captains were classified as fitting one of three profiles along a battery of personality assessment scales. The performances of 23 crews led by captains fitting each profile were contrasted over a one-and-one-half-day simulated trip. Crews led by captains fitting a positive Instrumental-Expressive profile (high achievement motivation and interpersonal skill) were consistently effective and made fewer errors. Crews led by captains fitting a Negative Expressive profile (below average achievement motivation, negative expressive style, such as complaining) were consistently less effective and made more errors. Crews led by captains fitting a Negative Instrumental profile (high levels of competitiveness, verbal aggressiveness, and impatience and irritability) were less effective on the first day but equal to the best on the second day. These results underscore the importance of stable personality variables as predictors of team coordination and performance.

INTRODUCTION

The effectiveness of an aircrew is a joint product of the technical skills, attitudes, and personality characteristics of its individual members (Sells, 1955) and the process by which they plan, execute, and solve problems. As obvious as this might seem, traditional approaches to the optimization of crew performance in air transport operations have emphasized the skills dimensions almost exclusively. As pilots are selected, trained, and evaluated, the primary emphasis is on ensuring that each has the technical skills necessary to perform his or her role in the cockpit. More recently, many training programs have been expanded to include the influence of interpersonal characteristics associated with crew coordination, in the form of Cockpit Resource Management (CRM) programs (see Orlady and Foushee, 1987). While it is certainly encouraging that efforts are now underway to improve both technical and interpersonal skills, the impact of other factors, particularly stable personality characteristics, may provide an additional contribution to the crew performance process.

The search for personality predictors of performance in aviation operations has been plagued by an historic failure to validate links between those dimensions and performance criteria (Melton, 1947, 1957; Ellis and Conrad, 1948; R. S. Melton, 1954). The reasons for these failures are complex, and include the strategies of performance evaluation or assessment employed and problems with conceptualizations of the role of personality. For example, the vast majority of personality research in aviation settings has examined performance during initial pilot training and has employed criteria such as completing training or obtaining a pilot rating (Dolgin and Gibb, 1988). These types of criteria probably do not capture the range of performance occurring in day-to-day flight. A focus on training-completion misses the complex setting in which crews perform.

Judging the operational significance of personality solely on the basis of training-completion research implicitly assumes that the relationship between personality and performance remains constant from training through initial experience to routine performance. Helmreich, Sawin, and Carsrud (1986) have presented evidence that performance during and shortly after training is much less sensitive to personality effects than performance after the effects of training have begun to subside. As a result, it is necessary to assess the relationship at the level of routine performance, or in critical situations, to answer questions of operational relevance. Unfortunately, criteria available from routine flight settings in the form of regular performance checks typically consist of pass-fail evaluations, with failures being extremely rare, and tend to focus on very standard tasks like control-manipulation smoothness or completion of specific maneuvers. Using flight checks as criteria would mask a broader range of performance variability and minimize the power of any predictor of performance (Hackman and Helmreich, 1987).

In any event, the research conducted to date has not fully explored the *range* of valid research questions bearing on the link between personality and performance. This range includes the following issues: First, one can ask whether those attracted to aviation or space settings differ on average from the general population. Second, among those sufficiently attracted to present themselves as candidates, does personality predict successful completion of training? Third, do crewmembers who are more successful or proficient in their duties over the course of a career differ in personality from those who are less successful? Fourth, does personality predict individual performance or do particular combinations of crew personalities predict crew performance in critical flight situations? Fifth, how does personality interact with training, task design, or other variables in the prediction of performance? Though some research has suggested pilot population differences in personality characteristics (e.g., Fry and Reinhardt, 1969), only the training-completion question has been adequately explored. The present study focuses on the impact of personality in critical flight situations.

Our strategy for evaluating the potential impact of personality characteristics on crew effectiveness was (1) to define critical elements of performance, (2) to identify dimensions of personality theoretically linked to these elements, (3) to identify a selection algorithm to classify or differentiate individual subjects along these dimensions, and (4) to conduct a high-fidelity validation study to determine whether these theoretical relationships translate into practical performance consequences. As a starting point for research, we chose to focus on the impact of crew leaders and the characteristics that contribute to leadership. Even though our understanding of group phenomena is not what it should be, there are a number of possible elements of a successful crew that have been suggested by previous research. First, high levels of individual technical skill, proficiency, and the motivation to work hard are the foundation upon which effective crew coordination is built. Second, past research (e.g., Kanki, Lozito, and Foushee, 1988) has demonstrated that effective crews are characterized by communications patterns that tend to be both predictable and responsively linked. In addition, we would suggest that effective leadership is a joint function of: 1) maintaining effective task delegation and definition, 2) encouraging cross-checking and feedback, and 3) creating an atmosphere where subordinates feel free to offer suggestions and counter-proposals to leader-prescribed courses of action (e.g., Ginnett, 1987). Our task was to seek personality measures that captured these elements.

A great deal of emphasis has been placed upon the first element, individual technical skills and the motivation to achieve. For research purposes, we chose to focus on dimensions underlying the motivational component of individual performance, one's overall level of "instrumentality," which we define operationally as a person's level of goal orientation and independence. We also chose to emphasize "achievement striving" as an additional measure of an individual's dispositional orientation toward task performance situations. A second dimension, oriented toward communication and interpersonal exchange, is commonly defined as "expressivity," or interpersonal warmth and sensitivity. Communication would be expected to be facilitated in groups led by or composed of individuals characterized by high levels of expressivity and inhibited in groups composed of individuals characterized by both negative expressive (e.g., frequent complaining) and negative instrumental traits such as verbal aggressiveness, competitiveness, or impatience and irritability. In summary, we theorized that effective leaders are more often characterized by relatively high levels of both positive instrumentality and positive expressivity (high levels of both concern for people and concern for performance), and that this type of leadership style would facilitate crew performance. Lower levels of positive expressivity and higher levels of negative expressive and negative instrumental traits were expected to lead to less effective crew communication, coordination, and performance overall. Definitions of each of these dimensions and the instruments measuring each are displayed in appendix A.

These dimensions were chosen for several reasons. First, a number of personality theorists and researchers have focused on dimensions reflecting instrumentality and expressivity as one central set of personality characteristics. Influential theorists (e.g., Spence and Helmreich, 1978; Fiedler, 1967) have all in one way or another identified these characteristics as core components of human personality with strong behavioral relationships. Moreover, many popular management theories have espoused concern for people balanced with concern for performance as the key to leader success. Blake and Mouton's

(1978) "managerial grid" is perhaps the most widely applied example of these notions, and it has been incorporated into many training programs, including a number in aviation. Second, these dimensions capture theoretically relevant traits corresponding to core elements of performance in aviation. Third, a great deal of real-world performance data has been collected for the instruments assessing these dimensions (Spence and Helmreich, 1983; Helmreich, Spence, and Pred, 1988).

So far, we have suggested possible relationships between dimensions of personality and elements of crew performance. But if one wishes to apply personality research operationally, either by selecting individuals with desirable characteristics or tailoring training to individual personality profiles, one must have some means for classifying individuals. Typically in research, one concentrates on one or two dimensions of personality. A frequent approach is to determine, through multiple regression techniques, the unique portion of behavioral variance contributed by a particular trait. But in applied settings, a researcher may need to consider the distribution of combinations of many different traits within individual subjects, or in other words, to look at the constellation of traits that exist in people working in the applied setting. This is often made necessary by the small numbers of subjects available for study. Providing a research design combining possible levels (even simple median splits) of each trait under study quickly becomes impractical.

Chidester (1987) and Gregorich et al. (1989) have employed the technique of cluster analysis within several samples of pilots to determine the distributions of differing combinations of positive and negative personal attributes along the personality dimensions described above. Cluster analysis is a statistical technique which combines subjects into groups or clusters based upon each subject's similarity to other subjects along any specified set of dimensions (Anderberg, 1973). Those subjects that fell into each cluster are similar to one another along the dimensions analyzed and different from individuals in other clusters. The subgroups that were identified through this process reflect meaningful constellations of traits as they are distributed across individuals. Three distinct clusters have been found across samples of pilots, one with high levels of positive traits and two others with different constellations of negative traits. Pilots in the positive cluster are characterized by high levels of instrumentality, expressivity, achievement striving, work, and mastery and are designated the positive instrumental-expressive or "IE+" cluster. One of the negative clusters is defined by high levels of negative expressivity and low levels of instrumentality and achievement striving. This cluster is characterized by traits associated with tendencies to express oneself in a negative fashion (e.g., complaining) and lower than average goal orientation. It has been labeled the negative expressive or "EC–" cluster. The second negative cluster is characterized by higher than average levels of verbal aggressiveness, negative instrumentality, and competitiveness. This cluster comprises a more "authoritarian" orientation and may well be associated with elements of a profile popularly known as "the right stuff." It has been labeled the negative instrumental or "I–" cluster. Chidester (1987) found some evidence that these clusters may be related to determinants of crew performance. Pilots responded differentially to training in crew coordination as a function of these profiles. IE+ pilots benefitted the most from training as assessed by changes in attitudes concerning cockpit management.

Having specified a set of elements critical to crew performance, a set of personality characteristics linked to these elements, and a means for classifying individuals along multiple dimensions, our task was to assess whether these characteristics were operationally relevant. That is, do these theoretical relationships translate into real-world performance differences? We put these characteristics to the test in a full-mission simulation research experiment.

METHOD

2.1 Study Design Overview

The current study was designed to (1) evaluate whether the personality characteristics of aircraft commanders significantly impact the crew performance process and (2) evaluate the experimental classification algorithm as a possible countermeasure for the prevention of crew coordination problems. A two-day full-mission simulation study was designed, in which crews flew five flight segments under varying conditions of workload and in which crews were chosen according to personality criteria. Three different types of crews were composed. The crew types contrasted were based upon cluster membership as described by Chidester (1987). Crew types represent selection for leadership; that is, only the captain's personality characteristics were considered when crews were chosen. The first crew type was composed of a randomly assigned first officer and flight engineer flying with a captain from the IE+ cluster. We hypothesized that these captains would be good leaders and their leadership would translate into effective crew performance. The second type was composed of a randomly assigned first officer and flight engineer flying with a captain from the I– cluster. The third type was composed of a randomly assigned first officer and flight engineer flying with a captain from the EC– cluster. We hypothesized that these two negative leader crews would be less effective at crew coordination in the high workload flight segments. These hypotheses were tested using data collected from four sources: self-reports, expert observation, video-based coding of errors, and aircraft handling parameters.

