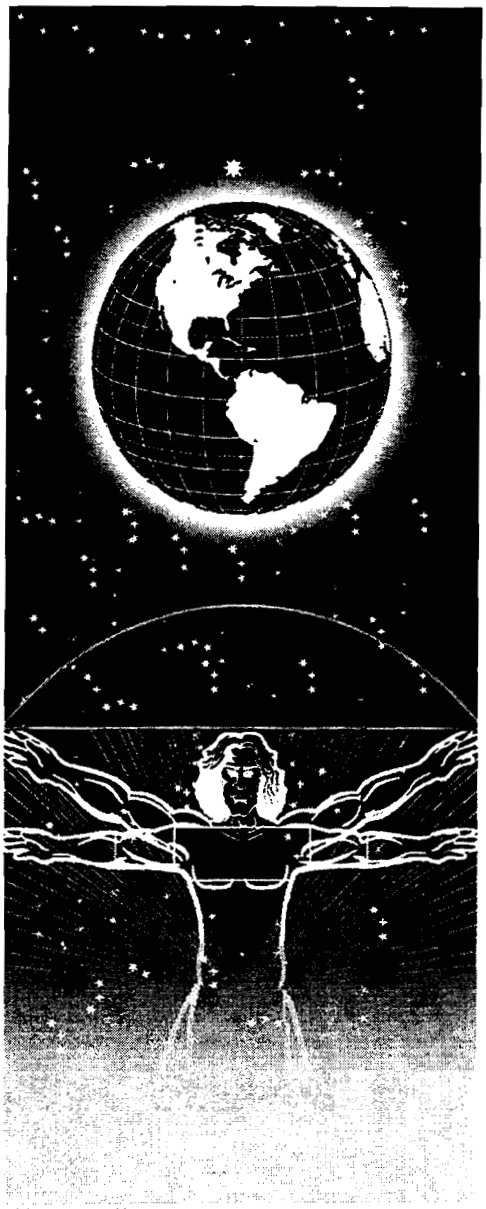
AL/HR-TR-1997-0015



UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

ANNOTATED BIBLIOGRAPHY OF RESEARCH RELEVANT TO THE DEVELOPMENT AND VALIDATION OF THE SITUATIONAL TEST OF AIRCREW RESPONSE STYLES INVENTORY

Kenneth T. Bruskiewicz

Kristi K. Logan

Jerry W. Hedge

Mary Ann Hanson

Personnel Decisions Research Institutes
100 South Ashley Drive, Suite 1230
Tampa FL 33602

August 1997

AIR FORCE MATERIEL COMMAND
ARMSTRONG LABORATORY
HUMAN RESOURCES DIRECTORATE
AIRCREW TRAINING RESEARCH DIVISION
6001 South Power Road, Building 561
Mesa AZ 85206-0904

NOTICES

This research was conducted under the Small Business Innovation Research (SBIR) Program as a Phase II effort.

The views expressed in this report are those of the authors and do not necessarily reflect official views of the US Air Force or the Department of Defense.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

FREDERICK M. SIEM
Contract Monitor

DEE H. ANDREWS
Technical Director

LYNN A. CARROLL, Colonel, USAF
Chief, Aircrew Training Research Division

Please notify AL/HRPP, 7909 Lindbergh Drive, Brooks AFB, TX 78235-5352, if your address changes, or if you no longer want to receive our technical reports. You may write or call the STINFO Office at DSN 240-3853 or commercial (210) 536-3853.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1997	Final - May 1994 to May 1996

4. TITLE AND SUBTITLE

Annotated Bibliography of Research Relevant to the Development and Validation of the Situational Test of Aircrew Response Styles Inventory

5. FUNDING NUMBERS

C - F41624-94-C-5004
PE - 65502F
PR - 3005
TA - HM
WU - 41

6. AUTHOR(S)

Kenneth T. Bruskiewicz; Kristi K. Logan; Jerry W. Hedge; Mary Ann Hanson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Personnel Decisions Research Institutes
100 South Ashley Drive, Suite 1230
Tampa FL 33602

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Human Resources Directorate
Aircrew Training Research Division, Aircrew Performance Branch
7909 Lindbergh Drive
Brooks Air Force Base TX 78235-5352

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AL/HR-TR-1997-0015

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr Frederick M. Siem (210) 536-6806
This research was conducted under the Small Business Innovation Research (SBIR) program as a Phase II effort.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This annotated bibliography was constructed as a resource aid to be used in the development and validation of the Situational Test of Aircrew Response Styles (STARS). It reviews research reported before approximately 1993 and had three primary goals to identify: (a) those personality or interpersonal skills constructs relevant for performing effectively as an aircraft commander; (b) which of these constructs could be measured using a situational judgment test (SJT) format; and (c) criteria appropriate for validating a test such as the STARS. Research reports from three general areas were targeted for inclusion in this bibliography: (a) pilot personality; (b) situational judgment tests; and (c) crew resource management. Both computerized and manual literature searches were conducted. The computerized searches were conducted in several different databases: (a) Books in Print; (b) ERIC; (c) the National Technical Information Service (NTIS); (d) Psychological Abstracts; (e) Dissertation Abstracts; and (f) the Government Printing Office (GPO) Publications Reference File. The Defense RDT&E On-Line System (DROLS) was also used to access the Defense Technical Information Center technical report database. In addition, the reference sections of all relevant articles were manually searched. Using these search methods, 93 pilot personality, 27 situational judgment test, and 150 crew resource management articles were identified and reviewed.

14. SUBJECT TERMS

Aircrews; Bibliography; Crew resource management; CRM; FY94 SBIR Program; Military; Phase II SBIR; Pilot personality; Pilot selection; Pilot training; Situational judgment; Training

15. NUMBER OF PAGES
170

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION ABSTRACT
Unclassified	Unclassified	Unclassified	UL

Table of Contents

PREFACE

This report documents work performed by Personnel Decisions Research Institutes, Inc. for the Air Force Armstrong Laboratory's Aircrew Training Research Division, Aircrew Performance Branch (AL/HRAA) located at Brooks Air Force Base, Texas, under Small Business Innovation Research (SBIR) Contract F41624-94-C-5004. This effort was conducted under SBIR Work Unit Number 3005-HM-41, Computer-Based Assessment of Pilot Personality. The laboratory contract monitor was Dr Frederick M. Siem. The contractor principal investigator was Mr Jerry W. Hedge.

The authors would like to thank Patti Haas for her assistance in preparing this manuscript.

ANNOTATED BIBLIOGRAPHY OF RESEARCH RELEVANT TO THE DEVELOPMENT AND VALIDATION OF THE SITUATIONAL TEST OF AIRCREW RESPONSE STYLES INVENTORY

I. Introduction

This annotated bibliography was constructed as a resource aid to be used in the development and validation of the Situational Test of Aircrew Response Styles (STARS). It reviews research reported before about 1993 and had three primary goals to identify: (a) Those personality or interpersonal skills constructs relevant for performing effectively as an aircraft commander; (b) which of these constructs could be measured using a situational judgment test (SJT) format; and (c) criteria appropriate for validating a test such as the STARS. Briefly, the STARS is a situational judgment test designed to measure the interpersonal effectiveness and personality attributes most critical for success as an Air Force pilot. The STARS will present respondents with realistic but difficult aircrew situations, and five possible responses in each situation. Respondents will be asked to identify the one response that would be most effective and the one response that would be least effective in that situation. Research reports from three general areas were targeted for inclusion in this bibliography: (a) Pilot personality; (b) situational judgment tests; and (c) crew resource management. Both computerized and manual literature searches were conducted. The computerized searches were conducted in several different databases:

1. Books in Print
2. ERIC
3. National Technical Information Service (NTIS)
4. Psychological Abstracts
5. Dissertation Abstracts
6. Government Printing Office (GPO) Publications Reference File.

The Defense RDT&E On-Line System (DROLS) was also used to access the Defense Technical Information Center (DTIC) database. In addition, the reference sections of all relevant articles were manually searched in order to ensure the inclusion of any articles which may not have been identified as a result of the computerized searches. Using these search methods, 92 pilot personality, 27 situational judgment test, and 151 crew resource management articles were identified and reviewed. For each annotation, a brief description of the article is presented. In addition, all predictor and criterion measures are identified, along with corresponding sample sizes (or significance values) and obtained validity coefficients. However, more emphasis has been placed on describing relevant personality constructs and measures, as opposed to other types of measures (e.g., cognitive ability, psychomotor). In addition, an executive review briefly summarizing the literature from each of the aforementioned sections is provided.

II. Pilot Personality Research Reviews

Executive Review

The prediction of pilot performance has played a prominent role in the military research and development arena for most of this century. In addition, military researchers in this area have a long and rich history of examining the relationships between personality measures and pilot performance. Generally, most of this research has taken one of two approaches. The first has concentrated on predicting performance in undergraduate pilot training (UPT) using traditional, off-the-shelf measures of personality. Many of these personality tests have been developed for use with abnormal populations (e.g., the Minnesota Multiphasic Personality Inventory) and thus, may not be appropriate for use in a normal population, especially a group as homogeneous as pilots. The second approach has largely focused on identifying personality constructs that differentiate successful pilots from unsuccessful pilots (or from non-pilot norm groups). These two avenues of research have generated mixed results with respect to identifying persons who will be effective as aircraft commanders. Most obtained validity coefficients have been either not significantly different from zero or too small to be practically significant. The second avenue of research, however, has resulted in more encouraging findings. Several researchers (Lardent, 1991; Reztlaff & Gibertini, 1987) have identified personality constructs (e.g., Instrumentality, Extraversion) that differentiated between effective and ineffective aircraft commanders.

Although this research has resulted in some positive outcomes, most researchers have had little success in using personality measures for selecting pilots. Carretta and Siem (1988) attribute some responsibility for these poor findings to the focus on "selecting out" abnormal personalities rather than "selecting in" people who are likely to perform effectively as aircraft commanders. Another serious problem that has led to discomfort with traditional measures of personality is their susceptibility to deliberate response distortion. Such instruments generally rely on candidate self-reports of feelings, attitudes, opinions, or unverifiable personal experiences. Thus, in some personnel settings, it is likely that some candidates will distort their responses in either a negative or positive fashion. As a result of this, the military has more recently focused its efforts on tests for which the dimension of interest is not readily apparent to respondents. One such test is the Dot Estimation Test (DOT). The DOT is a behaviorally based personality instrument designed to measure compulsivity and decisiveness. The DOT has displayed acceptable test-retest reliability, but little construct validity (Lambirth, Gibb, & Alcorn, 1986). Another complaint associated with traditional personality measures is that the items often have no apparent relevance to job duties or responsibilities and are thus perceived as intrusive by many test takers.

The lack of a satisfactory criterion for validating measures of personality has also been problematic for research in this area. The vast majority of these studies have used dichotomous UPT pass/fail as the criterion. Several researchers (Dailey, 1951; Sells, 1956, 1957) have noted that the dichotomous pass/fail criterion may have restricted the obtained validities. In addition, Dailey (1951) suggested that this UPT pass/fail criterion may be mostly saturated with cognitive ability factors, such that there is little variance left over to be predicted by personality measures. These researchers have further argued that the relative importance of personality and cognitive ability factors is very likely to shift after completion of training, with the personality factors

increasing in importance. Sells (1956) demonstrated that while cognitive aptitude measures are much more highly correlated with the training pass/fail criterion than personality measures, the validity of the aptitude measures drops significantly when correlated with on-the-job performance. The validities of the personality measures remains relatively unchanged. Thus, any real gains in validity to be obtained using measures of personality may only be realized through the evaluation of these measures against on-the-job criteria. Few studies have examined the relationships between personality measures and post-UPT criteria, and even fewer studies have examined the incremental validity of personality measures above that obtained using only ability measures. It is not likely that personality measures will exhibit much incremental validity for success in UPT, because most of the variance in this criterion can be accounted for by aptitude and motivational factors. It is likely, however, that personality variables will add to the prediction of pilot performance, once the pilots have been on-the-job for an extended period of time.

Annotated References

Adams, I. C. (1985). Personality and somatotype of trainee pilots. *Psychological Reports, 56(3),* 835-840.

Article type: Group Comparison Study
Keywords: 16 PF; Civilian; Personality; Pilot Selection

These researchers administered the Cattell 16 PF (Form A) to thirty-one student aviators who already had their pilot's certificates and were enrolled in an advanced aviation course. The student pilots scored significantly higher than the 16 PF college male norm group on several scales: Dominance, Surgency, Conscientiousness, Parmis and Self-Sentient.

Adams, R. R., & Jones, D. R. (1987). The healthy motivation to fly: No psychiatric diagnosis. *Aviation, Space and Environmental Medicine, 58,* 350-354.