2.2 Subjects and Recruitment

Twenty-three, three-person crews (69 pilots) completed a one and one-half day full-mission simulation of airline operations in the Ames Man-Vehicle Systems Research Facility (MVSRF) Boeing 727 simulator. All crews were employed by the same major U.S. air carrier, all crewmembers were currently operating the B-727 exclusively in passenger operations, and all crewmembers were at the time qualified in the B-727 crew position (captain, first officer, second officer, or flight engineer) they occupied in the simulation.

All crewmembers had completed an initial course in crew coordination or CRM and all had participated in at least one line-oriented flight training (LOFT) session during recurrent training. LOFT (Lauber and Foushee, 1981) is a form of training in which a crew completes a full-mission simulation very similar to that in this research project. Unlike traditional simulator training, the focus is not on the completion of a specified set of maneuvers or procedures, but on training the crew to deal with problems in the manner required in the line environment; as a team working and exchanging information with each other, with air traffic control, and with company dispatch and maintenance services.

Subjects were recruited through an announcement letter delivered to their company mailboxes at their local domicile. This letter described the study as a major simulation examining the factors influencing crew coordination in routine line operations and described the degree of participation requested. Subjects were asked to notify a local union council member if they did not wish to be contacted by NASA investigators by telephone. Twelve of the 394 pilots and flight engineers in the domicile declined to participate at this point. Telephone numbers for the remaining 382 pilots were released to the research team by the company through the local union executive committee. One member of the research team then attempted to contact all of these pilots by telephone and mailed a copy of the pretest to those pilots contacted and willing to participate. A total of 161 questionnaires were mailed; 121 pilots returned usable pretests. Of that number, 69 subjects (18% of those in the domicile) were subsequently scheduled for the simulation. While this may appear to be a low percentage of participation, the degree of participation required (traveling to the research center on two consecutive days off duty), the geographic dispersion of pilots assigned to the domicile (many pilots commute great distances to begin each duty cycle), and the limited amount of available simulator time made a higher complete-participation rate unlikely. We were able to fill virtually all of the simulator time slots available to us despite three last-minute cancellations over the course of the study. Given these constraints, the rate of return of pretest questionnaires (31% of the domicile) is a more accurate reflection of willingness of subjects to participate.

2.2.1 Pretesting– Prior to scheduling for the simulation, candidate subject pilots completed a battery of personality instruments composed of the Expanded Personal Attributes Questionnaire (EPAQ; Spence, Helmreich, and Holahan, 1979), the Work and Family Orientation Questionnaire (WOFO; Spence and Helmreich, 1978), and the Achievement Striving and Impatience/Irritability scales (A/S, I/I; Pred, Spence, and Helmreich, 1986) derived from the Jenkins Activity Survey (JAS) measure of the Type A Behavior Pattern (Jenkins, Zyzanski, and Rosenman, 1971) along with a number of items focusing on flight experience. Scoring of these instruments by their published instructions results in 10 scale-scores: instrumentality, expressivity, negative instrumentality, verbal aggressiveness, negative communion, work, mastery, competitiveness, achievement striving, and impatience/irritability (see appendix A).

2.2.2 Profile Classification– The personality battery was scored so that cluster profile could be determined for each subject. A scoring routine developed by Chidester (1987) was utilized in lieu of conducting a new cluster analysis in this small sample. This routine compares a subject's score on each dimension to norms (sample median) based on a sample of over 400 airline pilots (Chidester, 1990). Each subject is then considered for inclusion into a cluster based upon his/her relative standing (above or below the median) on each dimension as compared to the pattern of median scores found for that cluster in Chidester's sample. Individuals were assigned to (1) the IE+ cluster if they scored above the median on three of the following dimensions: instrumentality, expressivity, mastery, and work, (2) the I-cluster if they scored above the median on negative instrumentality and verbal aggressiveness and below the median on expressivity, or (3) the EC– cluster if they scored above the median on Negative Communion and below the median on three of the following dimensions: instrumentality, achievement striving, mastery, work, and impatience/irritability. If an individual met none of these criteria, he/she was listed as unclassifiable. Unclassifiable captains were not pursued further for recruitment, but unclassifiable first and second officers were recruited since their assignment to crews was intended to be random.

The careful reader will recognize a number of alternative ways of assigning new individuals to previously-defined clusters. For example, the scores of each individual may be compared to cluster centroids (see Norusis, 1988) and included in the cluster to which the individual is closest. Chidester's (1987) scoring routine was chosen to emphasize distinction between clusters. Since we would contrast a small sample, large personality differences between clusters were desirable and borderline-case leaders (the unclassifiable) were not recruited. Most alternative classification schemes would result in less distinct clusters.

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 participating pilots complete confidentiality. All data are identified by a five-digit code number. Thus, it is not possible for anyone, including the NASA investigators, to identify any participating pilot by name. The code number merely provides a link between each subject's pretest data and his or her participation in the simulator.

2.4 Experimental Equipment

A Boeing 727-200 simulator operated by the MVSRF at Ames Research Center was utilized. This simulator has a six degree-of-freedom motion platform and four-window visual system. It was manufactured by Singer-Link Corporation, and is equipped with special effects and programmed with the aircraft performance data required to meet Federal Aviation Administration (FAA) Phase II certification. The MVSRF was constructed specifically for research purposes and is configured for detailed data collection. Simulator computers record aircraft configuration and handling information, and multi-camera videotape and multichannel audiotape recording systems are installed for capturing crew communication and action. A remotely-located Air Traffic Control (ATC) facility with flight-progress monitoring displays and voice-disguising equipment provides for highly realistic ATC support. The facility also supports equipment to provide Automated Terminal Information System (ATIS) information to the crew over VHF radio and background recordings of ATC communication provided by the FAA with aircraft operating in airspace controlled by Air Route Traffic Control Centers (ARTCC). Both of these capabilities greatly enhanced the realism of flights conducted in the simulator. We gratefully acknowledge the assistance of the FAA in obtaining recordings of Oakland and Los Angeles ARTCC communications.

2.5 Personnel

Simulator operations were accomplished with a basic staff of six people. An expert observer (described in section 2.8.3.1) rode in the simulator cab along with the subject crew. A simulator operator and the experimenter occupied an Experimenter Operations Station (EOS) located remotely from the simulator cab. The EOS room included a data-entry terminal through which all simulator functions were initiated, modified, and terminated. (Aircraft setup and events were pre-programmed and the operator needed only to initiate the program and make any changes necessitated by crew decisions, such as adding fuel). Video and audio monitors and recording equipment were also located in the EOS, allowing the experimenter and operator to monitor crew communications and to detect simulator problems. The experimenter started and stopped audio-video recording, communicated with the observer in event of simulator problems, served as ground-crewman and company dispatcher, and supervised the operation of the simulator and support facilities. The remote ATC facility housed an air traffic controller, a pseudo-pilot, and an ATC facility manager. The controller was a retired FAA-certified traffic controller who provided all clearances and ATC services to the subject crew. The pseudopilot initiated calls to ATC over the radio frequency in use by the subject crew. These calls were scripted to simulate "traffic" as it would normally occur in the operating environment. The pseudopilot also operated "target aircraft" which were visible to the crew though the four-window visual system. A number of aircraft operating along approach paths and airways along the route of flight were simulated in an attempt to increase realism. An ATC facility manager worked with the controller and pseudopilot, operating the ATC facility computers and monitoring the course of the simulation. The ATC manager also served as the interface to facility maintenance and support personnel.

2.6 Experimental Procedure

Crews were met upon arrival at the simulator facility on the first day by the experimenter. In an initial briefing, the importance of operational realism was emphasized and pilots were urged to treat the simulation just as if it were an actual two-day trip. Crews 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. Crewmembers then received a detailed briefing on the differences between the simulator's configuration and that of their company's aircraft. This briefing was conducted by the expert observer. Following the briefing, crews were given a schedule of operations, weather information and flight plans for two familiarization flight segments, and an opportunity to hold an initial crew briefing. Crews were then escorted to the simulator cab, where they completed a round trip between San Francisco and Stockton, California. After completing the familiarization segments, crewmembers had lunch, then returned to initiate the experimental flights.

2.7 Simulation Scenarios

Past full-mission simulation research (Ruffell-Smith, 1979; Lauber and Foushee, 1981; Foushee et al. (1986) has shown that successful simulation scenarios have at least five essential elements. First, they are designed to be completely representative of the actual operational environment, and all details are faithfully represented. Second, they are complicated enough to require the coordinated action of all crewmembers for successful completion, but not to the extent that they induce complete crew failure such as a "crash." Third, problems presented to crews have ongoing consequences which must be dealt with in flight, but cannot be fixed in flight. Fourth, the problems involved are very ambiguous, and there is usually no simple corrective "by the book" solution. And fifth, the original problem is usually compounded by other events such as weather-induced complications (e.g., landing on a rain-slick runway with partial brake failure). It is also interesting to note that these characteristics have been seen in past incidents and accidents.

In the process of scenario design, outlines of potential events were developed by the principal investigators using accident case studies and incident reports. These were reviewed by investigators with checking and training experience in the particular aircraft type being simulated and by simulator operational personnel from the MVSRF. Typical environmental conditions for the proposed area of flight (November-February weather patterns for coastal and central California) were considered in great detail, so that weather patterns and scenario events would seem realistic to the experimental flight crews. Aircraft documentation and airline dispatch procedures were assembled for each flight segment in cooperation with airline management and members of the local pilot labor union executive committee. Following this development process, selected scenario outlines were programmed into the simulator computer and eight pretest runs were conducted using qualified flight crews to refine procedures, train facility personnel and the experiment staff, and test scenario events. These pretest crews were carefully debriefed to assess the realism of the scenarios and procedures used by the experiment staff. This feedback allowed continual refinement until the scenarios were finalized.

Crews flew five experimental flight segments (legs). Each segment was planned and flown as closely as possible to real operations. Crews were provided with all of their normal flight documentation, completed all normal flight and cockpit preparations, and communicated with all ground support personnel normally available to them. Flight routings corresponded to typical clearances along routes in central and southern California and were adapted from flight plans created for an earlier NASA investigation. Airfields available to the crew included San Francisco (SFO), Stockton (SCK), Sacramento (SMF), and Los Angeles (LAX). Segments 1 through 3 were flown on the first day, segments 4 and 5 occurred on the second day. Routine levels of workload were designed into segments 1, 2, and 4, but segments 3 and 5 were far more demanding than normal and involved continuing abnormal conditions that could not be resolved completely in flight. Crews were led to believe, however, that they would fly six legs. The last leg (SMF-SFO) was not intended to occur, except as a part of preflight planning, because of scenario events. This deception was intended to counteract suspicions that might be associated with the last segment of the study, particularly since the last segment of day one involved an abnormality.

2.7.1 Day One Scenarios– On the first day, all legs except the third were relatively routine. However, one irregular item was included in the dispatch paperwork for the full-mission segments. The deferred item list and accompanying minimum equipment list (MEL), indicated that the plane was to be dispatched with the no. 3 generator inoperative. This was a legal procedure, but given the weather conditions (night, fog, and low cloud ceilings), prudent crews should have considered delaying departure until either the weather improved or the generator was repaired. If the crew elected to request the repair, the generator was "repaired," but the dispatcher warned the crew that they might expect it to malfunction again requiring it to be reset. If the crew did not request repair, the generator was "repaired" later when the crew reached the maintenance base at SFO. In either case, each time the no. 3 engine was started following the "repair," the generator field light illuminated, and at least one field reset was required to bring the generator on line. This manipulation was intended to complicate the decision-making process on a later high workload segment.

On segment 3, crews flew from Sacramento to Los Angeles. Following a normal takeoff and climb, a combination of system failures were activated automatically. First, the vertical stabilizer trim system began running uncommanded and jammed in a nose-down condition at a predetermined point. Second, and shortly thereafter, when the aircraft crossed a specified navigational point, the no. 2 engine low-oil-pressure warning light illuminated and the indicated pressure fell to the cautionary range. This combination represented a relatively high-workload situation, but was compounded by neither weather, traffic, nor ATC problems. No diversion from the flight plan was necessary, except for actions required to stabilize the aircraft and land safely. Segment 3 is displayed graphically in figure 1.

The scenario presented the crew with two independent failures which could have impacted flight safety. The jammed stabilizer trim system was a serious control problem, disabling the autopilot, and requiring (due to the descent configuration) constant nose-up control inputs and considerable back pressure from the flying pilot. This made the approach and landing much more difficult and was physically fatiguing (given the need to hold constant, firm back pressure). The procedure for dealing with the

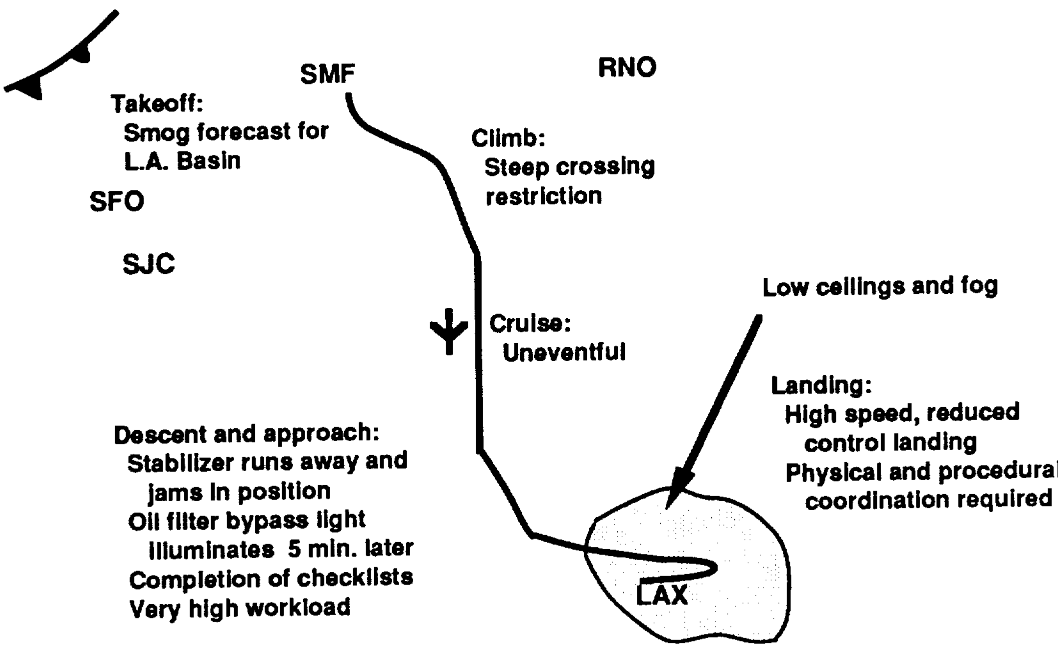

Figure 1.– Scenario for segment 3.

jammed stabilizer trim, once it was identified, was to limit flap setting to 15° for landing, increase approach speed by 15 knots, and to establish the landing configuration as early as possible so as to get a feel for the control forces necessary for landing. This procedure is difficult but not unreasonable since it is required for all type ratings in the B-727 aircraft. The low oil pressure light and corresponding cautionary gauge reading were also covered by a checklist procedure, but the outcome of the checklist left the crew a choice. At the captain's discretion, the crew could either shut the engine down or reduce thrust on the engine to idle. Shutting the engine down required the completion of more checklists (greatly increasing workload and time requirements), may have required the dumping of fuel (again increasing workload), and removed one generator from operation. Recall that the crew had already encountered minor problems with the no. 3 generator, so the aircraft could have easily been down to one should conditions have deteriorated further. The most prudent course would appear to be to keep the engine operating as a reserve while continually monitoring its operating parameters. However, a pilot could reasonably argue that the possibility of the engine failing catastrophically was sufficient to warrant the workload consequences of shutting it down. Once this decision was made, the crew had to land the aircraft in an abnormal configuration.

2.7.2 Day Two Scenarios– On the second day, the weather continued to be characterized by fog, overcast, and minimal visibilities in the central California valley. The first leg (SCK-LAX) went without programmed incident except for fog in the Los Angeles area (ceiling 400 ft, visibility one mile). By the time of the second leg (LAX-SMF), weather had deteriorated further in the central valley, resulting in poor visibility in the Sacramento area (ceiling 300 ft, visibility one-half mile; approaching the legal minimums for ILS approaches). Weather in Los Angeles remained foggy but above legal minimums. Following a normal departure and climb out of the Los Angeles area and normal cruise, the crew received clearance for the Wraps Four Arrival to Sacramento. As the crew entered the Sacramento terminal area, runway visual range (RVR) was reported by the approach controller as 2000 ft, just above the Category 1 approach minimums of 1800 ft. Prudent crews should, at this point, have considered and briefed for a possible Category 2 approach and landing. The minimums for this approach were RVR 1200 with a decision height of 126 ft. However, after the aircraft crossed the outer marker, RVR was reported by the tower as less than 1000 ft, which was below the minimum for the initiation of an approach. However, since the aircraft was inside the marker, it was legal to continue until the published decision height and to land if the captain had an adequate view of the runway environment. If the crew attempted the landing, the runway was not visible at the decision height, requiring the approach to be aborted. During the missed approach, as the crew retracted the flaps, a hydraulic system A failure occurred, caused by a leak that depleted all of the hydraulic fluid. At this point, it was immediately apparent that a diversion to an alternate airport would be necessary. Weather conditions at various nearby alternates were poor (all ceilings less than 800 ft and visibilities less than one mile). Reno (RNO), San Francisco, Oakland, and San Jose (SJC) were the best alternates. Reno was unacceptable because the prevailing wind direction made its shortest runway the active runway and the use of this runway for B-727 operations was prohibited by company policy. SFO had the best weather of the remaining alternates, with clearing conditions, a 2000-ft ceiling, and 2 miles visibility with light rain. SFO also provided a very long runway at just under 12,000 ft. However, SJC was listed as the flight's legal alternate, because SFO weather conditions were below alternate minimums at the time of dispatch. Segment five is displayed graphically in figure 2.

This scenario confronted the crew with a number of hazards and limitations. First, the hydraulic failure disabled a number of aircraft systems. The landing gear had to be extended by hand crank, and once extended could not be retracted. Flaps also had to be extended by alternate means, and this system does not allow leading edge flaps to be retracted once extended. In addition, the trailing edge flaps were not protected against asymmetric extension using the alternate system. Alternate extension required more time than extension by normal means, and was limited to 15° in case of a missed approach (flap retraction from the normal 30° setting by the alternate system following a missed approach would be too slow to reduce drag sufficiently to allow the aircraft to climb to obstacle clearance altitude). The hydraulic failure also disabled nosewheel steering, ground spoilers, and outboard flight spoilers. All of these limitations caused a combination of higher than normal approach speeds and reduced stopping ability. Finally, the crew had to select an airport (SFO was suggested by the circumstances) and execute an approach and landing under adverse circumstances. When the crew extended the flaps on approach to SFO, they received an outboard trailing-edge flap asymmetry indication resulting from the lack of protection discussed above. This condition changed the handling characteristics of the aircraft and required that crews discuss and estimate a landing speed, because none is given in the flight manuals for the condition of split inboard/outboard flaps combined with a hydraulic system failure.

2.8 Measurement

A variety of workload and performance variables were measured. Measurement instruments and procedures are described in this section.

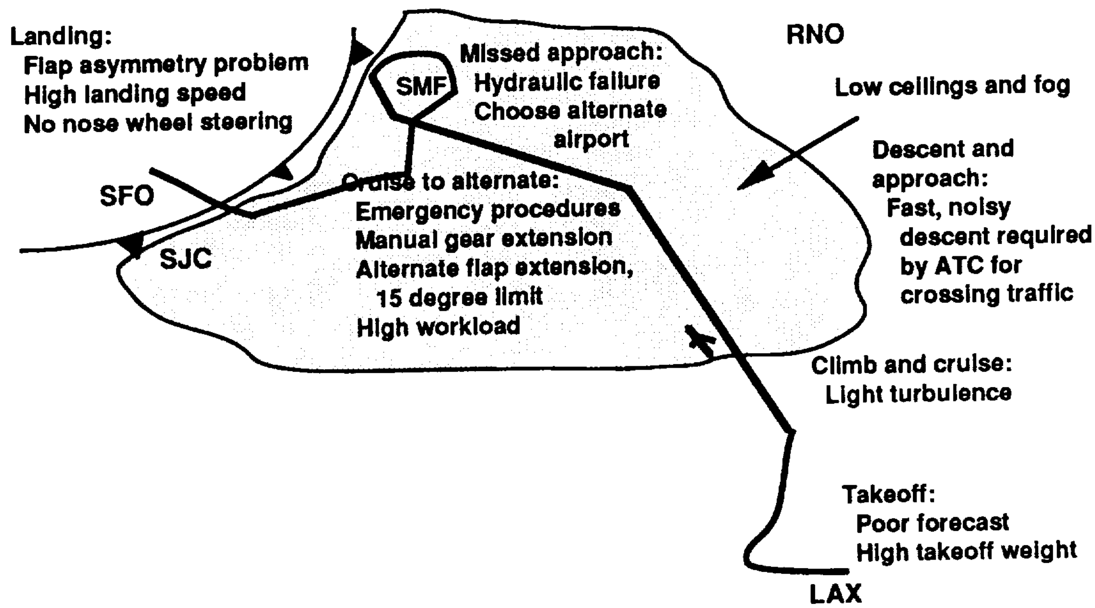

Figure 2.– Scenario for segment 5.