Article type: Review
Keywords: Military; Motivation; Personality; Psychiatric

This article reviews theoretical approaches for distinguishing a "normal" from a "neurotic" motivation to fly. The purpose of this review was to develop a definition of a "healthy" motivation to fly that could aid in decisions concerning whether or not grounded pilots should be returned to duty. The authors used a dynamic perspective and noted that, in their experience, grounded pilots were usually psychologically defended and not introspective. Because pilots have difficulty in expressing their motivations to fly, the authors concluded that the examiner's sensitivity to countertransferential feelings may be the best diagnostic method. One profile of healthy pilots showed that they tended to be first-born children who had close relationships with their fathers. Other research suggests that "successful risk-takers" should be selected, while an alternative view is that conservative compulsives may be preferable in the context of most flying

tasks. Research also suggests that an Oedipus syndrome may point to an unhealthy motivation to fly, because pilots may be motivated to seek success as an overcompensation for perceived inadequacies.

Alkov, R. A., & Borowsky, M. S (1980). A questionnaire study of psychological background factors in U.S. Navy aircraft accidents. *Aviation, Space and Environmental Medicine, 51(9),* 860-863.

Article type: Validation
Keywords: Accidents; Combat/Fighter Pilots; Military; Personality

This article describes an empirical study of the influences of crewmember personalities and life changes on accident behavior. The sample included crews from all squadrons in the U.S. Navy that had reported major aircraft accidents during a one-year period (1977-78). Cross-validation was completed using a similar sample from the following year (1979). Personality and life change data were collected using a 50-item questionnaire containing 25 personality items (e.g., judgment, maturity, leadership, professionalism) and 25 life change items, condensed from an existing inventory. The flight surgeon assigned to the accident board for each squadron filled out these questionnaires using information obtained from the commanding officer (for the personality items) and the aviator involved (for the life change items). The criterion was whether or not the aviator was judged to have played a causal role in the accident. In the sample, 91 crewmembers had played at least some causal role and 38 had not. Discriminating items included maturity, a recent engagement or marriage, recent decisions regarding the future, professionalism, difficulty with relationships, a sense of humor, quick thinking, a recent death in family and trouble in dealing with peers, superiors, or subordinates. The authors did not report the direction of the item-criterion relationships (e.g., it is not reported whether humor was positively or negatively related to playing a causal role in accidents, only that it is related). A major flaw in this study is the possibility of contamination in the personality (and some life change) reports due to knowledge of the accident. The authors called for more research with better instruments and a non-accident control group.

Ambler, R. K., Bair, J. T., & Wherry, R. J. (1960). Factorial structure and validity of naval aviator selector variables. *Aerospace Medicine, 31,* 456-461.

Article type: Validation
Keywords: AQT; Attitudes; Cognitive Ability; Military; Personality; Pilot Selection; Training Performance

This study assessed the effectiveness of the Navy's Aviation Score Sheet for selecting flight trainees. The Aviation Score Sheet consists of a total score and seven component scores: (1) a personality rating (composed of ratings on five traits, which are assigned by a senior naval officer [traits were not mentioned]); (2) the Aviation Qualification Test (AQT; cognitive ability); (3) the Flight Aptitude Rating (FAR; attitudes); (4) the Selection Board Rating (SBR); (5) past

scholastic performance; (6) credit hours in math and physics and (7) the board evaluation. Specifically, the validity of the Aviation Score Sheet for predicting a variety of criteria was examined. Four criteria were used: pre-flight ground grade (GG), pre-flight Officer-like qualities (OLQ), flight failure (FF) and voluntary attrition (VA). The personality rating correlated significantly with GG (r = .11) and VA (r = -.12). The AQT correlated significantly with GG (r = .44) and OLQ (r = .12). The FAR correlated significantly with all four criteria, validities ranged from .11 for OLQ to .43 for FF (median =.30). Neither the selection board rating nor the board evaluation were significantly related to any of the criteria. Past scholastic performance predicted GG (r = .21) and OLQ (r = .15). Finally, credit hours in math and physics were significantly correlated with GG. These seven Aviation Score Sheet component scores were also factor analyzed, along with the four criteria and a variety of other selection measures (rotation method unspecified). Five factors were retained and labeled: (1) flight ability; (2) appearance of maturity; (3) military conduct; (4) motivation to take risks; and (5) academic interest.

Ambler, R. K., & Smith, M. J. (1974). *Differentiating aptitude factors among current aviation specialties* (NAMRL-1207). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Article type: Validation
Keywords: Biodata; Cognitive Ability; Military; Motivation; Personality; Pilot Selection; Training Performance

These researchers created a special criterion measure reflecting training success for a sample of 1700 Navy aviation trainees in three pilot and four non-pilot Air Force specialties. They then factor analyzed several personality and cognitive ability measures along with several different criterion measures, in order to determine which of these tests to include in a selection and classification battery. These measures included the Guilford-Zimmerman Aptitude Survey, the Hidden Figures test and four tests from the Navy and Marine Corps aviation selection battery (including a biographical inventory). An initial factor analysis of the predictors alone yielded six factors: mechanical, spatial manipulation, perceptual flexibility, verbal intelligence, numerical intelligence and flight motivation. Other factor analyses were then conducted which included various combinations of the predictor measures and selected criterion measures (a pilot/non-pilot dichotomous variable, a training completion/attrition dichotomous variable and the special criterion measure of training success). Several principal axis factor solutions were examined. The most salient results included relatively weak loadings for verbal intelligence across all criteria, greater utility of numerical intelligence for non-pilot versus pilot specialties and large loadings for flight motivation across all criterion combinations. Based on these results, the authors suggest reducing the emphasis on the verbal intelligence component in selection and further exploring the construct of flight motivation.

Ashman, A., & Telfer, R. (1983). Personality profiles of pilots. *Aviation, Space and Environmental Medicine, 54(10),* 940-943.

Article type: Group Comparison Study
Keywords: Civilian; Combat/Fighter Pilots; EPPS; Military; Personality

This study compared the personality profiles of fighter pilots with two different groups: commercial pilots and the general public. The sample consisted of 14 Australian Air Force pilots, 67 commercial pilot trainees and 21 members of various civilian occupations. Subjects were administered all subscales of the Edwards Personal Preference Schedule (EPPS). Air Force pilots scored significantly higher than the other two groups on the Achievement, Affiliation and Nurturance subscales. Commercial pilot trainees scored *below* the general public on the Succorance and Nurturance subscales. The authors concluded that the general dimension of Sociability best summarized the differences between combat pilots, commercial pilots and the general population.

Bale, R. M., Rickus, G. M., & Ambler, R. K. (1973). Prediction of advanced level aviation performance criteria from early training and selection variables. *Journal of Applied Psychology, 58,* 347-350.

Article type: Validation
Keywords: Biodata; Cognitive Ability; Job Sample; Military; Pilot Selection; Psychomotor Ability; Training Performance

This study attempted to predict performance in advanced naval aviation training using early training performance and ability measures. The authors noted that some Navy pilots successfully completed undergraduate pilot training (UPT) only to fail postgraduate training, or more specifically, replacement air group (RAG) training. This research thus attempted to predict success in RAG training using several UPT training performance scores and several ability measures. The tasks performed in RAG training are nearly identical to those actually performed on-the-job, so the authors believed that it would be an important and appropriate criterion for evaluating selection measures. Thirty-three predictor variables were used, including cognitive/spatial ability measures, job sample measures, a variety of UPT and advanced flight training performance scores and ground school grades. The criterion was dichotomous, that is, whether the aviator passed or failed RAG training. Subjects were 592 naval jet aviator trainees. Researchers examined multiple correlations and *F* tests and then selected 15 of the 33 measures which best predicted the criterion. The initial multiple correlation coefficient for these 15 tests was .43 ($p<.01$), with a cross-validity in a different, but comparable, sample of .36 ($p<.01$). The authors discuss the results in terms of the proportion of the explained variance in RAG training success accounted for by each of the predictors: mission/combat skills (36%), flight skills (28%) and instrument skills (19%). The UPT selection tests all together (including a biographical inventory) accounted for only 6 percent of the variance in RAG success. This study is important because these researchers recognized how problematic UPT pass/fail score is as a criterion measure for validating pilot selection measures. They identified and utilized a more appropriate

criterion measure. However, the criterion still focused on training success and not job performance.

Bale, R. M., & Ambler, R. K. (1971). Application of college and flight background questionnaires as supplementary noncognitive measures for use in the selection of student naval aviators. *Aerospace Medicine, 42,* 1178-1181.

Article type: Group Comparison Study
Keywords: Background; Biodata; Military; Motivation; Pilot Selection

These researchers conducted exit interviews with naval aviation students who had voluntarily withdrawn from flight training and found that the reasons students withdrew were often unrelated to mental or physical ability. Accordingly, this study examined the effectiveness of two background questionnaires, the College Background Questionnaire (CBQ) and the Flight Background Questionnaire (FBQ), in decreasing the attrition rate and thus improving the cost effectiveness of naval flight training. The CBQ is comprised of questions about applicants' college experiences (e.g., number of times college major changed, type of school attended). The FBQ is comprised of questions about applicants' experiences with flying (e.g., actual flight experience, experience as an airline passenger). A total of 22 items from both inventories were included in this study. Subjects were 1207 aviation officer candidates who entered flight training during 1966 and 1967. Of this sample, 769 completed training and 438 attrited for various reasons. The sample was split in half to create an initial validation sample and a cross-validation sample. Two multiple regressions were conducted in the validation sample to predict training completion. The first regression used the four selection tests currently used by the Navy (i.e., the Aviation Qualification Test, the Mechanical Comprehension Test, the Spatial Apperception Test and the Biographical Inventory) and the second included the CBQ and FBQ items as well. Regression weights generated using the validation sample were then applied to the cross-validation sample to generate predicted criterion scores. These scores correlated .19 with actual training completion/attrition. A cut-off score was then selected to eliminate a maximum number of attrites while allowing for the retention of the greatest number of "completes." Using this cut-off score in the cross-validation sample, the inventory would have eliminated 34 percent of the attrites, at the expense of only eliminating 16 percent of the completes.

Bartram, D., & Dale, H. C. A. (1982). The Eysenck Personality Inventory as a selection test for military pilots. *Journal of Occupational Psychology, 55,* 287-296.

Article type: Validation
Keywords: EPI; Military; Personality; Pilot Selection; Training Performance

This research study empirically examined validities of personality constructs from the Eysenck Personality Inventory (EPI) for predicting pilot training success. Subjects were 432 British Army Air Corps (AAC) soldiers and 205 British Royal Air Force (RAF) cadets, both selected into training on the basis of cognitive ability tests. The EPI Extraversion, Neuroticism and Lie scales

were administered in the AAC sample, while only the Extraversion and Neuroticism scores were available for the RAF sample. The criterion was pass/fail in their respective basic flight courses. Results for the AAC sample showed correlations with the training criterion of -.11 for Neuroticism and .37 for Extraversion. Similar results were obtained for the RAF sample, except the Neuroticism effect size was larger than the Extraversion effect size. The authors concluded that these personality constructs showed promise for prediction, especially in light of the small observed correlations with existing selection devices. Thus, the prospects for incremental validity are good.

Bennett, G. (1983). Psychiatric disorders in civilian pilots. *Aviation, Space and Environmental Medicine, 54(7),* 588-589.

Article type: Anecdotal/Descriptive
Keywords: Accidents; Civilian; Psychiatric

This author suggests that one possible explanation for aircraft accidents is the influence of psychiatric disorders on such piloting tasks as judgment and attention. Less effective selection and stresses of business and personal affairs are seen by this author as contributing more to psychiatric illness than if a pilot is having difficulty in training due to cognitive ability. One difficulty in diagnosing psychiatric illnesses is that peers would frequently cover up for the affected pilot based on the misguided notion that they are assisting this person. This may result in the affliction going untreated and thus, the illness may progress to the point where rehabilitation becomes difficult and the prognosis poor. The author notes that airline doctors and aviation medical examiners are probably best suited to alleviate this problem.

Besco, R. O. (1991). The myths of pilot personality stereotypes. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 2* (pp. 685-693).

Article type: Review
Keywords: Pilot Selection; Personality

There has been a resurgence in the use of personality assessment in the aviation industry. Many personality profiles which have been subsequently developed have been heavily influenced by the conventional pilot stereotype. According to the author, these personality stereotypes are unable to distinguish good performers from poor performers. Aviation managers are looking to personality researchers for explanations of the causes of pilot errors and according to this author, are being misled. The author notes three factors that cause him to believe that the current state of the art personality research is flawed: (1) the lack of replication or cross validation; (2) biases and contamination in the performance evaluations; and (3) the transparency and fake-ability of the testing instruments.

Braun, P. Wiegand, D., & Aschenbrenner, H. (1991). The assessment of complex skills and of personality characteristics in military services. In R. Gal and A. D. Mangelsdorff (Eds.), *Handbook of military psychology*. John Wiley & Sons Ltd.

Article type: Review
Keywords: Background; Biodata; Cognitive Ability; Military; Motivation; Personality; Pilot Selection; Psychomotor Ability

These authors reviewed the personnel selection and measurement literatures, focusing on complex skills and personality in the military services. They discuss several issues, especially how one can evaluate the methodological quality of psychodiagnostic procedures and issues concerning when one would expect measures to demonstrate predictive validity. The review explores research pertaining to application documents, biographical data, scholastic achievement, personality inventories, ratings, interviews and computer-based tests. The authors note that, historically, biodata has played a very important role in the prediction of organizationally relevant criteria for the military. The validities for these types of predictors and a wide variety of criteria are reported. The authors conclude that reality-close assessment situations (assessment centers) hold much promise as selection measures for the military services and their demonstrated validity justifies their high development costs.

Bucky, S. F., & Ridley, S. L. (1972). California Psychological Inventory as a predictor of success in the naval flight program. *Aerospace Medicine, 43*, 971-973.

Article type: Validation
Keywords: CPI; Military; Personality; Pilot Selection

The authors state that although personality inventories had previously differentiated outstanding from average pilots, they were less successful in predicting drop-outs. The authors propose that pilots may be a homogeneous group with respect to personality and that the mental "set" they adopt when completing instruments may muddle the results. In this study, they proposed that they could change the mental set by including a condition in which respondents are told to complete the inventory "as you would like to be." The sample was composed of 315 U.S. Navy aviation officer candidates with 283 of these then taking the "as you would like to be" condition. All subscales of the California Psychological Inventory (CPI) were administered and the criterion was dichotomous (finish or voluntary withdraw from flight training). The only scale that differentiated between the finish and withdraw groups was the Communality scale. Nearly all scale scores were elevated in the "as you would like to be" condition. The authors recommended that an item analysis of the Communality scale be completed.

Carretta, T. R. (1987). *Basic attributes tests (BAT) system: Development of an automated test battery for pilot selection* (AFHRL-TR-87-9). Brooks AFB, TX: Air Force Human Resources Laboratory.

Article type: Test Development
Keywords: BAT; Cognitive Ability; Military; Personality; Pilot Selection; Psychomotor Ability; Training Performance

This report documents the development of the Basic Attributes Test (BAT) by the Air Force and provided some preliminary validation results. A detailed description is given of the BAT hardware and software, as well as descriptions of each of the tests that were currently part of the selection system when this report was written. Several personality measures were included in the system: (1) a Risk Taking test; (2) Self-Crediting Word Knowledge Test (measures self-assessment ability and self-confidence); (3) Activities Interest Inventory (measures survival attitudes); (4) the Embedded Figures Test (measures field dependence/independence); and (5) the Automated Aircrew Personality Profile (measures personality factors that had not been determined at the time of the report). Preliminary results for the personality measures were not very encouraging, with only the Self-Crediting Word Knowledge test shown to be predictive for either training outcome (pass/fail) or advanced training assignment.

Carretta, T. R. (1989). USAF pilot selection and classification systems. *Aviation, Space and Environmental Medicine, 60*, 46-49.

Article type: Validation
Keywords: AFOQT; Attitudes; BAT; Cognitive Ability; Military; Personality; Pilot Selection; Training Performance

The purpose of this study was twofold: (1) to identify research methodologies (i.e., the Basic Attributes Test; BAT) that add to the predictive validity of currently used pilot selection procedures; and (2) to determine how accurately the recommendations made by the Advanced Training Recommendation Board (ATRB; fighter/non-fighter) could be duplicated without using flight training performance data. Four hundred seventy-eight USAF officer candidates from the AFROTC and OTS who had been tested on the BAT participated. Subjects had already been chosen for Undergraduate Pilot Training (UPT) based on their Air Force Officer Qualifying Test (AFOQT) scores. The AFOQT is a paper and pencil battery that consists of 16 subtests and produces five composite scores and the BAT consists of 12 computer-administered tests that measure psychomotor skills as well as a variety of cognitive abilities, perceptual abilities, personality and attitudinal characteristics. Three regression models were evaluated against UPT final outcome and ATRB recommendations: (1) Model I included AFOQT Pilot and Navigator-Technical composite percentile scores along with the number of times the AFOQT was taken by the subject; (2) Model II included scores from the BAT subtests; and (3) Model III combined the first two models. Results showed that Model I was significantly related to both UPT ($R = .17, p < .01$) and ATRB ($R = .17, p < .05$). Subjects who took the AFOQT only once were more likely to complete UPT successfully and to be recommended for follow-on training with fighter aircraft

(FAR). The results for Model II showed that of the BAT tests, the two psychomotor skills tests demonstrated the strongest relationship to UPT ($R = .26$, $p < .01$), however, they were only marginally related to ATRB ($R = .16$, $p < .10$). The cognitive/perceptual abilities subtests were also significantly related to UPT and ATRB. For the personality/attitudinal subtests, results showed that those subjects who were more cautious on the test of self-confidence (Self-Crediting Word Knowledge) and chose fewer high risk activities were more likely to complete UPT, but these tests were not related to ATRB. Finally, the results for Model III showed the strongest relationships to the criteria (UPT, $R = .50$, $p < .01$; ATRB, $R = .44$, $p < .05$), but some of the variables did not contribute at all. Thus, stepwise regression was used to develop a simpler model. The AFOQT scores were entered first, followed by the remaining 39 variables. The final model contained 11 variables from 8 different tests. These variables included the AFOQT, both psychomotor tests, 3 of the cognitive perceptual abilities tests (Encoding Speed, Item Recognition and Time Sharing) and 2 of the personality/attitudinal tests (S-CWK, AII) ($R = .31$, $p < .01$). The author recommends using AFOQT and BAT scores to both classify and select students for UPT and advanced flight training.

Carretta, T. R. (1990). *Basic attributes test (BAT): A preliminary comparison between Reserve Officer Training Corps (ROTC) and Officer Training School (OTS) pilot candidates* (AFHRL-TR-89-50). Brooks AFB, TX: Air Force Human Resources Laboratory.

Article type: Group Comparison Study
Keywords: AFOQT; BAT; Cognitive Ability; Military

This study compared Air Force Reserve Officer Training Corps (ROTC) candidates' and Officer Training School (OTS) candidates' scores on the Air Force Officer Qualifying Test (AFOQT) and the Basic Attributes Test (BAT). The purpose was to develop an interim scoring profile for BAT score interpretation based on pilot candidates commissioned through the ROTC and OTS programs. The author focused on the 8 most promising BAT tests in a sample of 350 ROTC and 705 OTS officer candidates. The ROTC candidates completed the AFOQT sometime between their senior year in high school and their junior year in college and took the BAT the summer after their junior year in college. OTS candidates took the AFOQT during their final year in college or after completion of college and took the BAT at the beginning of their participation in a Flight Screening Program (FSP). OTS pilot candidates scored consistently higher than ROTC candidates on most of the composites and scales of the AFOQT. The ROTC group scored significantly higher only on the Math Knowledge subtest. There were no consistent or meaningful differences between the two groups' scores on the BAT.

Carretta, T. R. (1990). *Cross-validation of experimental USAF pilot training performance models* (AFHRL-TR-90-68). Brooks AFB, TX: Air Force Human Resources Laboratory.

Article type: Validation
Keywords: AFOQT; Attitudes; BAT; Cognitive Ability; Military; Personality; Pilot Selection; Training Performance

The purpose of this study was to cross-validate the results of Carretta's (1989) validation study of the Air Force Officer Qualifying Test (AFOQT) battery and the Basic Attributes Test (BAT). Subjects were 885 U.S. Air Force Reserve Officer Training Corps (ROTC) and Officer Training School (OTS) cadets, 478 of whom were included in the original validation study. The total sample of 885 cadets was randomly divided into two samples: a validation sample and a cross-validation sample (matched such that they were similar in UPT pass/fail rate). For each sample, two different regression approaches were used: (1) a stepwise approach; and (2) simultaneously entering all 19 AFOQT and BAT test scores. The two sets of results were then combined in order to compute a "best estimate" of the regression weights for the AFOQT and the BAT. Two different criteria were used. The first was the Undergraduate Pilot Training (UPT) overall pass/fail outcome and the second was a refined version of UPT pass/fail, using only those subjects that failed UPT due to Flying Training Deficiency. Because the AFOQT and the BAT were not developed to predict failures due to non-flying reasons (e.g., medical failures), it was believed they should not be included. Thus, only those results pertaining to the refined criterion will be described. When the stepwise regression weights from each half sample were cross-validated in the other sample, the resulting validity coefficients were significant in both cases (Group 1: $r = .26, p < .01$; Group 2: $r = .22, p < .01$). The final stepwise model including all cases resulted in a multiple R of .33 ($p < .01$). The results of the individual group cross-validations using the simultaneous regression approach were very similar (Group 1: $r = .25, p < .01$; Group 2: $r = .26, p < .01$), as was the simultaneous model applied to all cases ($R = .34, p < .01$).

Carretta, T. R. (1992). *Predicting pilot training performance: Does the criterion make a difference?* (AFHRL-TR-91-55). Brooks AFB, TX: Air Force Human Resources Laboratory.

Article type: Validation
Keywords: AFOQT; BAT; Military; Pilot Selection; Training Performance

The purpose of this study was to examine different procedures for generating performance criteria in order to: (a) reflect the relative quality of USAF pilot candidates based on flying performance scores and academic grades; and (b) evaluate the utility of these criteria for improving the understanding of the relationship between selection test scores and training performance. Seven hundred fifty-five USAF students between the ages of 21 and 31 years old who were completing Undergraduate Pilot Training (UPT) participated. Each subject had been administered the Air Force Officer Qualifying Test (AFOQT) and the Basic Attributes Test (BAT) prior to entry into UPT (subjects had already been chosen, in part, on the basis of their AFOQT scores). The AFOQT battery consists of 16 subtests that assessed 5 ability domains:

verbal, quantitative, spatial, perceptual speed and aircrew interests/aptitude. Fourteen of the 16 subtests were used to compute the Pilot and Navigator-Technical composite scores used in the selection of pilot candidates. The BAT consists of 8 computerized tests that assessed individual differences in psychomotor coordination (rotary pursuit, stick and rudder, compensatory tracking), information processing ability (reasoning, spatial transformation, short-term memory, perceptual speed), personality (self-confidence) and attitudes toward risk-taking. A variety of performance criteria were examined: (1) UPT final outcome (pass/fail); (2) academic grades; (3) daily flying grades; (4) check flight grades; and (5) number of flying hours. The Air Training Command (ATC) uses a weighted evaluation score based on three phases: Phase I (academic), Phase II (T-37, flying performance grades), and Phase III (T-38, flying performance grades). Several different equations were developed each dealing with the eliminees in a different way. Generally speaking, the ranking of the candidates was nearly identical for equations based on all of the criteria. For pilot training candidates, the criterion used did not make a difference as to who would have been selected. Additionally, alternative criteria demonstrated little utility for understanding the relationship between preselection personnel test scores and training performance. However, the rankings generated from the weighted evaluation scores were shown to be closely related to advanced training recommendations (fighter vs. nonfighter aircraft).

Carretta, T. R. (1992). Recent developments in U.S. Air Force pilot candidate selection and classification. *Aviation, Space and Environmental Medicine, 63*, 1112-1114.

Article type: Review
Keywords: AFOQT; Background; BAT; Biodata; Cognitive Ability; Military; Personality; Pilot Selection; Psychomotor Ability; Training Performance

This article discusses the implementation of a new system for classifying pilots. At the time this article was written, Air Force pilots were selected using a positively weighted composite of the following measures: the AFOQT pilot composite, composite psychomotor response speed scores from Mental Rotation (spatial transformation) and Item Recognition (short-term memory), tracking difficulty from Time Sharing, response speed and response choice from Activities Interest Inventory (attitudes toward risk) and previous flying experience. The Air Force had proposed to implement the Pilot Selection and Classification System (PSACS) which would change the process by which pilot candidates are selected and classified. The first plan was to replace Undergraduate Pilot Training (UPT) with Specialized Undergraduate Pilot Training (SUPT), which would classify pilot candidates into one of four major weapon systems categories (bomber, fighter, tanker or transport aircraft) using the AFOQT and the BAT. This plan was revised and classification into SUPT was to occur at the completion of T-37 training and be based on T-37 flying and academic performance, pilot candidate preferences and aircraft availability. The new plan included the Pilot Candidate Selection Method (PCSM), which includes the Pilot and Navigator-Technical AFOQT composites, BAT psychomotor scores, biographical information, information processing and personality measures. In a sample of 1,112 U.S. Air Force UPT students, the regression of these scores onto UPT final outcome (pass/fail) was 0.31. The author described this as a significant improvement in operational suitability. This system was scheduled to be operationally implemented by 1992.