2.8.1 Flight Experience and Personality Data– Flight experience and personality data were collected in the pretest described earlier. Crewmembers were asked to report their age, gender, total of number of years they had been with their airline; total number of flight hours completed with their airline, in the military, in private aircraft, and any other setting; total flight hours in the B-727; and flight hours in their present crew position (see appendix B). Personality characteristics were assessed using the EPAQ, WOFO, and the A/S and I/I scales. These questions from these scales are listed in appendix C. The scales are described in appendix A.

2.8.2 Subjective Workload– Following each flight segment, all crewmembers completed the NASA Task Load Index (TLX; Hart and Staveland, 1988). This rating form consists of 6 workload-related dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. Subjects reported their responses along a seven-point Likert scales with high numbers indicating greater amounts of the dimension in question. Hart and Staveland (1988) also advocate the completion of a scale-weighting measure wherein subjects evaluate the importance of each dimension to their perception of workload in the performance setting. However, completion of the TLX by pretest crews used in the scenario development process (24 subjects) indicated that a simple summed composite of the six items correlated 0.92 with the weighted scale scores. As a result, the forced-choice procedure was eliminated and only the summed rating will be reported. Items forming the TLX are listed in appendix D.

2.8.3 Crew Performance– Crew performance data were collected from three sources: expert observation, video coding of crew errors, and computer recording of aircraft handling parameters.

2.8.3.1 Observer ratings: A recently-retired, highly experienced (39 yr in Part 121 operations, 30,750 total flight hours) airline captain served as an expert observer and was present in the simulator cab with every flight crew. He was blind to the experimental condition, and evaluated crew performance following every flight segment, and individual performance during specific phases of the high-workload segments.

The observer completed an overall rating form immediately after every segment. This form was adapted directly from Helmreich and Wilhelm's (1987) Line/LOFT Checklist, which was, in turn, derived from previous NASA simulation studies. The observer was instructed to complete the form evaluating the crew as a unit. Ratings were made on five-point Likert scales and were intended to assess the expert observer's overall impression of performance on each dimension for each segment. These ratings were summed and averaged to form a crew-level composite for each segment. Items forming this scale are listed in appendix E, and will be referred to as observer ratings of crew performance.

The observer completed a second rating form for the high workload segments, three and five. This form was adapted from the Foushee et al. (1986) study and was organized by phase of flight and abnormal events during segments three and five. Within each section of the form (e.g., preflight, taxi/takeoff, climb, and cruise), the observer rated the performance of each individual crewmember. Ratings were completed in real time as the flight progressed using a five-point Likert scale and were summed and averaged to form a composite for each crewmember during each phase of flight. Items forming this scale are listed in appendix F, and will be referred to as observer ratings of individual phase performance.

The ratings of crew and individual performance presented are limited to evaluations by a single expert observer. The psychometric ambiguities and potential biases resulting from use of a single observer or rater are a problem for this study. However, as in previous NASA investigations, we have elected to collect what may be very important data in the only manner in which it could be collected, rather than simply not collect it, and to bolster that data with information from other sources. For safety reasons, only one observer can be in the MVSRF simulator cab at a time, so inter-rater reliability could not be established for these ratings. We have chosen to err on the side of maximizing operational credibility over psychometric sophistication in these performance ratings, in a situation where maximizing both was not possible. The observer is a highly experienced pilot in terms of both operational experience and aircraft accident investigation, and as such, his evaluations would be considered highly relevant by the aviation community. Much of the information required to make these ratings is available only in real time and inside the simulator cab. The situational awareness of an experienced pilot sitting directly behind the action of a flightcrew solving a problem simply cannot be matched or replaced by the evaluations using multichannel audio and multicamera video presentations. The use of multiple observers to evaluate crews using these media would allow assessment of rater consistency, but at the expense of lost information and reduced operational credibility. We have chosen to collect and analyze this operationally significant information, recognizing its psychometric limitations. At the same time, we have attempted to compensate for these limitations by collecting data from other sources which reflect on the same domain, crew effectiveness, but capture only smaller portions.

2.8.3.2 Crew errors: Error analyses were undertaken using two independent sources of data to assure the reliability of performance assessment. First, during test runs, the expert observer kept a record of all errors he observed. The second source of error data required a complete review of the videotape records. Using these records, two condition-blind observers reviewed each flight for operational errors. The first observer was a retired pilot, the second a NASA researcher. Both observers studied the B-727 training and flight operations manuals provided by the airline prior to initiating error identification.

When an error was recognized by one or both observers, the tape was stopped and the segment containing an alleged error reviewed. The error was counted in the analysis only if both observers agreed that the error occurred and agreed on a description of the error. All errors identified by the videotape observers were then presented to the expert observer, who had the option of eliminating an error on the basis of his notes taken during the simulations or his operational experience. This was a conservative error tabulation process and assured that every error data point was reviewed at least twice. Since some performance errors were more operationally significant than others, errors were categorized according to level of severity. This process was accomplished by the expert observer and by both of the observers involved in the videotape error analysis. A three-level classification was utilized. Type 1 errors were defined as minor, with a low probability of serious flight safety consequences. Type 2 errors were defined as moderately severe, with a stronger potential for flight safety consequences. Type 3 errors were defined as major, operationally significant errors having a direct negative impact upon flight safety.

2.8.3.3 Aircraft handling: Data collected and stored by the simulator computer allowed the assessment of aircraft handling. Deviations from prescribed paths during specific flight phases (instrument approaches) were calculated for the high workload segments. Deviations from glide slope, localizer, and target approach airspeed were collected by the simulator computer once per second from the outer marker to the runway threshold on final approach during segments 3 and 5. The absolute value of the deviations per second was summed for each dimension during each approach.

RESULTS

3.1 Pretest Data

Crewmembers who participated in the study are described by their years and hours of flight experience and by their mean scores on personality scales in comparison to other samples of pilots.

3.1.1 Flight Experience– Flight experience differed, predictably, by crew position. Captains had been with their airline an average of 22.5 yr and had completed an average of 12,493 flight hours in airline operations (not necessarily all with the same carrier), with 3,259 of those hours as Captain. Captains averaged 51 yr of age. First Officers had been with their airline an average of 13.8 yr and had completed 4,011 airline flight hours, with 2,145 of those hours as First Officer. First Officers averaged 43 yr of age. Second Officers had been with their airline an average of 2 yr and had completed 2,680 airline hours, with 1,505 of those hours as Second Officer. It was not possible to determine with certainty where the Second Officer's had completed airline hours other than as B-727 Second Officers, because the background questionnaire did not request these details. However, because the B-727 was a relatively junior (lower seniority) aircraft at the airline, these hours most likely consisted of Pilot-in-Command or Copilot hours completed with another airline, perhaps a commuter or regional airline. Second Officers averaged 33 yr of age.

3.1.2 Personality Data– Average scores on each of the personality characteristics assessed are displayed in table 1. No differences among crew positions were found in the personality data. These sample means can be compared with means among pilots sampled by Chidester (1990) and norms reported by the authors of each instrument (male college students appear to be the closest comparison group reported). Comparisons of these means suggests that this sample falls somewhere between the norms for male college students and norms for pilots reported by Chidester (1990).

Table 1. Sample means and norms for personality scales

Personality characteristics by test administered	Sample	Pilots[a]	College males
Characteristics		PAQ	
Instrumentality	40.47	43.87	38.92
Expressivity	22.55	24.24	22.08
Negative instrumentality	11.89	10.30	13.69
Verbal aggressiveness	5.26	4.14	5.55
Negative communion	5.74	5.84	6.36
Characteristics		WOFO	
Mastery	20.32	22.87	19.26
Work	20.48	22.00	19.80
Competitiveness	12.31	13.18	13.63
Characteristics		JAS (derived scales)	
Achievement striving	23.73	26.17	22.89
Impatience/irritability	14.01	14.91	16.48

[a]Based on a sample of 469 pilots in domestic short-haul operations (Chidester, 1990).

Table 2 displays the mean personality scale scores of captains classified into each cluster profile for the purposes of this experiment. As would be expected given the use of an algorithm to assign pilots on the basis of previously-defined clusters, each cluster closely resembled its profile as defined by Chidester (1987) and Gregorich et al. (1989). IE+ captains were distinguished by elevated scores on instrumentality, expressivity, mastery, work, achievement striving, and unexpectedly, competitiveness. In simple terms, IE+ captains can be described as being both instrumental and expressive relative to other pilots. I– captains were distinguished by elevated scores on negative instrumentality, verbal aggressiveness, impatience/irritability, and mastery. Perhaps most importantly, I– captains scored very low on expressivity. I– captains can be described as being instrumental, but not at all expressive. EC– captains were distinguished by elevated scores on negative communion and expressivity and by very low scores on instrumentality, mastery, work, and achievement striving. EC– captains can be described as at least moderately expressive, but not at all instrumental, relative to other pilots.

Table 2. Personality scales means for each profile among captains

Personality dimensions by test administered	IE+	I–	EC–
Characteristics		PAQ	
Instrumentality	43.09	40.30	36.11
Expressivity	24.18	19.20	23.72
Negative instrumentality	10.18	14.10	11.50
Verbal aggressiveness	4.40	6.40	5.28
Negative communion	4.00	6.00	6.50
Characteristics		WOFO	
Mastery	24.09	21.30	18.39
Work	22.18	19.60	18.89
Competitiveness	14.55	12.60	11.89
Characteristics		JAS (derived scales)	
Achievement striving	24.27	24.00	22.17
Impatience/irritability	13.09	15.40	14.39

3.2 Subjective Workload

Analysis of TLX workload ratings revealed flight segment as a significant main effect ($F(4,80) = 54.26$, $p < 0.01$). Post hoc comparisons revealed that segments 3 and 5 were rated as significantly higher in workload than segments 1, 2, and 4. This served as a manipulation check, indicating that the test segments required greater effort than the normal segments. Mean workload ratings for each segment are presented in table 3. While this analysis suggested no impact of captain personality on average crew workload ratings, it has been argued that agreement or disagreement among crewmembers in workload assessment may be sensitive to differences in crew coordination (R. D. Blomberg, personal communication). However, disagreement contributes only to error variance in a repeated measures analysis. An index of disagreement among crewmembers was calculated for each flight segment. Analysis of disagreement revealed a significant flight segment by captain personality interaction ($F(8,80) = 2.27$, $p < 0.05$). Simple-effects tests suggested that crews led by IE+ captains disagreed more concerning workload in both segments 3 and 5 than did crews led by I– or EC– captains. Follow up analyses using analysis of covariance revealed that this was attributable to disagreement between IE+ captains and their subordinate crewmembers. Specifically, IE+ captains rated segments 3 and 5 as less demanding than did their subordinates, but their subordinates rated workload the same as subordinates or captains of I– or EC– crews.

3.3 Observer Ratings

Analyses of the observer's ratings of crew performance during the full-mission segments revealed a significant interaction between leader personality and flight segment ($F(8,80) = 2.80$, $p < 0.01$). Means for each group of captains during each segment are presented in Table 4 and these results are presented graphically in Figure 3. Examination of these means revealed that crews led by IE+ captains were rated as consistently effective, and these ratings were higher than the other crew types for the segments overall (though not every comparison for every segment was statistically significant). Crews led by EC– captains were rated as consistently less effective over all segments than those led by IE+ captains (though not all comparisons were statistically significant). Crews led by I– captains received ratings that varied considerably across segments. For segments one, two, and three, I– led crews were similar to EC– crews; they were rated as less effective than IE+ crews. However, on segment five, I– led crews were rated as performing as well as IE+ led crews, and significantly more effectively than EC– led crews.

Table 3. Average ratings of workload

Segment	Task load index score
Segment 1	2.97
Segment 2	3.27
Segment 3	4.51[a]
Segment 4	2.69
Segment 5	4.56[a]

[a]Means for segments 3 and 5 differ from all others at the 0.05 level.

Table 4. Observer ratings of crew performance

	Captain Personality profile		
Segment	IE+	I–[a]	EC–
Segment 1	3.85	3.40	3.23
Segment 2	3.88	3.55	3.59
Segment 3	3.95[b]	2.97	2.90
Segment 4	4.37	3.79	3.12
Segment 5	4.22	3.98	2.90[b]

[a]The following within-crew comparisons are significant among I– led crews: 5 vs. 1, 2, and 3; 3 vs. 4.

[b]Indicates significant between-crew differences

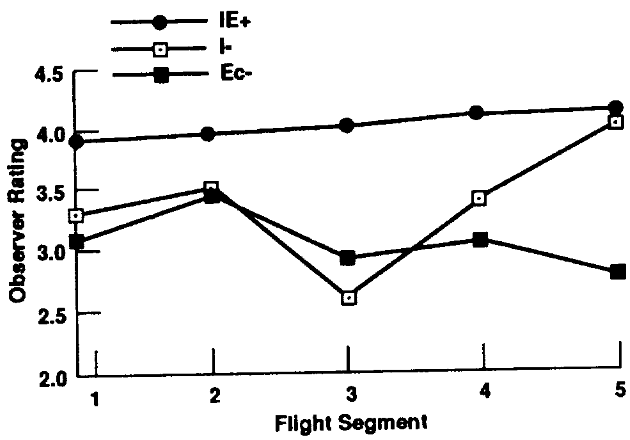

Figure 3.– Observer ratings over five flight segments

Observer ratings of individual Phase Performance during the high workload segments were consistent with the crew-level ratings. That is, crews led by IE+ captains were rated as outperforming I– and EC– led crews throughout segment three ($F(2,20) = 3.69$, $p < 0.05$, post-hoc comparisons via Tukey's Honestly Significant Difference (HSD)), and EC– crews were rated as performing more poorly than IE+ or I– led crews on segment five ($F(2,20) = 4.27$, $p < 0.05$). There were mean differences in ratings between phases, but no reliable differences among crew positions nor interactions between phase or position and captain personality profile. Means for each phase are displayed for segment three in table 5 and for segment five in table 6.

Table 5. Observer ratings of phase performance—segment 3

	Captain personality profile		
Phase of flight	IE+	I–	EC–
Prestart	3.93	3.62	3.51
Taxi/takeoff	3.83	3.59	3.42
Climb	3.69	3.38	3.40
Cruise	3.69	3.24	3.04
Stabilizer	3.46	2.72	3.00
Oil pressure	3.63	2.99	2.51
Approach	4.15	3.55	3.27
Mean	3.76[a]	3.29	3.16

[a]Mean for IE+ differs from I– and EC– at the 0.05 level.

Table 6. Observer ratings of phase performance—segment 5

	Captain personality profile		
Phase of flight	IE+	I–	EC–
Prestart	4.42	3.79	3.40
Taxi/takeoff	4.30	3.64	3.41
Climb	4.39	3.94	3.41
Cruise	4.23	3.59	3.44
Missed approach	4.26	4.13	3.25
Hydraulic failure	4.14	3.91	3.24
Cruise to Altitude	3.96	3.89	3.12
Approach	3.63	3.99	3.06
Mean	4.16	3.86	3.29[a]

[a]Mean for EC– differs from IE+ and I– at the 0.05 level.

3.4 Error Analyses

A total of 998 errors were identified by the video observers across the 23 crews. Of that number, 85 (8.5%) were eliminated by the expert observer as logical choices made by the crew in response to available information. Of those errors that were eliminated, 60 (71%) were initially rated as type 1 errors by the video observers; the remainder were initially rated as type 2 errors. None of the eliminated errors was initially rated as a type 3.

A consensus severity classification was reached by the video and expert observers for the remaining 913 errors. Table 7 shows the relative frequency of each type of error. Examples of type 1 errors included: failure of the pilot not flying to make and announce a crosscheck dictated by company policy, incorrect briefing of a missed approach procedure which was immediately corrected by the pilot not flying, and failure to notice an "expect" provision on a standard arrival procedure. Examples of type 2 errors included: delayed discovery of a system failure (hydraulic loss) following aural and visual indications of the failure, flying at airspeeds above or below those required for the aircraft configuration, and failure to use a proper climb airspeed to make an ATC crossing restriction imposed due to intersecting traffic. Examples of type 3 errors included: airspeed deviations resulting in stick-shaker (stall) or mach-overspeed warnings, identification and initiation of an incorrect checklist or failure to run a required emergency checklist (jammed stabilizer) resulting in an improper landing configuration, and failure to consider an alternate airport other than that filed (SJC) or to determine alternate weather following a weather-induced missed approach.

Table 7. Frequency of errors by severity classification

Flight segment	Type 1	Type 2	Type 3
Segment 3	168	137	81
Segment 5	242	161	124

Error analyses revealed a pattern of findings similar to that seen for observer ratings. A repeated measures analysis including type 1, 2, and 3 error counts for each day, with severity level entered as a design factor, revealed three main effects. First, EC– led crews tended to make more errors than IE+ or I– led crews ($F(2,20) = 4.03$, $p < 0.05$; means = 23.2, 17.7, and 18.3 per day, respectively). Second, as is apparent in table 7, crews tended to make more type 1 (minor) than type 2 (moderate) errors, and more type 2 than type 3 (major) errors ($F(2,40) = 22.60$, $p < 0.01$; means= 9, 6.6, and 4.5 per day, respectively). Third, crews tended to make more errors on day two than on day 1 ($F(1,20)=20.64$, $p < 0.01$; means= 16.5 and 22.9, respectively). However, this difference between the two days appeared to be entirely attributable to a difference in the length of segments three and five ($F(1,20) = 151.75$; $p < 0.01$; means 66.5 and 100.1 minutes, respectively). An analysis of covariance, that controlled for this time difference eliminated the day effect, but left the captain personality ($F = 4.05$) and error severity ($F = 22.6$) main effects intact. This indicates that the error rates were statistically equivalent on both days; given equal flight times, equivalent numbers of errors would occur.

While error analyses were consistent with ratings by the expert observer in discriminating performance of EC– led crews from IE+ or I– led crews, they did not reveal a change in I– led crew performance over the course of the simulation. That is, the captain personality by flight segment interaction, which was significant for expert observer ratings, was not significant for crew errors. One possible explanation for this discrepancy is that in the process of systematically rating crew performance, observers may (and probably should) take more than errors into account. This might be expected, since the observer has access not only to errors, but also to the crew communications process which is significantly related to crew performance (Kanki, Lozito, and Foushee, 1989).

3.5 Aircraft Handling

Deviations from glide slope, localizer, and target approach airspeed were collected by the simulator computer once per second from the outer marker to the runway threshold on final approach during segments 3 and 5. The absolute value of the deviation per second was summed for each dimension during each approach. These sums were submitted to a multivariate analysis of variance, which revealed a main effect for flight segment (multivariate approximate $F(4,17) = 7.55$, $p < 0.01$). Univariate analyses revealed that this difference was due to greater glide slope deviation on segment three than segment five ($F(1,20) = 22.25$, $p < 0.01$). Follow-up analyses indicated that this effect was accompanied by greater airspeed variability on segment three. These deviations could be expected to result from operation of an aircraft with a jammed stabilizer, the problem presented to the crew on segment three. Aircraft handling appeared to be unrelated to captain personality.