Carretta, T. R., & Siem, F. M. (1988). *Personality, attitudes and pilot training performance: Final analysis* (AFHRL-TP-88-23). Brooks AFB, TX: Air Force Human Resources Laboratory.

Article type: Validation
Keywords: AFOQT; BAT; Military; Personality; Pilot Selection; Training Performance

The authors cite two possible reasons for the weak relationships that have been found in past research between personality tests and various outcome criteria: (1) the use of these tests has focused on differentiating normal individuals from abnormal individuals; and (2) these tests are prone to response biases. As a result, the Air Force has explored the use of personality tests in which the dimension being measured is not immediately apparent to the respondent (e.g., the Dot Estimation Task and Self-Crediting Work Knowledge) and which include response alternatives that are equivalent in terms of social desirability (e.g., the Activities Interest Inventory). This study examined the predictive and incremental validity of five tests included in the Basic Attributes Test (BAT) battery: the Dot Estimation Task (DOT), a measure of Risk Taking, Self-Crediting Word Knowledge, Activities Interest Inventory and the Embedded Figures Test. All of these predictors produce a reaction time measure, in addition to at least one other score. Scores from the Air Force Officer Qualifying Test (AFOQT) were also examined. Multiple regression was used to examine the relationships between these predictors (both singly and in combination) and two performance criteria [Undergraduate Pilot Training (UPT) final outcome (pass/fail) and Advanced Training Recommendation Board (ATRB) recommendation (fighter/non-fighter)]. Results showed that the AFOQT tests that contributed most to the prediction of UPT outcome were Instrument Comprehension, Aviation Information, Rotated Blocks and Arithmetic Reasoning. Tests that contributed most to the prediction of ATRB were Instrument Comprehension, Block Counting and Table Reading, while Arithmetic Reasoning and Word Knowledge contributed only marginally. The authors argue that instrument comprehension appears to be more important early in training, while information processing abilities become more important later in training. Only two scores from the BAT demonstrated good internal consistency: Dot Estimation and Risk-Taking. However, the intercorrelations between scores from the different tests were not strong. Multiple regression including all of these variables in addition to the AFOQT did not improve prediction over the AFOQT and Self-Crediting Word Knowledge alone. Interestingly, in the Self-Crediting Word Knowledge test, subjects' average performance (67.1% correct) far exceeded their expectations (39.0% correct). Also, subjects who made more correct responses also responded more quickly. The authors recommend that other measures of personality and attitudinal characteristics be evaluated for possible inclusion in a subsequent version of the BAT battery.

Carretta, T. R., & Walters, L. C. (1991). *The development of Behaviorally Anchored Rating Scales (BARS) for evaluating USAF pilot training performance* (AL-TP-1991-0022). Brooks AFB, TX: Armstrong Laboratory.

Article type: Test Development and/or Evaluation
Keywords: Military; Personality; Pilot Selection

The purpose of this study was twofold. First, the authors wanted to determine whether pilots could make reliable judgments concerning personality traits relevant for effective pilot performance. Second, they wanted to determine whether these judgments could be used to develop valid Behaviorally Anchored Rating Scales (BARS) to be used as criterion measures. Three different groups of pilots participated in various phases of this study, resulting in a total of 85 participants. A structured questionnaire approach was utilized with Group A to identify personality characteristics relevant for several flying related tasks. This resulted in the identification of eight constructs: Achievement Motivation, Aggressiveness, Stress Tolerance, Risk-Taking, Cooperativeness, Assertiveness, Leadership and Decisiveness. Group B generated critical incidents targeted toward each of these constructs. Finally, Group C performed the retranslation/allocation and scaling phase. Seventy-three critical incidents were retained to be used in the retranslation phase. Of these, SMEs were able to reach satisfactory agreement for 37 incidents. Behavioral examples for aggressiveness, decisiveness, leadership and risk-taking were often mis-classified, so these constructs were dropped from consideration for final BARS development. BARS were then developed for the achievement motivation, assertiveness, cooperativeness and stress tolerance constructs. Interested readers can find the complete scales and critical incidents presented in appendices of the report.

Carretta, T. R., Walters, L. C., & Siem, F. M. (1991). Personality assessment in proposed USAF pilot selection and classification systems. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 1148-1153). Columbus, OH: Ohio State University.

Article Type: Review
Keywords: Attitudes; BAT; Military; Motivation; Performance Ratings; Personality

These authors note that since 1955 the United States Air Force (USAF) has employed essentially the same basic approach to selecting pilot candidates. This paper describes proposed changes to this selection process. The two major changes in the system were: (1) converting from a generalized undergraduate pilot training (UPT) system to a specialized undergraduate pilot training (SUPT) system; and (2) classifying candidates into one of two major weapon system categories (bomber/fighter or tanker/transport) after completing T-37 training. Broadly, this proposed Pilot Selection and Classification System (PSACS) consists of two types of methodologies. The first methodology relies on a computerized test device (the Basic Attributes Test [BAT]) to measure individual differences in hand-eye coordination, information processing ability, personality and attitudes. The BAT currently has four subtests that measure personality. Two of these tests (Self-Crediting Word Knowledge and ABCD Working Memory) are

considered cognitive ability tests that include performance-based personality measures (e.g., self-confidence). The other two tests (Activities Interest Inventory and Aircrew Personality Profiler) are traditional self-report personality measures. These four personality measures have been shown to correlate significantly with UPT pass/fail scores, but they have not demonstrated any incremental validity over cognitive ability measures. The second methodology was a structured interview developed by the Air Training Command (ATC) designed to collect three types of information: (1) background data (e.g., academic experience); (2) motivation and self-confidence; and (3) job-related knowledge. These interview ratings have been shown to be significantly related to performance in a light aircraft, flight screening program.

Cetinguc, M. (1992). An assessment of Turkish Air Force pilots' anxiety and depression levels. *Aviation, Space and Environmental Medicine, 63*, 905-907.

Article type: Group Comparison Study
Keywords: Anxiety; Depression; Military; Stress

In this study, the author attempted to objectively measure (and compare) stress levels in pilots and non-pilots through the use of anxiety and depression scores. The sample included 345 pilots and 70 non-flying Air Force officers who had been eliminated from flight training. Subjects were administered the Speilberger State-Trait Personality Inventory (STPI) A-scale and the Zung Depression Scale (ZDS). Results showed that pilots scored significantly lower than non-pilots on both the anxiety and depression scales. The author concludes that pilots are relatively happy, have better morale and feel less stress, in spite of the dangerous nature of their profession.

Chidester, T. R., & Foushee, H. C. (1989). Leader personality and crew effectiveness: A full-mission simulation experiment. In R. S. Jensen (Ed), *Proceedings of the Fifth international symposium on aviation psychology*. Columbus, OH: Ohio State University.

Article type: Group Comparison Study
Keywords: Attitudes; Civilian; Communication; CRM; Job Performance; Leadership; Motivation; Performance Ratings; Personality; Pilot Selection; Simulation

The authors examined the impact of crew leader personality characteristics on crew performance in a flight simulation. Leaders (i.e., Captains) in this study were classified into one of three personality "types" based in part on past research. The IE+ cluster consisted of pilots who were characterized by high levels of instrumentality, expressivity, achievement striving and work and mastery orientation. Captains designated as Ec- were characterized by traits associated with tendencies to express oneself in a negative fashion and lower than average goal orientation. The third cluster was designated I- and consisted of pilots with higher levels of verbal aggressiveness, negative instrumentality and competitiveness (this profile is most like that of the authoritarian pilot). Twenty-three, three-person crews completed a full-mission simulation that consisted of five different legs of varying difficulty. The Captain of each crew was randomly selected from one of the three leader types. This classification was based on the results from a battery of

personality tests that consisted of the Personal Attributes Questionnaire, the Work and Family Orientation Questionnaire and the Achievement and Impatience scales of the Jenkins activity scale. Task workload was measured via the NASA Task Load Index (TLX). Finally, three crew performance measures were used: Expert observation, video-coding of crew errors and computer records of aircraft handling parameters. The TLX was used, in part, as a manipulation check and confirmed that legs 3 and 5 of the simulation resulted in a higher workload than legs 1, 2 and 4. Analyses of the expert observer ratings revealed a significant interaction between leader personality and flight segment. Specifically, for leg 3, the crews led by the IE+ Captains were rated as significantly more effective than the crews led by either the I- or Ec- Captains. Performance shifted somewhat for leg 5, with the crews led by Ec- Captains being rated as significantly less effective than crews led by either of the other two Captain types. The error analyses revealed that Ec- led crews tended to make more errors than either the IE+ or I- led crews. The computer records of aircraft parameters only revealed a main effect for flight segment. The authors conclude that the crews led by the IE+ Captains performed more effectively than those led by Ec- Captains. A surprising finding was that while the IE+ led crews performed better than the I- crews early in the flight simulation, by the end of the mission these crews were performing in a manner comparable to that of those crews led by the IE+ Captains. One possible explanation for this finding is that the crews gradually adapted to their Captain's leadership style and were then able to perform effectively. In addition, the personality characteristics of the Captain seemed to have the most effect on crew performance during those legs of the trip that had the highest workload.

Christy, R. L. (1975). Personality factors in selection and flight proficiency. *Aviation, Space and Environmental Medicine, 46*, 309-311.

Article type: Review
Keywords: Civilian; Military; Personality; Pilot Selection; Psychiatric

This author views personality development as a "dynamic process involving life forces, instinct drives, growth processes, life stresses, reactions to significant persons and reactions to the environment in general." The author goes on to provide a general, albeit mostly anecdotal, description of the successful pilot as intelligent, emotionally mature, stable, action-oriented, adaptable and perfectionistic. He also makes several interesting points concerning pilot selection. For example, some personality traits that are predictive of success in training (namely, perfectionism) may be maladaptive once the pilot is on the job for an extended period of time. A compulsive and perfectionistic person may have difficulty in completing complex tasks that do not lend themselves to total control (e.g., flying under instrument flight rule conditions). The author further argues that the more marginal certain "personality limitations or defects" are, the more likely they are to have a detrimental impact on performance (and safety) later on in a pilot's career. Additional characteristics of pilots and implications for performance are also discussed, but most of these were in the context of the development of phobias and treatment of pilots.

Cooper, C., & Kline, P. (1989). A new objectively scored version of the Defence Mechanism Test. *Scandinavian Journal of Psychology, 30*(3), 228-238.

Article type: Validation
Keywords: Military; Personality; Pilot Selection; Psychiatric; Stress; Training Performance

The authors conducted a predictive validity study of a newer, more objective version of the Defence Mechanism Test (DMT). The DMT involves tachistoscopically presenting successively longer exposures to some threatening picture. At each time interval, subjects are asked to draw what they thought they saw and to also write a description of it. The general logic is that high levels of "defensive activity" should be indicative of a person who might perform poorly under situations of high stress and/or cognitive demand. The authors discuss some of the past criticisms of the DMT, namely, the relatively labor intensive administration procedures, lack of demonstration of construct validity and the fact that the scoring scheme does not provide for a large number of commonly observed perceptual distortions. In the present study, the authors conducted a Q-factor analysis of the DMT in a sample of 70 male civilian Royal Air Force officer applicants. In essence, the authors were trying to develop a "profile" of successful and unsuccessful pilots based on the results of the Q-factor analysis. Scores on the DMT were compared to a dichotomous pass/fail flight training criterion two years later. Results from this study were then compared to a study including 28 college students. The authors conclude that the three factor solution is "best," but because only Factor I demonstrated predictive validity they only report results for this factor (the rest are described elsewhere). They report a mean loading on Factor I for the successful pilots of .46, whereas, the mean loading for the unsuccessful pilots (those who either failed training due to non-medical reasons or voluntarily withdrew) was .37 (difference is significant at $p < .05$). This translates into a point-biserial correlation coefficient of .46 ($p < .01$) with the pass/fail criterion in flight training. In the sample of college students, the authors report that only Factor I bore a "strong resemblance" to that found with the pilot sample.

Corey, M. A. (1988). Delta Airline's problems as a function of self-fulfilling prophecy. *Psychology: A Journal of Human Behavior, 25*(2), 59-64.

Article type: Anecdotal/Descriptive
Keywords: Civilian; Motivation; Personality; Psychiatric; Stress; Training

The author attempts to tie the recent string of near misses involving Delta aircraft to one incident in which a Delta pilot accidentally turned off the engines and did not notice his mistake until the plane was within 600 feet of crashing into the Pacific ocean. The framework for this argument comes from motivational psychology. Specifically, the author states that these other incidents were due to the fact that Delta's pilots were afraid to make mistakes, resulting in self-fulfilling prophecies. The premise is that the pilots were so fearful of making a humiliating mistake, that they paid too much attention to certain aspects of flying to the exclusion of others (e.g., ensuring they were landing at the correct airport). The author supports this interpretation using the "inverted U syndrome" from motivational psychology. This theory states that there is an optimal

level of motivation for any given task which corresponds to an inverted U on a graph. As the subjects become under- or over-motivated, their performance suffers. The author recommends that Delta management take more proactive steps to put their pilots at ease in an effort to end this streak of bad luck.

Dailey, J. T. (1951). Criteria for the evaluation of tests of motivation and temperament. In A.W. Melton (Chair), *Research planning conference on objective measurement of motivation and temperament* (Conference Report 51-3). San Antonio, TX: Human Resources Research Center, Air Training Command, Lackland Air Force Base.

Article type: Review
Keywords: Military; Motivation; Training Evaluation

The author discusses the importance of adequate criteria for the successful development of measures of motivation and temperament. Past research pertaining to the validity of noncognitive measures is criticized as having little bearing on the usefulness of objective measures of temperament and motivation because the criterion most often used has been dichotomous training pass/fail. Since this criterion is mostly weighted with aptitude factors, there is little variance left over to be predicted by non-aptitude measures. The author emphasizes that any objective measures of temperament and/or motivation need to be evaluated in terms of two relatively distinct types of criteria. First, the construct validity of any new measures must be established. More importantly, for the military's purposes, is the establishment of criterion-related validity. For this, a more refined criterion (other than a pass or fail in training) needs to be developed. The author describes various other training criteria that could be used in validity studies and suggests the application of Form 77A ratings (official USAF instrument for obtaining officer evaluations) made in the field as an appropriate measure of on-the-job performance.

Davis, R. A. (1989). *Personality: Its use in selecting candidates for US Air Force Undergraduate Pilot Training* (AU-ARI-88-8). Maxwell AFB, AL: Airpower Research Institute Force.

Article type: Validation
Keywords: Background; Biodata; Military; Personality; Pilot Selection; Training Performance

The author states that pilot candidates washed out of Undergraduate Pilot Training (UPT) between 1980-1986 at a higher rate than anytime before. This study attempted to predict whether students would pass or fail UPT using a variety of personality measures. Subjects selected for UPT came from four different sources: Air Force Reserve Officer Training Corps (AFROTC), Officer Training School (OTS), the Air Force Academy (AFA) and active duty officers. The predictor measures used included the Myers-Briggs Type Indicator (MBTI), a 12-item biographical questionnaire, a Social Desirability (SD) Scale, the Extended Personal Attributes Questionnaire (EPAQ), the Work and Family Orientation Questionnaire (WOFO) and Locus of

Control (LOC). All 250 items from these instruments are presented in Appendices. Of the 1550 subjects who were mailed questionnaires containing the predictor measures, 666 returned usable questionnaires. Responses were analyzed using analysis of variance (ANOVA) and few significant differences were found in scores on the 15 personality scales between those who successfully completed UPT and those who did not. There were differences between the groups on the assertiveness, extroversion/introversion and sensing/intuition scales, but the author concludes that these score were not practically different across the two groups. The multiple correlation calculated using assertiveness, extroversion/introversion and mastery motivation as predictors and UPT pass/fail as the criterion was .23 (no significance test is reported). The author concludes that these personality measures would not be useful for selecting candidates for UPT. However, they may be useful for assessing the student's "preferred learning style." Students could then be matched with instructors, thus improving the UPT educational process. The author also recommends further analyses using the 12-item biographical inventory and using advanced training assignment recommendations as a more refined criterion measure.

Dolgin, D. L., & Gibb, G. D. (1988). *A review of personality measurement in aircrew selection.* (NAMRL-Monograph-36). Pensacola, FL: Naval Aeromedical Research Laboratory.

Article type: Review
Keywords: Civilian; Military; Motivation; Personality; Pilot Selection; Training Performance

The authors reviewed research pertaining to the personality factors used to predict performance in aviation. With few exceptions, personality constructs have not predicted success in primary flight training as well as other measures have (e.g., cognitive ability). Guilford (1947) attributed this failure to three factors: (1) the tests were not designed to predict flight performance; (2) motivational factors compensated for weakness in personality traits during training; and (3) subject biases yielded inaccurate measures of the personality traits under study. The authors discuss some other difficulties with this research. First, many efforts to increase the predictive validity of pilot selection systems have had inherent methodological problems. Second, the use of success in flight training as a criterion is seen as problematic. Third, relatively few studies employ multiple regression models of the initial candidate selection variables, thus, it is difficult to assess whether non-cognitive variables add unique variance (above that provided by other selection measures) to the prediction of training success (or pilot performance in the long-term). Fourth, the candidate population tends to be relatively homogeneous, whereas, most personality instruments were developed for testing heterogeneous groups. The implication of this is that few personality differences exist for this population. The authors proceed to review the success of almost every existing personality measure in predicting a variety of aviator criteria. For example the PRF, 16 PF, LOC, WOFO and the SVIB have all demonstrated at least minimal validity for predicting flight training performance. The authors conclude that two directions hold the most promise for future personality testing in pilot selection: (1) computer administration of personality inventories; and (2) concealing the personality trait of interest from respondents.

English, A., & Rodgers, M. (1992). Deja vu? Cultural influences on aviator selection. *Military Psychology*, *4*(1), 35-47.

Article type: Review
Keywords: Cognitive Ability; Military; Motivation; Personality; Pilot Selection; Psychiatric; Psychomotor Ability

The authors provide an interesting discussion of the influences of national and organizational cultures on the methods used to select aviators. They reviewed the selection methods of the Royal Canadian Air Force (RCAF), the British Royal Air Force (RAF), the United States Air Force (USAF) and the German Luftwaffe. For example, the USAF is characterized as basing their selection of pilots almost entirely on numerical tests. This is supposedly a reflection of American psychologists' efforts to present their profession as a numerically based "hard science" in order to be accepted by other areas of American science. The German psychologists are characterized as slipping from the "masters of research" in the early 20th century, to the "abode of charlatans and sycophants under the Nazis," because of an environment which forced psychologists to adopt theories that were in accordance with Nazi doctrine.

Fassbender, C. (1991). Culture-fairness of test methods: Problems in the selection of aviation personnel. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 1148-1153). Columbus, OH: Ohio State University.

Article Type: Group Comparison Study
Keywords: Anxiety; Civilian; Cognitive Ability; Personality

This study examined cultural differences on several general aptitude and personality measures. The author notes that the international standardization of psychological selection tests, although very critical, can only be achieved if the selection methods employed are culturally fair. This study resulted from an attempt to develop norms for an English version of a psychological test battery, specifically focusing on two aspects of culture fairness: nationality and language differences. Subjects were scientists (15 women and 82 men) working in a variety of areas (e.g., Physics, Astronomy) in several different countries. The aptitude tests in this battery were presented in either a paper-and-pencil format or as apparatus tests, and were typical of most aptitude batteries (e.g., the battery measured attention, memory, psychomotor function, etc.). The personality inventory was the Temperament Structure Scales, which consists of 11 scales: motivation, emotional stability, rigidity, extroversion, aggressiveness, vitality, dominance, empathy, spoiltness, mobility and openness. In addition, the State Trait Anxiety Inventory and the Fear Survey Schedule (which measures phobic behavior and generalized anxiety) were administered. The results showed that there were significant differences in average scores obtained across nationalities (Northern Europe vs. Southern Europe) for scales from the FSS, the STAI, and for the openness scale of the TSS. An analysis of covariance revealed that there was a strong effect for knowledge of English and that nationality alone had only a negligible effect. The author discusses the implications of culturally biased tests for selecting aviation personnel.

Fiske, D. W. (1947). Validation of naval aviation cadet selection tests against training criteria. *Journal of Applied Psychology, 31*, 601-614.

Article type: Validation
Keywords: Background; Biodata; Cognitive Ability; Military; Motivation; Personality; Pilot Selection

This article summarizes the work conducted by various government agencies to improve the selection of aviation cadets during the period around World War II. Three different predictors were tested using three different samples, each of whom joined the service under somewhat different circumstances (e.g., before or after the war). The first predictor was the Wonderlic Personnel Test (PT), which was later to become the Aviation Classification Test (ACT). It was a group-administered intelligence test (112 items) that tapped vocabulary, following directions, arithmetic reasoning, etc. The Mechanical Comprehension Test (MCT) was the second predictor and it is a 76-item test that measures knowledge of "barnyard physics." Finally, the Biographical Inventory (BI) was used, which has 150 items that measure background variables, interests, habits, preferences, etc. The BI and the MCT were later combined to form an overall Flight Aptitude Rating (FAR). This paper contains some of the first published validity data for these tests, which are still in use today (e.g., the BI and MCT). A variety of criteria were examined including training outcome (pass/fail) both for flight school and ground school, reason for failure in flight training and number of flight hours (although validities for number of flight hours are not reported). The validities for the PT were generally low across all three samples, although validities for predicting performance in ground school were moderately high (.20 to .31; sample sizes were very small; *Ns* ranged from 24-45). The validities of the MCT for predicting flight/ground school training were moderately high (ranging from .25 to .33 across the three samples). Finally, the BI did not predict ground school failures (or "Other Types of Failures") very well, but it did do a good job of predicting flight training failures (validities ranging from .29 to .34 across these samples).

Fry, G. E., & Reinhardt, R. F. (1969). Personality characteristics of jet pilots as measured by the Edwards personal preference schedule. *Aerospace Medicine, 40*, 484-486.

Article type: Group Comparison Study
Keywords: EPPS; Military; Personality

The authors state that much data has been collected on the student military aviator, but significantly less attention has been focused on the operational military pilot. This study compared the performance of three groups (jet aviators, a general population normative sample and a normative group of college-educated males) on several psychological tests. Of 298 pilots reporting to several Carrier Replacement Air Wing squadrons, 288 participated. Subjects completed a personal history questionnaire, the Maudsley Personality Inventory and the Edwards Personal Preference Schedule (EPPS). None of the inventories are described in any detail in the article. The jet aviator group differed significantly from the EPPS General Adult Male norm group on all scales except Intraception. This group differed significantly from the EPPS College

Male norm group on all scales except for Need for Achievement and Exhibition. The EPPS has also been given to several classes of student flight surgeons and, with a few conceptually meaningful exceptions, the pilots in this study tended to score similarly to the student flight surgeons. The differences were that the student flight surgeons scored higher on Nurturance and lower on Dominance. As compared to the general male population, the jet aviators scored higher on the Heterosexuality, Dominance, Change, Achievement and Exhibition scales and lower on the Nurturance, Abasement, Deference, Order and Succorance scales.

Geist, C. R., & Boyd, S. T. (1980, August). Personality characteristics of Army helicopter pilots. *Perceptual and Motor Skills*, *51*(1), 253-254.

Article type: Group Comparison Study
Keywords: Military; MMPI; Personality

This study used the Minnesota Multiphasic Personality Inventory (MMPI) to compare the scores of 15 male Army helicopter pilots to 16 male non-pilot Army officers. Differences in mean standard scores obtained by these two groups on the MMPI scales were compared using t-tests. Results showed that pilots scored significantly higher than non-pilots on several scales: Depression, Hypochondriasis, Hysteria, Social I-E (all significant at $p < .05$) and Psychopathic Deviate ($p < .01$). However, the authors used t-tests to compare scores on across the two groups on 13 separate scales. Thus if a Bonferroni correction was used to counteract the increased probability of committing a Type I error that comes from making this many comparisons (.05/10 comparisons = .005 [this alpha estimate does not include the Lie, F and K-correction MMPI scales]), the pilots would have only scored significantly higher on the Psychopathic Deviate scale.

Gerbert, K., & Kemmler, R. (1986). The causes of causes: Determinants and background variables of human factor incidents and accidents. *Ergonomics*, *29*(11), 1439-1453.