DISCUSSION

These results present a relatively consistent picture of the impact of leader personality on crew performance, particularly in critical high-workload situations. As might be expected, consistently effective performance was found among crews led by IE+ captains and less effective performance among crews led by EC– captains. Also, IE+ captains rated their workload during segments three and five as lower than did their own crews or I– or EC– leaders and crews. This finding may be a clue to understanding all the other results, but it requires the testing of hypotheses based upon process data. Specifically, IE+ captains may evaluate the abnormal segments as less demanding due to specific strategies they apply in task delegation, or they may simply adopt differing standards of evaluating workload. Since we predicted that IE+ captains would better organize both people and tasks, a finding of differing strategies would strongly validate this theoretical conceptualization. If IE+ captains do indeed delegate tasks more effectively, one would expect their workload to be lower, but not necessarily that of their subordinates. These junior officers may have a number of tasks to accomplish equivalent to those led by less effective captains, yet the crew may have a greater capacity to deal with that workload. Discriminating between these two competing explanations will require further research into the process by which these crews accomplished their tasks.

Somewhat surprisingly, I– led crews performed comparably with IE+ crews on the second day and apparently improved over time. From previous research (Foushee et al., 1986), one might expect that increasing crew familiarity would result in better crew performance in the later flight segments, regardless of crew type. Instead, familiarity apparently facilitated performance only among I– led crews.

These findings are intriguing because they support logical hypotheses about crew effectiveness and are consistent with recent research examining individual pilot performance. Foushee and Helmreich (1988) have hypothesized that relatively high levels of both instrumental and expressive traits might facilitate crew performance. IE+ individuals have elevated levels of both of these traits, and as noted earlier, Chidester (1987) found that IE+ pilots appeared to benefit most from CRM training. This group showed the most positive and enduring attitude change up to six months after participating in the training program. Results from the present study suggest that personality factors, in general, may contribute significantly to crew effectiveness, and provide further support for the notion that both instrumentality and expressivity are important predictors of team performance in aerospace environments.

Somewhat surprising was the lack of relationship between captain personality and approach deviations. In the Foushee et al. (1986) simulation study, approach deviations were sensitive to differences in crew familiarity, just as were ratings and errors. This would appear to argue for a reliable correspondence between approach smoothness and crew errors or observer ratings, and lead to an expectation of similar patterns of personality effects in the handling data. One could argue that at the least, familiarity influenced approach smoothness, since segment five approaches were smoother than segment three approaches. But this explanation is confounded in the experiment with the jammed stabilizer manipulation, and the handling differences identified would appear to be attributable to that manipulation. That approaches flown in this study were not influenced by experimental variables may be due to a number of factors, ranging from the aircraft type to the meaning of personality variables. This study utilized a B-727 simulator, while Foushee et al. (1986) used a B-737 simulator. The variety of differences between these two aircraft types might explain differences in findings concerning aircraft handling. Additionally, personality characteristics should not be expected to influence all aspects of performance. Aircraft handling examines only the most individual physical and technical portions of the flying task, while the personality battery emphasizes differences in motivation towards organizing and performing tasks. Predictions of influence should be carefully defined to correspond to these motivational and organizational portions of the task.

Perhaps this is the primary lesson to be drawn from decades of personality research in aviation (cf. Dolgin and Gibb, 1988): that the variety of tasks comprising crew performance may be differentially influenced by a number of factors. These include characteristics of individuals (i.e., skill, personality, and experience), their organizations (military or airline), their mission (tactical, commuter, or long-haul), training (technical proficiency and crew coordination) and task design (old vs. new technology). For example, while the literature on personality effects on pilot performance is overwhelmingly dominated by failures to demonstrate reliable linkages, most of this research examines only whether personality predicts completion of initial pilot training. The variety of other possible links between personality and performance-relevant factors have been left relatively unexplored. One exception is that personality characteristics appear to reliably differentiate those who are attracted to aviation from the general population (Fry and Reinhardt, 1969; Novello and Youssef, 1974a, 1974b). The present study suggests that personality may play a critical role in day-to-day and, especially high-workload task performance. Characteristics of the captain and probably the crew impact the effectiveness with which the crew plans and responds to inflight problems. Perhaps aviation psychologists interested in personality can now begin to distinguish more precisely where and when these characteristics influence performance outcomes.

At least two other issues are raised by the results: defining the limits of crew familiarity effects, and devising strategies for conducting operationally valid research. The remainder of this discussion highlights those issues.

4.1 Crew Familiarity

The familiarity effect among I– led crews raises a number of important questions. Why was the pattern of lower performance levels seen in the observer ratings not reflected in the errors committed by I– led crews on the first day? The expert observer saw these crews as relatively ineffective, giving them the lowest average ratings for segment three. However, these crews made no more errors than IE+ led crews. One reason may be that process observers are, by definition, integrating more than errors into their observations in any group task situation. In previous studies (Ruffell-Smith, 1979; Foushee and Manos, 1981; Foushee et al., 1986; Kanki, Lozito, and Foushee, 1989), patterns of flight crew communications were significantly related to crew performance. As a result, we have incorporated communications dimensions into our observer rating scales. These scales require the observer to use his experience and professional judgment to evaluate how crews make decisions, handle inter- and intracrew communications, prioritize problems, deal with distractions, and distribute workload. In short, these ratings seek to evaluate the process by which crewmembers coordinate their activities. Since problems of crew coordination do not always produce observable errors, we should not expect ratings reflecting process to be perfectly correlated with performance outcomes. They are related, but they are not the same. Problems reflected in observer ratings are important in their own right because they may raise the probability that errors will be committed or not corrected quickly. We would argue that our observer ratings reflect the fact that there were significant process problems within I– crews.

This argument is also consistent with the conceptual framework proposed by McGrath (1964, 1984). In this model, the link between input variables (such as personality profiles) and group outcomes (such as errors of communication or action) is mediated by the process of group activities (such as patterns of communication). While it is possible to identify links between input and process or process and outcome variables, these relationships may diverge somewhat from input-outcome relationships. The observer ratings may be viewed as integrating both process and outcome information.

The idea that familiarity may have affected the process of crew interaction in I– crews raises another question. What behaviors were changing over the course of the simulation? There are at least three possibilities. (1) The captains may have altered their behavior, (2) the junior crewmembers may have adapted to the captain's style, or (3) the performance situations on each day may have emphasized differing elements of leadership or performance. Adaptation by subordinates may be the most plausible explanation. It is not unusual for a crew to be composed of individuals who have never met prior to the beginning of a trip. Accordingly, an adjustment period is likely during the first few flight segments, and it is probable that subordinate crewmembers often attempt to tailor their behaviors to the captain's expectations. Since the captain's role is more central to cockpit organization than other positions and because it carries final decision authority, we suspect that a captain is less likely to change his behavior to adapt to the crew (although this may occur to a lesser degree as well).

Foushee et al. (1986) demonstrated significant process differences between crews that had recently flown together vs. those that had not. The most likely explanation is that, all subordinate crew members are initially tentative in their behavior because they are awaiting signals from the leader about how he or she expects the cockpit crew to operate. In general, IE+ leaders would be expected to very quickly create an atmosphere in which open communication is encouraged. Theoretically, I– leaders would not be as likely to do so and might by nature tend to discourage questioning by subordinates. After the initial adjustment process, subordinates in I– crews may have been able to work more effectively because they knew what to expect. On the other hand, EC– led crews never seemed to make any adjustments over the course of the simulation. That I– and EC– leaders differ substantially in task motivation or instrumentality may be sufficient to explain the differences in the effects of familiarity, but tangible evidence of both change among the I– crews and differences between EC–and I– crews will have to await the process analyses of data generated by this study.

Another important idea is related to the generality of this familiarity effect and its application over time. The current study and the Foushee et al. (1986) study compared crews who had worked together for only two or three days, and familiarity seemed to provide a performance benefit in a substantial number of these cases. However, we know little about team performance over longer durations, and it is quite possible that increasing familiarity could ultimately result in worse performance. In this report, we have suggested that a number of characteristics of I– leaders might be viewed as aversive under certain circumstances. I– leaders are characterized by high levels of impatience and irritability, competitiveness, and verbal aggressiveness combined with low levels of expressivity. So far, we have shown that leaders possessing such characteristics are capable of operating in crews that perform at relatively high levels after an initial adjustment period, but it may not be possible to maintain these levels over long periods of time. Over time, we would predict that individuals possessing these attributes would have difficulty maintaining effective crew performance. If this is the case, we would be particularly concerned about crewmembers fitting this profile and participating in long-duration operations such as ship, submarine, or space station operations. The two-day time period of this simulation study was not sufficient to explore these limits, and it is important that research on longer duration flights be accomplished.

4.2 Implications for Designing Operationally-Valid Research

Foushee (1984) argued that flight simulation provides an ideal environment in which to conduct research meeting both basic and applied criteria. Since aircraft simulators provide high levels of both realism and experimental control, they make ideal laboratories for experimentation, and there are clearly fewer problems in generalizing to real-world behavior. On the negative side, high-fidelity simulation demands a large investment, and sound use of these expensive facilities demands that proposed theories or models be conceptualized with real-world applications in mind. This suggests in turn that researchers should collect evidence for the ecological validity of their theories prior to testing in a high-fidelity environment. The tradeoff identified here argues for a program of research moving from lower-fidelity to increasingly high-fidelity environments.

We believe that high-fidelity studies represent an important direction for psychological research. Helmreich (1983) has argued that researchers have overly restricted their work to the extent that it does not apply to real-world phenomena. The failure of personality researchers to demonstrate strong links with important behavioral dimensions may be in large part the result of the "sterile" laboratory tradition so predominant in psychological research. The structuring of artificial tasks for laboratory experimentation, a process viewed as necessary for both control and assessment, may also tend to create artificial conditions that account for far more behavioral variation than the experimental variables themselves.

We chose the personality battery utilized in this study in large part because of the substantial body of real-world performance-relevant data collected by Spence, Helmreich, and their colleagues (Spence and Helmreich, 1983; Helmreich, 1982, 1986; Chidester, 1987). For example, Spence and Helmreich's (1983) measures of achievement motivation were shown to predict performance among academic scientists and engineers. Moreover, Helmreich (1986) found instrumentality and expressivity to be significantly correlated with check airman evaluations of individual pilot performance. This study has provided further evidence for the validity of instrumentality and expressivity as meaningful and important components of individual personality. As we consider the development of selection criteria for future aerospace operations, these dimensions appear to be strong candidates for representation. In the absence of such preliminary validation work, the application of personality models or measures to predicting pilot performance amounts to what Ellis and Conrad (1948) referred to as a criterion shift in the absence of empirical justification. Given the cost of high-fidelity research, applying untested measures or models becomes highly risky. This phenomenon may have been a factor in the problem-plagued search for personality predictors of performance in past research. Thus, high fidelity research environments may put us in a better position to resume our search, but only if researchers accomplish preliminary validation work. Personality theory and research should move towards a closer association with real-world performance.

REFERENCES

Anderberg, M. R.: Cluster Analysis for Applications. New York: Academic, 1973.

Blake, R. R.; and Mouton, J. S: The New Managerial Grid. Houston: Gulf, 1978.

Bluth, B. J.: Soviet space stress. Science, 81, 30-35,1981.

Chidester, T. R.: Trends and individual differences in response to short-haul flight operations. Aviation, Space, and Environmental Medicine, vol. 61, Feb. 1990, pp. 132-138.

Chidester, T. R.: Selection for optimal crew performance: Relative impact of selection and training. In R. S. Jensen, ed., Proceedings of the Fourth International Symposium on Aviation Psychology, Columbus, OH, 473-479, 1987.

Dolgin, D. L.; and Gibb, G. D.: A review of personality measurement in aircrew selection. NAMRL Monograph no. 36, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1988.

Ellis, A.; and Conrad, H. S.: The validity of personality inventories in military practice. Psychological Bulletin, vol. 5, 385-426, 1948.

Fiedler, F. E.: A theory of leadership effectiveness. New York: McGraw-Hill, 1967.

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

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

Foushee, H. C.; Lauber, J. K.; Baetge, M. M.; and Acomb, D. B.: Crew factors in flight operations III: The operational significance of exposure to short-haul air transport operations. NASA TM-88322, 1986.

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 in the aviation system. NASA TP-1875, 1981.

Fry, G. E.; and Reinhardt, R. F.: Personality characteristics of jet pilots as measured by the Edwards Personal Preference Schedule. Aerospace Medicine, vol. 40, 484-486, 1969.

Ginnett, R.: The formation process of airline flight crews. In R. S. Jensen, ed., Proceedings of the Fourth International Symposium on Aviation Psychology, Columbus, OH, 399-405, 1987.

Gregorich, S.; Helmreich, R. L.; Wilhelm, J. A.; and Chidester, T. R.: Personality based clusters as predictors of aviator attitudes and performance. In R. S. Jensen, ed., Proceedings of the Fifth International Symposium on Aviation Psychology, Columbus, OH, 1989.

Hackman, J. R.; and Helmreich, R. L.: Assessing the behavior and performance of teams in organizations: The case of air transport crews. In D. R. Peterson and D. B. Fishman, eds., Assessment for decision. New Brunswick, NJ.: Rutgers University Press, 1987.

Hart, S. G.; and Staveland, L. E.: Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, eds., Human Mental Workload. Elsevier, Amsterdam, The Netherlands, 1988.

Helmreich, R. L.: Pilot selection and training. Paper presented at the annual meeting of the American Psychological Association, Washington, DC, 1982.

Helmreich, R. L.: Applying psychology in outer space: Unfulfilled promises revisited. American Psychologist, vol. 38, 1983, pp. 445-450.

Helmreich, R. L.: Studying flight crew behavior: A social psychologist encounters the real world. Invited address to the annual meeting of the American Psychological Association, Washington, DC, 1986.

Helmreich, R. L.; Sawin, L. L.; and Carsrud, A. L.: The honeymoon effect in job performance: Delayed predictive power of achievement motivation. Journal of Applied Psychology, vol. 71, 1986, pp. 1085-1088.

Helmreich, R. L.; Spence, J. T.; and Pred, R. S.: Making it without losing it: type A, achievement motivation, and scientific attainment revisited. Personality and Social Psychology Bulletin, vol. 14, 1988, pp. 495-504.

Helmreich, R. L.; and Wilhelm, J. A.: Evaluating cockpit resource management training. In R. S. Jensen, ed., Proceedings of the Fourth International Symposium on Aviation Psychology, Columbus, OH, 1987, pp. 440-446.

Jenkins, C. D.; Zyzanski, S. J.; and Rosenman, R. H.: Progress toward validation of a computer scored test for the Type A coronary prone behavior pattern. Psychometric Medicine, vol. 33, 1971 pp. 193-202.

Kanki, B. G.; Lozito, S. C.; and Foushee, H. C.: Communication Indexes of Crew Coordination. Aviation, Space, and Environmental Medicine, vol. 60, 1989, pp. 56-60.

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

McGrath, J. E.: Social psychology: A brief introduction. Holt (New York), 1964.

McGrath, J. E.: Groups: Interaction and performance. Prentice-Hall (Englewood Cliffs, NJ), 1984.

Melton, A. W., ed.: Apparatus Tests, Air Forces Aviation Psychology Program Research Reports, No. 4. Washington: U. S. Government Printing Office, 1947.

Melton, A. W.: (1957). Military psychology in the United States of America. American Psychologist, vol. 12, 1957, pp. 740-746.

Melton, R. S.: Studies in the evaluation of the personality characteristics of successful naval aviators. Aviation Medicine, vol. 25, 1954, pp. 600-604,.

Norusis, M. J.: Chapter 3: Cluster Analysis. In M. J. Norusis, ed.,. SPSS/PC+ Advanced Statistics V2.0. Chicago: SPSS Inc., 1988.

Novello, J. R.; and Youssef, Z. I.: Psycho-social studies in general aviation: I. Personality profile of male pilots. Aerospace Medicine, vol. 45, 1974a, pp. 185-188.

Novello, J. R.; and Youssef, Z.I.: Psycho-social studies in general aviation: II. Personality profile of female pilots. Aerospace Medicine, vol. 45, 1974b, pp. 630-633.

Orlady, H. W.; and Foushee, H. C., eds.: Proceedings of the NASA/MAC workshop on cockpit resource management. NASA CP-2455, 1987.

Pred, R. S., Spence, J. T.; and Helmreich, R. L.: The development of new scales for the Jenkins Activity Survey measure of the Type A construct. Social and Behavioral Science Documents, 1986.

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

Sells, S. B.: Development of a personality test battery for psychiatric screening of flying personnel. Aviation Medicine, vol. 25, 1955, pp. 35-45.

Spence, J. T.; and Helmreich, R. L.: Masculinity and femininity: Their psychological dimensions, correlates, and antecedents. University of Texas Press (Austin), 1978.

Spence, J. T.; and Helmreich, R. L.: Achievement-related motives and behavior. In J. T. Spence, ed., Achievement and achievement motives: Psychological and sociological approaches. W. H. Freeman and Co. (New York), 1983.

Spence, J. T.; Helmreich, R. L.; and Holahan, C. K.: Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behaviors. J. Personality and Social Psychology, vol. 37, 1979, pp. 1673-1682.

APPENDIX A

DEFINITIONS OF PERSONALITY SCALES

Extended Personal Attributes Questionnaire (EPAQ; Spence, Helmreich, and Holahan, 1979)

Instrumentality- a cluster of positive attributes reflecting goal-orientation and independence (active, self-confident, can stand up to pressure)

Expressivity- a cluster of positive attributes reflecting interpersonal warmth and sensitivity (gentle, kind, aware of the feelings of others)

Negative Instrumentality- negative characteristics reflecting arrogance, hostility, and interpersonal invulnerability (boastful, egotistical, dictatorial)

Negative Communion- self-subordinating, subservient, or unassertive characteristics (gullible, spineless, subordinates self to others)

Verbal Aggressiveness- verbal passive-aggressive characteristics (complaining, nagging, fussy) Negative Instrumentality- cluster of negative attributes reflecting emotional invulnerability, cynicism, and hostility (arrogant, boastful, egotistical, dictatorial)

Work and Family Orientation Questionnaire (WOFO; Spence and Helmreich, 1978)

Mastery- a preference for challenging tasks and striving for excellence ("If I am not good at something, I would rather keep struggling to master it than move on to something I may be good at")

Work- a desire to work hard and do a good job ("I find satisfaction in working as well as I can")

Competitiveness- a preference for tasks with clear winners and losers and a desire to outperform others ("It annoys me when other people perform better than I do")

Achievement and Impatience Scales (Pred, Helmreich, and Spence, 1986)

Achievement Striving- a cluster of characteristics related to hard work, activity, and seriousness in approaching work tasks ("How much does your job 'stir you into action?" "Compared to others, how much effort do you put forth?")

Impatience/Irritability- ("How easily do you get irritated?" "When a person is talking and takes too long to come to a point, how often do you feel like hurrying the person along?")

APPENDIX B

FLIGHT EXPERIENCE ITEMS

Crew Position (circle one) CAPT FO SO

How long have you been employed by your present airline? ____yrs. ____mnths.

How many hours have you logged as a pilot in the following categories?

Military __________

General Aviation __________

Airline __________

Other __________

Age __________ Sex __________

How many hours have you logged in your present crew position (Captain, First Officer)?

____

How many hours have you logged in the B-727? __________

APPENDIX C

PERSONALITY BATTERY

Extended Personal Attributes Questionnaire (EPAQ)

The items below inquire about what kind of a person you think you are. Each item consists of a pair of characteristics, with the letters A - E in between. For example:

Not at all artistic   Very artistic

A........B........C........D........E

Each pair describes contradictory characteristics -- that is, you cannot be both at the same time, such as very artistic and not at all artistic.

The letters form a scale between the two extremes. You are to choose a letter which describes where you fall on the scale. For example, if you think you have no artistic ability, you would choose A. If you think you are pretty good, you might choose D. If you are only medium, you might choose C, and so forth.

Be sure to answer every question, even if you're not sure.