Article type: Anecdotal/Descriptive
Keywords: Accidents; Military; Personality; Psychological Status; Stress

These authors characterize the activities of an aircrew as comprised of the following tasks: (1) searching for information; (2) receiving information; (3) information processing; (4) decision-making; (5) implementing decisions; and (6) monitoring/feedback. Further, crews have three types of "workload." The quantity and quality of incoming information constitutes the "mental workload" of the aircrew member. Physical factors (e.g., vibration, g-forces) comprise the "physical workload." Psychological stress factors (e.g., pressure from responsibility, risk-taking) comprise the "emotional workload." The purpose of this study was to explore the "causes of causes of accidents," that is, to examine which variables led up to those final behaviors which resulted in an accident (or an incident). Subjects were 1448 fighter, cargo plane and helicopter pilots who had been involved in near-accidents (incidents). Subjects completed a 315-item questionnaire that was targeted toward several key questions: when and under what conditions did the incident occur, what hazardous situation resulted and how was it corrected. Specifically,

the questionnaire was designed to collect a wide variety of information ranging from the date and time of the incident to various retrospective variables, such as the conditions surrounding the pre-flight briefing, the physical and psychological condition of the crew prior to and during the flight, the hazardous situation encountered, etc. Sixty-one error variables were collected and factor analyzed (varimax rotation) but the solution was "not satisfactory." Subsequently, a subset of the variables (27 of the original 61) were chosen to be factor-analyzed and this yielded a four-factor solution. Factor I was labeled Vigilance Errors. These errors were generally associated with intra-personal background conditions, for example, nervousness, high tension, oversaturation of information channels and carelessness. Factor II was labeled Information Processing Errors. These errors were generally associated with a lack of flying experience, task overload, inadequate briefings, etc. Factor III was labeled Perception Errors. These errors were closely associated with environmental cues such as visibility problems (e.g., due to bad weather) or terrain peculiarities. Factor IV was labeled Sensorimotor/Handling errors. These errors were mainly associated with individual difference variables such as tension, nervousness during flight, motivation to succeed (excessive), lack of flying skill and insufficient confidence. The authors conclude that many "pilot error" statements in accident reports oversimplified the actual causes of accidents.

Goeters, K. M., Timmermann, B., & Maschke, P. (1993). The construction of personality questionnaires for selection of aviation personnel. *The International Journal of Aviation Psychology, 3*(2), 123-141.

Article type: Test Development and/or Evaluation
Keywords: Attitudes; Civilian; Motivation; Personality; Pilot Selection

The authors describe the Temperament Structure Scales (TSS), a German personality instrument designed for selection in a pilot context. It was designed by Kirsch (1976) using a sample of student pilots and airline pilots. This study compared results from the TSS to results from the Jackson Personality Research Form (PRF) using a sample of 300 applicants for ab initio pilot training at a major German airline. Subjects believed that their scores on the personality inventories would be used in making selection decisions. These TSS and PRF data were factor analyzed (orthogonal rotation). The factor analysis indicated a very high degree of correspondence between the TSS and the PRF. However, "because the factors were mostly defined by several PRF scales but by only a few TSS scales," the authors conclude that the TSS was more closely constructed to a clear second order factor as compared to the PRF. They also note that the frequency distributions of almost all of the PRF scales were severely negatively skewed (i.e., ceiling effect) and therefore exhibited a faking response bias. Whereas, the scores on the TSS, while not resulting in perfect approximations of normality, tended to be rather close and were definitely not skewed to the same extremes as the PRF scales. The authors conclude that it is possible to construct personality scales that have psychometric properties that are ideal for selection (although they concede that the PRF was *not* intended for selection, but has been used in that manner). No validity data are presented.

Guinn, N., Vitola, B. M., & Leisey, S. A. (1976). *Background and interest measures as predictors of success in undergraduate pilot training* (AFHRL-TR-76-9). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Article type: Validation
Keywords: AFOQT; Attitudes; Background; Biodata; Military; Pilot Selection; SVIB; Training; Training Performance

The purpose of this study was to develop empirical keys to predict a variety of training outcomes using the Strong Vocational Interest Blank (SVIB) and the Officer Biographical and Attitudinal Survey (OBAS). Subjects were 593 pilot candidates who entered pilot training through Officer Training School (OTS). (It is interesting to note that UPT trainees from OTS represented the largest percentage of eliminees.) The SVIB has 54 occupational keys and five supplemental keys. The OBAS is a 116-item inventory which contains a variety of background and attitudinal items. It also contains the importance-possibility scale, which measures the importance of certain career needs and how likely it is that those career needs can be met by the Air Force. The predictive utility of the Air Force Officer Qualifying Test (AFOQT) score was also examined. Four criteria were used: total elimination (UPT pass/fail), Flying Deficiency Elimination (FLY), Self-Initiated Elimination (SIE) and Motivational Deficiency Elimination (SIE/MOA). Regression analyses were conducted to examine the prediction obtained using the keys and the number of pilots correctly classified as passing or failing UPT using several different prediction models. The OBAS keys demonstrated significant cross-validity only for UPT pass/fail ($R = .13$, $p < .05$) and the FLY ($R = .14$, $p < .05$) criteria. Seventeen of the SVIB occupational scales demonstrated significant zero-order validities with UPT pass/fail resulting in a multiple correlation of .45 (not cross-validated). Three different models combining a variety of these predictors were tested. Model 1 consisted of the 17 SVIB scales, the four OBAS keys and the AFOQT (cross-validity = .14, $p < .05$). Model 2 consisted of just the 17 SVIB scales and the AFOQT (cross-validity = .11 ns). Finally, Model 3 consisted of the four OBAS keys and the AFOQT (cross-validity = .14, $p < .05$). With respect to classification of pass or fail, Model 1 was the most efficient, resulting in an overall correct classification of 71 percent. The authors conclude that non-cognitive data have practical value in the selection of pilot trainees.

Hemphill, J. K. (1951). Criteria of crew effectiveness. In A. W. Melton (Chair), *Research planning conference on objective measurement of motivation and temperament* (Conference Report 51-3). San Antonio, TX: Human Resources Research Center, Air Training Command, Lackland Air Force Base.

Article type: Single Group Study
Keywords: Combat/Fighter Pilots; Military; Motivation; Training Performance

The author evaluated performance criteria for B-29 bomber crews. The criteria were simulated radar bomb releases, five optical bomb releases, target identification and navigation problems. The bomber crews were also rated by an evaluation team and observations were made in the form of flight check ratings. Specific data are not presented by the author. In general, results showed

that flight check ratings tended to have positive correlations with more objective measures of bombing performance. The author recommend developing more refined criteria. Specifically, three types of measures were suggested: measures derived from sociometric ratings, measures derived from self-reports of motivation and measures which reflect crew morale.

Hilton, T. F., & Dolgin, D. L. (1991). Pilot selection in the military of the free world. In R. Gal and A. D. Mangelsdorff (Eds.), *Handbook of military psychology*. John Wiley & Sons, Ltd.

Article type: Review
Keywords: AFOQT; Cognitive Ability; Military; Motivation; Personality; Pilot Selection; Psychiatric; Psychomotor Ability

This chapter documents the 80-year history of military pilot selection, both theory and methodology, for a variety of different countries with a focus on the U.S. The authors reviewed a great deal of research on psychomotor skill/quickness and intelligence/aptitude, but only the portions of the chapter dealing with "personality/character" are reviewed. Self-selection and rigorous physical examinations played a major role in early military pilot selection, with only the most adventurous people volunteering. During the period prior to 1919, the measurement of emotional coolness under pressure was the focus of pilot screening. Later, ingenuity and courage were added in screening for air-to-air combat, and interviews and observation were the main approaches for tapping these characteristics. From 1919 to 1938, successful combat pilots were characterized as quiet, methodical men, not given to emotional excitement, who were able to inhibit instincts of self-preservation. At this time, Germany pioneered the use graphology to form judgments of pilot characterological suitability (Wyatt & Teuber, 1944). The U.S. Army and Navy used interviews to judge psychiatric well-being, while Germany continued to emphasize observation of facial expressions and emotional reactions to stresses. From 1938 to 1945, the U.S. Army Air Force found that commercially available paper-and-pencil personality tests did not contribute to pilot selection, while European countries "emphasized screening pilots for the proper character" (at that time a term referring primarily to motivational potential, personal presence/domination and emotional calmness under stress). Interviews were common during this period and were usually guided by an applicant's answers to a paper-and-pencil biographical inventory. A great deal of research dealing with personality was conducted during the period from 1945 to 1970. Some of this research showed that personality inventories could accurately identify psychiatrically unsuitable applicants, but did not distinguish between successful and unsuccessful aviation cadets. Sells (1956) found that motivational factors were more predictive of long-term success as a pilot than of training performance. Prior to this, some countries were already screening based on nonclinical personality traits. Both Israel and Denmark screened for nonauthoritarian, egalitarian leadership style. Many European countries were also either using or evaluating the Defense Mechanism Test, which is a psychoanalytic projective instrument designed to measure stress tolerance.

Holtzman, W. H., & Sells, S. B. (1954). Prediction of flying success by clinical analysis of test protocols. *Journal of Abnormal and Social Psychology, 49*, 485-490.

Article type: Validation
Keywords: Background; Biodata; Military; Personality; Pilot Selection; Psychiatric; Training Performance

This study examined the ability of expert clinical psychologists to predict advanced flight training success using an experimental battery of psychological tests. The tests were similar to those used by clinical psychologists, but had been adapted for group administration. The researchers used an extreme groups design. Fifty successful (e.g., passed flight training, pilot stanine of six or greater) and 50 unsuccessful (e.g., did not pass training, pilot stanine less than six) subjects were randomly selected from a much larger pool of 1504 test subjects. Subjects completed the Background Information Form (a biographical history inventory), the Ink-Blot test (group adaptation of the Rorschach), Feeling and Doing (a psychosomatic inventory), What Is He Saying (a sentence completion test), the L-D test (a group version of the Szondi test) and a group administered version of the Draw-A-Person test. Nineteen psychologists were given profiles of all subjects containing information from all of these tests. They were asked to judge whether a subject would pass or fail flight training. They made these judgments on a global level using all available information, and also make ratings of how confident they were in their judgments. A subset of these psychologists also made pass/fail judgments after examining each of the tests one at a time. Results showed that none of the psychologists were able to predict the pass/fail criterion at an accuracy level much better than chance using the global approach. Even when the researchers only examined those judgments that the psychologists were most confident in, only two were able to predict at a level better than chance. The results were very similar for the subset of psychologists who made pass/fail predictions after examining each test individually.

Hunter D. R. (1989). Aviator selection. In M. F. Wiskoff and G. M. Rampton (Eds.), *Military personnel measurement: Testing, assignment, evaluation.* New York: Praeger.

Article type: Review
Keywords: Background; Biodata; Cognitive Ability; Job Sample; Personality; Pilot Selection; Psychomotor Ability; Training Performance

This chapter provides a comprehensive review of aviator selection methods from 1903 to the present. These methods are grouped into four relatively broad categories: cognitive ability tests; personality, background and interest inventories; psychomotor and information processing tests; and light-plane/job sample tests. The author reports that general ability has produced inconsistent findings with respect to predicting pilot performance, with validities reaching as high as .34, but also falling to the negative and non-significant ranges. However, instrument comprehension and mechanical comprehension have both demonstrated consistent predictive validities for success in training, ranging from .20 to .40. With respect to noncognitive predictors of flight training success, the Biographical Inventory (BI) has had the most success. In a study by Fiske (1947), correlations between a shortened version of the BI and relevant criteria ranged from .30 to .34,

and it was concluded that much of the variance in predicting success in flight training was accounted for by cognitive ability variables, leaving little variance left to be explained by noncognitive variables. Thus, there has also been an emphasis on predicting post-training adaptability criteria. This research program has demonstrated that although cognitive ability measures are much better in predicting training performance, these validities drop to the same level as noncognitive measures when the criteria are measured later in the pilot's career. The validities for the non cognitive measure remained relatively constant around time (around .11). Psychomotor and information processing tests have remained quite valid as predictors of flight training success over the past 50 years. Validities have ranged from -.02 to .49, although there has been a trend for these validities to steadily drop. This may not be due to deficits in the design of current tests, but due to a decrease in the importance of psychomotor coordination in flying. The author also reports that most studies examining the relationship between performance in a light plane screening program or in a simulator and performance in flight training have found that these tests are highly reliable and valid, albeit very expensive to develop and administer.

Hunter, D. R., & Burke, E. F. (September, 1990). *An annotated bibliography of the aircrew selection literature* (Research Report 1575). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Article type: Review
Keywords: Anxiety; Civilian; Cognitive Ability; Depression; Job Performance; Job Sample; Military; Personality; Pilot Selection; Psychomotor Ability; Training Performance

This report provides descriptions of 254 references pertaining to aircrew selection through 1989. The references are classified into one of five categories: (1) paper-and-pencil general ability measures; (2) personality and interest measures; (3) psychomotor, perceptual and information processing measures; (4) job sample (i.e., light aircraft) measures; and (5) physiological measures. Many of these references were also subjected to meta-analytical procedures and these results are reported in Hunter and Burke (1992).

Hunter, D. R., & Burke, E. F. (June, 1992). *Meta analysis of aircraft pilot selection measures* (ARI Research Note 92-51). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Article type: Review
Keywords: Anxiety; Civilian; Cognitive Ability; Depression; Job Performance; Job Sample; Military; Personality; Pilot Selection; Psychomotor Ability; Training Performance

These authors applied meta-analytic procedures to a database of over 450 correlations that are based on overlapping samples of over 400,000 cases. Data were obtained from studies published between 1920 and 1990. This study constituted a "bare bones" analysis in that the validities were

only corrected for sampling error. The researchers found that there had been a definite downward trend in validities from the 1940s to the time of this study. This occurred both for aggregated predictors (i.e., batteries of tests) and when the data were disaggregated using only general ability with either dichotomous or continuous criterion variables. The authors propose that this may be due in part to several factors: the increasing homogeneity of aircrew personnel, the increased ability level that has resulted from increased selectivity, and the reduction in the number of "wash-outs" from training. Tables are presented that contain averaged validity coefficients classified by predictor type, criterion type, military service (i.e., Army vs. Air Force), nation and type of aircraft (fixed wing vs. rotary wing). For the Air Force (across nations) subsample, job sample measures were the best predictors (mean $r = .32$), followed by psychomotor coordination (mean $r = .31$) and then biographical inventories (mean $r = .29$). The average validity for personality tests was .14. All of these predictor measures reduced the amount of unexplained variance (although to a lesser extent for the biographical inventories and the personality inventories than the other predictors). The authors concludes that the relative equivalence of validities across nations, military services and aircraft types is evidence for the generalizability of the validity of aircrew selection measures.

Hunter, D. R., & Thompson, N. A. (1978). *Pilot selection system development* (AFHRL-TR-78-33). Brooks AFB, TX: Air Force Human Resources Laboratory.

Article type: Validation
Keywords: AFOQT; Cognitive ability; Military; Personality; Pilot selection; Psychomotor ability; SVIB; Training performance

This report describes an examination of the validity of measures included in the Ground-Based Screening (GBS) project using a sample of 1800 officers and officer trainees (although due to missing data actual sample sizes for individual measures were much smaller). Data was collected on the following measures: the Aircrew Psychomotor Test (APT), paper-and-pencil tests included in the Air Force Officer Qualifying Test (AFOQT), three tests with items similar in content to the Strong Vocational Interest Blank (SVIB; Scales A, B and C; no further description of these measures are provided), the Officer Background and Attitude Survey (OBAS) and the Automated Pilot Aptitude Measurement System (APAMS; designed to predict pilot training success). Two criterion measures were used, dealing with undergraduate pilot training (UPT): dichotomous pass/fail (for any reason) and dichotomous pass/fail (for Flight Training Deficiency). The researchers used multiple and zero-order correlation coefficients to examine the many predictor-criterion relationships. Generally, the APT and the APAMS were valid predictors of the dichotomous pass/fail criteria. The SVIB scales A and B also contributed to the prediction of UPT pass/fail.

Jessup, G., & Jessup, H. (1971). Validity of the Eysenck Personality Inventory in pilot selection. *Occupational Psychology, 45,* 111-123.

Article type: Group Comparison Study
Keywords: EPI; Military; Personality; Pilot Selection; Training Performance

These researchers administered the Eysenck Personality Inventory (EPI) to 167 pilot cadets in the British Royal Air Force (RAF). Their results showed that as expected, those who passed training were significantly less neurotic than those who failed, but these two groups of cadets did not differ in terms of extroversion (contrary to expectations). Researchers divided the sample at the population mean for each of the two scales (resulting in very uneven quadrants), and based on this dichotomy they showed that neuroticism is much more strongly related to success for introverted pilots than for extroverted pilots. They also found that the EPI scores did not correlate significantly with any of the other predictors (e.g., cognitive ability and psychomotor tests) of success in pilot training. Even the most experienced pilots have been shown to experience a considerable increase in arousal during take-off and before landing (Roman et al., 1966). Eysenck (1955, 1957) has predicted that introverts learn more quickly in any situation because inhibition at a cortical level is set up more quickly in extroverts. The present authors provide an elaborate explanation for why the quadrant differences occur. They argue that even though introverts do better in training, extroverts may do better on the job (although this is not exactly true!), particularly in managerial situations. They mention that initiative may be a second trait that could be negatively related to training success, but positively related to success on the job.

Johnston, N. (1985). Occupational stress and the professional pilot: The role of the pilot advisory group (PAG). *Aviation, Space and Environmental Medicine, 56,* 633-637.

Article type: Anecdotal/Descriptive
Keywords: Anxiety; Attitudes; Accidents; Civilian; CRM; Depression; Training Program Design

This article delineates the role that pilot advisory groups (PAG) could play in assisting pilots who manifest personal and professional problems, usually as a result of stressors. A PAG is essentially a pilots' self-help group, made up of pilots who are willing to assist their peers in times of need. PAGs are more common in the military than in the commercial airlines. Members of these PAGs are rarely given any sort of comprehensive training. Their main role is to assist pilots in recognizing problems and seeking help through appropriate specialists. One area in which PAGs seem particularly relevant involves helping pilots cope with psychological/behavioral problems (e.g., fear of flying, testitis, depression, skill decay, etc.). This article presents possible new opportunities for PAGs to be utilized in the context of helping pilots to deal with the aftermath of accidents or incidents in which they might have been involved. The author notes that helping pilots deal with these situations is very common in the military, but not as common for civilian airlines. They propose that a PAG might be very useful in this role. Generally, the conclusion is that PAGs will prove to be a useful tool in helping pilots deal with stressors of all types.

Kanki, B. G., Palmer, M. T., & Veinott, E. (1991). Communication variations related to leader personality. *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 253-259). Columbus: Ohio State University.

Article type: Group Comparison Study
Keywords: Civilian; Classification; Communication; Leadership; Motivation; Personality; Pilot Selection; Simulation

This paper attempts to establish a link between crew communication patterns and the personality profile of the crew's Captains. Captains took the Expanded Position Analysis Questionnaire and the Work and Family Orientation Questionnaire and their scores were used to classify them according to three major personality types. The IE+ Captains were highly motivated, goal-oriented achievers who were also very concerned about the interpersonal aspects of the crew. The I- Captain had a very high goal orientation, with little regard for interpersonal skills. Finally, the Ec- Captains had low achievement motivation and little regard for interpersonal relations with the crew. These researchers attempted to identify communication patterns that would help explain the differences in how the three types of Captains interacted with their crews. Transcripts from one leg of a five-leg full-mission simulation were content analyzed. The results indicated that Ec- Captains initiated less speech than did either the IE+ or the I- Captains. In addition, those crews that were led by Captains who initiated the least amount of speech typically performed the most poorly. A second set of analyses were conducted to examine the responses of crewmembers to Captains for each of the three personality types. First officers (FOs) initiated more commands to I- Captains than FOs paired with either IE+ or Ec- Captains. (This may explain results from an earlier study [Chidester & Foushee, 1989]). In addition, the FOs paired with the Ec- Captains also asked more questions than FOs paired with either IE+ or I- Captains.

Karlins, M., Koh, F., & McCully, L. (1989). The spousal factor in pilot stress. *Aviation, Space and Environmental Medicine, 60,* 1112-1115.

Article type: Review
Keywords: Accidents; Anxiety; Attitudes; Civilian; CRM; Depression; Stress; Training Program Design

The authors state that there is a growing recognition that many aviation accidents are caused by pilot error and that one source of this pilot error may be aviator stress. They point out that pilots are oftentimes under an enormous amount of stress from a variety of sources (e.g., traffic delays, job insecurity, passenger/equipment responsibility, etc.). They also note that spouses play a critical role in alleviating (or exacerbating) pilot stress. This article describes a program at a major airline that has attempted to deal with spousal influence on pilot stress as part of their Aircrew Resource Management (ARM) seminars. The purpose of this program was to provide spouses with an appreciation of the stress faced by pilots and to provide the couples the opportunity to examine their husband-wife relationship in the context of an aircrew marriage. The seminar was evaluated by asking for participants' reactions. Many participants believed that

their spouses would "understand them better" and would be more "sensitive and sympathetic" to the stressors they faced as a result of attending the seminar.

Koonce, J. M. (1982). Validation of a proposed pilot-trainee selection system. *Aviation, Space and Environmental Medicine*, *53*(12), 1166-1169.

Article type: Validation
Keywords: AFOQT; Military; Pilot Selection; Psychomotor Ability

In this article, the authors propose adding the aircrew psychomotor test (APT) to the Air Force Officer Qualifying Test (AFOQT; Form N) in an effort to reduce the number of cadets who fail to complete flight training. The APT consists of two psychomotor tasks. For the first task, subjects are required to track a target using two hand controls. The second task introduces the use of foot pedals. The subjects are required to use the hand controls and the foot pedals to keep a cross centered on the crosshairs that split the screen in half. Approximately 200 cadets from the Air Force Academy's classes of 1978 and 1979 were administered the AFOQT and the APT. They were ranked according to their scores and all who failed to score above the cut-off score were identified. All subjects were allowed to participate in flight training. After completion of flight training, their training performance was compared with their scores on the AFOQT and the APT. Using the APT as the screening device, 9.3 percent of the pilot-qualified cadets in the class of 1978 would have been denied access to pilot training, but 83 percent of these cadets would have succeeded; 6.0 percent of the class of 1979 would have been denied training, but 91.6 percent of these cadets would have succeeded. The correlations between the APT scores and UPT pass/fail were -.04 (class of '78) and -.09 (class of '79). Results of the analyses using the AFOQT were approximately the same as those obtained using the APT. Thus, neither the AFOQT or the APT discriminated very well between those who passed flight training and those who did not.

Kragh, U. (1960). The defense mechanism test: A new method for diagnosis and personnel selection. *Journal of Applied Psychology*, *44*, 303-309.

Article type: Validation
Keywords: Military; Pilot Selection; Psychiatric; Training Performance

This article examines the validity of the Defence [SIC] Mechanism Test (DMT). The general logic of this test is that pilots who consistently make use of defence mechanisms are likely to have difficulties dealing with the stress of the flight training environment. The DMT presents subjects with repeated subliminal exposure to a Thematic Apperception Test (TAT)-like picture. After each exposure, the subject makes a drawing of what he or she has seen and writes a short description. The drawings and comments are then evaluated by psychologists and their defensive organization is coded. Interrater reliabilities ranged from .59 to .90 across several different Swedish aviation classes. Point-biserial validities with a pass/fail criterion in training ranged from -.07 to .40. Validities varied across both individual raters and across classes for whom

ratings were made (these ratings were within psychologist). The authors speculated that the differences in validities across raters may be due to experience or skill differences between the different raters. Thus, additional rater training may increase the validity of the DMT.

Kreienkamp, R. A., & Luessenheide, H. D. (1985, October). Similarity of personalities of flight instructors and student-pilots: Effect on flight training time. *Psychological Reports*, *57*(2), 465-466.

Article type: Group Comparison Study
Keywords: Civilian; Personality; Training; Training Performance

The authors note that a wide range of factors influence the number of flight training hours necessary for obtaining a pilot's certificate. Sometimes a student reaches a learning plateau, and it is necessary for that student to change instructors before he or she is able to complete their training. The purpose of this study was to explore whether a significant relationship exists between differences in personality characteristics of instructors and those of their students and the number of flying hours required to complete flight training. The Myers Briggs Type Indicator (MBTI) was administered to 22 male and 10 female students and their instructors after completing flight training. Correlations were then computed between the difference scores of students and instructors on each of the MBTI scales and the number of flying hours required to complete training. No significant correlations were found for the Sensing/Intuition, Thinking/Feeling and the Judgment/Perception scales for either men or women. The correlation between number of flight hours and the difference score on the Extrovert/Introvert scale was .37 ($p < .05$) for men. A moderate correlation was found for women between flying hours and the difference score on the Sensing/Intuition scale, but this correlation was non-significant (mostly likely due to a lack of power).

Lambirth, T. T., Dolgin, D. L., & Carretta, T. R. (1989). Selected personality characteristics of student naval aviators and student naval flight officers. *Proceedings of the EURO-NATO Aircrew Selection Working Group* (12 Meeting). England: London.