Not at all aggressive   Very aggressive
A........B........C........D........E

Very whiny   Not at all whiny
A........B........C........D........E

Not at all independent   Very independent
A........B........C........D........E

Not at all arrogant   Very arrogant
A........B........C........D........E

Not at all emotional   Very emotional
A........B........C........D........E

Very submissive   Very dominant
A........B........C........D........E

Very boastful   Not at all boastful
A........B........C........D........E

Not at all excitable in a major crisis   Very excitable in a major crisis
A........B........C........D........E

Very passive Very active
A........B........C........D........E

Not at all egotistical Very egotistical
A........B........C........D.......E

Not at all able to devote self completely to others Able to devote self completely to others
A........B........C........D........E

Not at all spineless Very spineless
A........B........C........D........E

Very rough Very gentle
A........B........C........D........E

Not at all complaining Very complaining
A........B........C........D........E

Not at all helpful to others Very helpful to others
A........B........C........D........E

Not at all competitive Very competitive
A........B........C........D........E

Subordinates oneself to others Never subordinates self to others
A........B........C........D........E

Very home oriented Very worldly
A........B........C........D........E

Very greedy Not at all greedy
A........B........C........D........E

Not at all kind Very kind
A........B........C........D........E

Indifferent to others' approval Highly needful of others' approval
A........B........C........D........E

Very dictatorial Not at all dictatorial
A........B........C........D........E

Feelings not easily hurt Feelings easily hurt
A........B........C........D........E

Doesn't nag Nags a lot
A........B........C........D........E

Not at all aware of feelings of others Very aware of feelings of others
A........B........C........D........E

Can make decisions easily Has difficulty making decisions
A........B........C........D........E

Very fussy Not at all fussy
A........B........C........D........E

Give up very easily Never gives up easily
A........B........C........D........E

Very cynical Not at all cynical
A........B........C........D........E

Never cries Cries very easily
A........B........C........D........E

Not at all self-confident Very self-confident
A........B........C........D........E

Does not look out only for self Looks out only for self
A........B........C........D........E

Feels very inferior Feels very superior
A........B........C........D........E

Not at all hostile Very hostile
A........B........C........D........E

Not at all understanding of others Very understanding of others
A........B........C........D........E

Very cold in relations with others Very warm in relations with others
A........B.......C........D........E

Very servile Not at all servile
A........B........C........D........E

Very little need for security   Very strong need for security
A........B........C........D........E

Not at all gullible   Very gullible
A........B........C........D........E

Goes to pieces under pressure   Stands up well under pressure
A........B........C........D........E

Achievement and Impatience Scales

For each question below, please select the alternative that best describes yourself or your opinion. Indicate the alternative you choose by circling the appropriate letter on the scale, A, B, C, D, or E.
How much does your job "stir you into action"?

---A--------------B-------------C---------------D-------------E----

Much less than others   About average   Much more often than others

When a person is talking and takes too long to come to the point, how often do you feel like hurrying the person along?

---A--------------B-------------C---------------D-------------E----

Very frequently   Occasionally   Almost never

Nowadays, do you consider yourself to be:

---A--------------B-------------C---------------D-------------E----

Very hard-driving   Slightly hard-driving   Very relaxed and easy-going

How would your best friend or others who know you well rate your general level of activity?

---A--------------B-------------C---------------D-------------E----

Too slow, should be more active   About average   Very active, should slow down

Typically, how easily do you get irritated?

---A--------------B-------------C---------------D-------------E----

Extremely easily   Somewhat easily   Not at all easily

How seriously do you take your work?

---A--------------B-------------C---------------D-------------E----

Much less seriously than most	About average	Much more seriously than most

How often do you set deadlines or quotas for yourself at work or other activities?

---A--------------B-------------C---------------D-------------E----

Very often	Sometimes	Almost never

Do you tend to do most things in a hurry?

---A--------------B-------------C---------------D-------------E----

Not at all true	More true than not	Definitely true

Compared with others in my occupation, the amount of effort I put forth is:

---A--------------B-------------C---------------D-------------E----

Much more	About average	Much less

How is your "temper" these days?

---A--------------B-------------C---------------D-------------E----

Very hard to control	Sometimes get angry but easy to control	I seldom get angry

Compared with others in my occupation, I approach life in general:

---A--------------B-------------C---------------D-------------E----

Much more seriously	About average	Much less seriously

When you have to wait in line such as at a restaurant, the movies, or the post office, how do you usually feel?

---A--------------B-------------C---------------D-------------E----

Accept it calmly	Feel very impatient and refuse to stay long

Work and Family Orientation Questionnaire

The following statements describe reactions to conditions of work and challenging situations. For each item, indicate how much you agree or disagree with the statement, as it refers to yourself, by choosing the appropriate letter on the scale, A, B, C, D, or E. When you have decided on your answer, circle the letter that best describes your attitude. There are no right or wrong answers.
I would rather do something at which I feel confident and relaxed than something which is challenging and difficult.

---A---------------B-------------C---------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

It is important for me to do my work as well as I can even if it isn't popular with my co-workers.

---A---------------B-------------C---------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

I enjoy working in situations involving competition with others.

---A---------------B-------------C---------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

When a group I belong to plans an activity, I would rather direct it myself than just help out and have someone else organize it.

---A---------------B-------------C---------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

I would rather learn easy fun games than difficult thought games.

---A---------------B-------------C---------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

It is important to me to perform better than others on a task.

---A---------------B-------------C---------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

I find satisfaction in working as well as I can.

---A-------------B-------------C--------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

If I am not good at something I would rather keep struggling to master it than move on to something I may be good at.

---A-------------B-------------C--------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

Once I undertake a task, I persist.

---A-------------B-------------C--------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

I prefer to work in situations that require a high level of skill.

---A-------------B-------------C--------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

There is a satisfaction in a job well done.

---A-------------B-------------C--------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

I feel that winning is important in both work and games.

---A-------------B-------------C--------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

I more often attempt tasks that I am not sure I can do than tasks that I believe I can do.

---A-------------B-------------C--------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

I find satisfaction in exceeding my previous performance even if I don't outperform others.

---A-------------	B-------------	C---------------	D-------------	E----
Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

I like to work hard.

---A-------------	B-------------	C---------------	D-------------	E----
Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

Part of my enjoyment in doing things is improving my past performance.

---A-------------	B-------------	C---------------	D-------------	E----
Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

It annoys me when other people perform better than I do.

---A-------------	B-------------	C---------------	D-------------	E----
Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

I like to be busy all the time

---A-------------	B-------------	C---------------	D-------------	E----
Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

I try harder when I'm in competition with other people.

---A-------------	B-------------	C---------------	D-------------	E----
Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

It is important to me that my job offers opportunity for promotion and advancement.

---A-------------	B-------------	C---------------	D-------------	E----
Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

It is important to my future satisfaction that my job pays well.

---A--------------B-------------C---------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

It is important to me that my job brings me prestige and recognition from others.

---A--------------B-------------C---------------D-------------E----

Strongly agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Strongly disagree

APPENDIX D

TASK LOAD INDEX ITEMS AND DEFINITIONS

Mental Demand	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/High	How much physical activity was required (e.g., pushing, pulling turning, controlling, activating, etc.)?
Temporal Demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Perfect/Failure	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

All items were rated on a seven-point scale.

APPENDIX E

ITEMS COMPRISING THE OBSERVER'S RATING OF CREW PERFORMANCE

Items scored "not at all" to "very much" on a five-point scale:

Communications were thorough, addressing coordination, planning, and problems anticipated.

Open communications were established among crewmembers.

Timing of communications was proper.

Active participation in decision making process was encouraged and practiced.

Alternatives were weighed before decisions were made final.

Crewmembers showed concern with accomplishment of tasks at hand.

Crewmembers showed concern for the quality of interpersonal relationships in the cockpit.

Work overloads were reported and work prioritized or redistributed.

Crewmembers planned ahead for high workload situations.

Appropriate resources were used in planning.

Items scored on a five-point bipolar scale:

Overall vigilance	Inattentive	-	Alert
Interpersonal climate	Hostile	-	Friendly
Preparation and planning	Late	-	Well in Advance
Distractions avoided or prioritized	Poor	-	Excellent
Workload distributed and communicated	Poor	-	Excellent
Overall Crew Effectiveness	Poor	-	Excellent

APPENDIX F

OBSERVER RATINGS OF PHASE PERFORMANCE

Items (scored on a 5-point scale ranging from below average to above average):

- Procedures, Checklists, and Callouts
- ATC and Company Communications
- Planning and Situation Awareness
- Crew Coordination and Communication
- Overall Performance and Execution

Phases of flight in which ratings were completed:

Segment 3	Segment 5
Prestart	Prestart
Taxi/Takeoff	Taxi/Takeoff
Climb	Climb
Cruise	Cruise
Runaway/Jammed Stabilizer	Approach and Missed Approach
Oil Filter Bypass	System A Hydraulic Failure
	Cruise to Alternate
Approach/Landing	Approach/Landing

NASA
National Aeronautics and Space Administration

Report Documentation Page

1. Report No. NASA TM-102259	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Personality Factors in Flight Operations: Volume I. Leader Characteristics and Crew Performance in a Full-Mission Air Transport Simulation		5. Report Date April 1990
		6. Performing Organization Code
7. Author(s) Thomas R. Chidester, Barbara G. Kanki, H. Clayton Foushee (Federal Aviation Administration, Washington, DC), Cortlandt L. Dickinson (Menlo Park, CA), and Stephen V. Bowles (California Professional School for Psychology, Berkeley, CA)		8. Performing Organization Report No. A-90018
		10. Work Unit No. 199-06-12
9. Performing Organization Name and Address Ames Research Center Moffett Field, CA 94035-1000		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 20546-0001		14. Sponsoring Agency Code

15. Supplementary Notes

Point of Contact: Thomas R. Chidester, Ames Research Center, MS 262-5
Moffett Field, CA 94035-1000
(415) 604-5785 or FTS 464-5785

16. Abstract

Crew effectiveness is a joint product of the piloting skills, attitudes, and personality characteristics of team members. As obvious as this point might seem, both traditional approaches to optimizing crew performance and more recent training development highlighting crew coordination have emphasized only the skill and attitudinal dimensions. This volume is the first in a series of papers on this simulation. A subsequent volume will focus on patterns of communication within crews. This paper reports the results of a full-mission simulation research study assessing the impact of individual personality on crew performance. Using a selection algorithm described in previous research, captains were classified as fitting one of three profiles along a battery of personality assessment scales. The performances of 23 crews led by captains fitting each profile were contrasted over a one-and-one-half-day simulated trip. Crews led by captains fitting a positive Instrumental-Expressive profile (high achievement motivation and interpersonal skill) were consistently effective and made fewer errors. Crews led by captains fitting a Negative Expressive profile (below average achievement motivation, negative expressive style, such as complaining) were consistently less effective and made more errors. Crews led by captains fitting a Negative Instrumental profile (high levels of competitiveness, verbal aggressiveness, and impatience and irritability) were less effective on the first day but equal to the best on the second day. These results underscore the importance of stable personality variables as predictors of team coordination and performance.

17. Key Words (Suggested by Author(s)) Crew coordination, Personality factors, Pilot performance		18. Distribution Statement Unclassified-Unlimited; Subject Category – 03	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 42	22. Price A03

NASA FORM 1626 OCT 86

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