Article type: Group Comparison Study
Keywords: Military; Personality; Training; Training Performance

This study examined personality differences between student naval aviators, student naval flight officers and a normative college sample using the Tri-Dimensional Personality Questionnaire (TPQ). The TPQ is a 100 item true-false questionnaire designed to assess several adaptive stimulus-response characteristics of personality proposed by an integrated biosocial theory of personality. It measures three dimensions: Novelty Seeking (activation), Harm Avoidance (inhibition) and Reward Dependence (maintenance) [these dimensions correspond to the dopamine, serotonin and norepinephrine neural systems, respectively]. The authors hypothesized that personality characteristics, psychomotor skills and cognitive abilities play a significant role in the selection of aviation as a career goal. Thus, there should be both similarities and

differences in personality characteristics between pilots and flight officers. One hundred twenty-nine aviation candidates from the U.S. Navy flight training program were used as subjects. Seventy-nine were student naval aviators and 50 were student naval flight officers. Three hundred twenty-six college students comprised the normative group. A one-way analysis of variance indicated that there were significant differences between the groups in Novelty Seeking and Harm Avoidance. Specifically, pilot candidates scored higher in Novelty Seeking than the normative group, and they scored lower than flight officer candidates and the normative sample in Harm Avoidance. The authors conclude that student naval aviators and student naval flight officers are likely to pursue novel and unfamiliar experiences and to appear calm and uninhibited under normal circumstances. They characterize the results of their study as being a "snapshot" taken at that point of time between finishing basic officer training and the beginning of flight training.

Lambirth, T. T., Gibb, G. D., & Alcorn, J. D. (1986). Use of a behavior-based personality instrument in aviation selection. *Educational and Psychological Measurement, 46*, 973-978.

Article type: Validation
Keywords: Military; Personality; Pilot Selection

This study examined the reliability and construct validity of the Dot Estimation Task (DOT), which is a behaviorally-based personality instrument designed to measure compulsivity versus decisiveness. One hundred fifty-three psychology students participated in this study. On the DOT, subjects are simultaneously presented with two fields containing an arbitrary number of dots. One field contains one more dot than the other. Subjects are required to determine, as quickly as possible, which of the two boxes has the larger number of dots. The number of trials attempted and the mean reaction time for correct and incorrect responses are recorded. Subjects also took the Obsessive-Compulsive Scale (OCS) and the Obsessive-Compulsive Scale Revised (OCSR). The OCS is a 22-item true-false scale reflecting cognitions and behaviors commonly associated with an obsessive compulsive disorder. The OCSR is a 20-item scale; each item has four response options that reflect an increase in behaviors, thoughts, or feelings associated with obsessive-compulsiveness. Testing was conducted in two sessions. In the initial session, subjects completed the DOT and either the OCS or the OCSR and in the final session they completed the DOT and the alternative obsessive-compulsive measure. Test-retest reliabilities for the DOT were .64 for the number of trials attempted, .46 for the number of correct trials and .64 for the number of incorrect trials. All correlations between the DOT and the OCS and the OCSR were non-significant. Thus, the DOT displayed acceptable test-retest reliability, but no construct validity.

Lardent, C. L. (1991). Pilots who crash: Personality constructs underlying accident prone behavior of fighter pilots. *Multivariate Experimental Clinical Research, 10*(1), 1-25.

Article type: Validation
Keywords: Accidents; Attitudes; Anxiety; Combat/Fighter Pilots; Job Performance; Military; Personality; Pilot Selection; Stress

This article examines whether measures of certain personality constructs can predict pilot involvement in accidents. The author makes the assumption that there is not a unitary set of noncognitive measures that make one more likely to be involved in accidents. Rather, there may be some personal characteristics that, depending on the circumstances, predispose one person more than another to engage in behaviors that increase the probability that he or she will be involved in an accident. Specifically, the author attempts to answer the question of whether or not there are personality, motivational and/or other psychological variables associated with behavioral responses that, under various circumstances, are more or less likely to result in aircrew accidents. An extreme groups design was utilized including 89 F-4 fighter aircraft pilots who were involved in at least one class A mishap and 89 F-4 pilots who were considered to be accident-free (control group). Form A of the 16 PF was used to assess personality. This instrument measures 16 source or primary factors and eight secondary factors (in addition to including several validity scales). The impact of life changes on temporary accident proneness was also measured using the Holmes-Rahe Social Readjustment Rating Questionnaire (SRRQ). Accident involvement was a dichotomous variable, either "crash" or "safe." Relationships between the variables were examined using set correlations. None of the eight 16 PF secondary factors or the SRRQ scales yielded significant relationships with the crash criterion. For the 16 PF primary factors, pilots who crashed scored significantly higher than those who did not crash on conscientious and self-sufficient. They also scored significantly lower on trusting, naive and relaxed/tranquil. The multiple correlation between the predictor set (i.e., the significant 16 PF scales) and "crashing" was .52. The results were not cross-validated, but the estimated cross-validation coefficient was .23. The author interprets the results as supporting the need to conduct "limited domain" research because past research has convincingly debunked the notion of a "grand theory" of accident proneness.

Lester, L. F., & Bombaci, D. H. (1984). The relationship between personality and irrational judgment in civil pilots. *Human Factors, 26*(5), 565-572.

Article type: Validation
Keywords: 16 PF; Attitudes; Accidents; Civilian; Personality; Pilot Selection; Situational Judgment

The author discusses the increasing dependability and reliability of aircrafts and how this has made the role of the human operator much more critical. One effort to identify individual attributes that are important to aviation safety has been the delineation of five hazardous thought patterns (HTPs): anti-authority, impulsivity, invulnerability, macho and resignation. These HTPs are thought to have the status of constructs and are seen as mediating the link between basic

psychological processes and irrational pilot judgment. Even though there is basically no validity research on these HTPs, they are a core element of the judgment training currently being conducted by several U.S. and Canadian airlines. The purpose of this study was to examine the construct validity of these HTPs using the 16 PF and Rotter's Locus of Control scale. Thirty-five male pilots participated (4 held commercial ratings and 31 had private licenses). Subjects completed three scales from the 16 PF (impulsivity, superego strength and integration/self-concept control) and the Rotter Locus of Control scale. They also completed a self-assessment inventory designed to measure the five HTPs. When completing this inventory, subjects are presented with 10 flight scenarios that describe errors in pilot judgment and asked to rank (from most probable to least probable) the reasons why they might have made the error. Results showed that fourteen of the subjects did not have a predominant HTP, but the percentage of pilots showing a single dominant HTP was similar to that found in later studies (e.g., Lester & Connolly, 1987). Of the remaining subjects, 43 percent primarily exhibited the invulnerability HTP, impulsivity was dominant in 20 percent and macho was dominant in 14 percent. The remaining HTPs were predominant in only a few subjects. An analysis of variance (ANOVA) revealed a significant relationship between the three HTPs and the integration/self-concept control scale and Rotter's locus of control scale. Pilots with the invulnerable HTP scored significantly lower on the integration scale and pilots with the macho HTP scored significantly higher on the integration scale. Also, pilots with the dominant macho HTP were significantly more internally controlled than either the invulnerable or impulsive pilots. The authors conclude that this study provides some support for the construct validity of the HTPs, but notes that their sample was extremely small.

Lester, L. F., & Connolly, T. J. (1987). *The measurement of hazardous thought patterns and their relationship to pilot personality.* Paper presented at the 4th International Symposium on Aviation Psychology.

Article type: Test Development and/or Evaluation
Keywords: Accidents; Attitudes; Civilian; Personality

The authors state that errors in pilot decision-making (PDM) can be accounted for by five hazardous thought patterns (HTPs): anti-authority, impulsivity, invulnerability, macho and resignation. However, the evidence for these HTPs is largely anecdotal and their exact nature is not very clear. For example, some researchers consider them to be attitudes, while others consider them to be personality variables. This study was conducted with five purposes: (1) establish base rates for the different HTPs; (2) examine the relationships among the HTPs; (3) determine how the HTPs are related to pilot personality; (4) determine whether the HTPs are related to involvement in aviation accidents/incidents; and (5) examine the utility of two forms of the Pilot Decision Making Questionnaire (PDMQ). One hundred fifty-two males in their late teens or early twenties (all pilots) were used as subjects in this study. They had an average of 190 flight hours. Subjects completed the Rotter Locus of Control Scale, four scales from the 16 PF (Emotional Stability, Surgency, Conscientiousness and Integration) and two forms of the PDMQ. In the PDMQ Form A, subjects are presented with scenarios and asked to indicate whether the pilot's behavior was "very much like me" or "not at all like me." This is designed to be a measure

a subject's overall propensity toward irrational judgment. The PDMQ Form P describes irrational decisions and presents the subjects with reasons (each representing a HTP) for why that pilot behaved as he or she did. Subjects are asked to rate how likely it was that each of the reasons would have caused the pilot to behave in that manner. This was designed to measure the strength of each of the HTPs. These data were analyzed and subjects were classified according to their predominant HTP (39%-Invulnerable, 24%-Impulsive, 19%-Macho, very few-Resignation, 0% Anti-Authority). HTP intercorrelations ranged from .21 to .53 and all were positive. Subjects who displayed the Macho HTP were significantly more internally controlled than subjects who displayed the Invulnerable or Impulsive HTPs. They were also more Conscientious than those classified as Invulnerable. No other significant differences among the groups were found. Subjects were separated into groups with "better" and "poorer" judgment on the basis of their scores on the PDMQ Form A. Pilots with better judgment were more internally controlled and were better integrated than those with poorer judgment. No relationships were found between scores on the PDMQ Form A and involvement in aviation accidents/incidents.

Levine, J. B., Lee, J. O., Ryman, D. H., & Rahe, R. H. (1976, January). Attitudes and accidents aboard an aircraft carrier. *Aviation, Space and Environmental Medicine, 47*, 82-85.

Article type: Validation
Keywords: Accidents; Attitudes; Personality; Pilot Selection

This article examines the relationships between behavioral attitudes and accidents for aviators and enlisted air wing support personnel aboard an aircraft carrier. One hundred fifty-six aviators and 879 support personnel participated in this study. These subjects completed a 22-item attitude questionnaire which asked them to rate the extent to which they agreed or disagreed with each statement. The questionnaire was factor analyzed (varimax rotation) and items loading most highly on each factor were summed to form six scales: logic, adventurousness, discipline, concern with self and brashness. The accident criterion was the number of personal injuries (or deaths) which were sufficiently severe to require a visit to sick call. Results showed that the Adventurousness scale was significantly correlated with personal injuries among the enlisted air crew personnel ($r = .12, p < .01$) and with aircraft accidents among aviators ($r = .25, p < .01$). None of the other five factor-analytically derived scales were correlated with accidents for either sample.

Levine, A. S., & Tupes, E. C. (1952). Postwar research in pilot selection and classification. *Journal of Applied Psychology, 36*, 157-160.

Article type: Review
Keywords: Background; Biodata; Classification; Cognitive Ability; Military; Motivation; Pilot Selection; Situational Judgment; Training Performance

This article summarizes major findings of postwar research in pilot selection and classification in the United States Air Force (USAF). The authors discuss how the problem of selection and

classification has changed over time. For example, during World War II, most eliminations were for flying deficiency, but since then, a large number of eliminations have been for other reasons, with motivation being the most prominent. The validity of the Pilot Stanine (this is primarily a measure of flying aptitude) for predicting flying deficiency has remained relatively stable over time, at about .60. In contrast, the validity of the Pilot Stanine for predicting elimination due to motivational reasons has ranged across different years from about .34 to .56. The authors discuss several reasons why this may have occurred. One reason could be that the Aircrew Classification Battery (optimally keyed for predicting flying deficiency), which prior to World War II had only been used for classification, has since then been used to select cadets into flight training. This battery consists of two tests dealing with both general and more specific aviation information questions: the Biographical Data Blank and the Practical Judgment test. This battery correlated .62 with flying deficiency, .51 with administrative elimination (primarily physical elimination) and .61 with elimination for motivational reasons. The authors recommend including some other tests (namely, the Biographical Inventory) and revising current tests in order to improve their predictive validity. However, they also reported that both the Pilot Stanine and the Aircrew Classification Battery have had much success in predicting a variety of organizationally relevant criteria.

North, R. A., & Griffin, G. R. (1977). *Aviator selection 1919-1977* (Special Report 77-2). Pensacola, FL: Naval Aerospace Medical Research Laboratory, Naval Air Station.

Article type: Review
Keywords: Anxiety; Background; Biodata; Civilian; Cognitive Ability; Job Performance; Military; Personality; Pilot Selection; Psychomotor Ability; Stress; Training Performance

This report provides wide-ranging descriptions of tri-service aviator selection methods and their predictive validity. A secondary purpose of the report was to suggest ways to improve the prediction of aviator success. The authors note at the time this report was written, that the potential to increase the predictive validity of current selection procedures was high, because only about 40 percent of the variance associated with aviator success (mostly in training) was accounted for by currently available methods. They also note that there have not really been any breakthroughs in perceptual/cognitive paper-and-pencil testing since World War II. The report provides a thorough summary of the methods that have historically been used to select pilots. It also describes very thoroughly, the success of various perceptual, cognitive, aptitude, psychomotor, biographical, personality and neurological approaches to aviator selection. The authors make the point that at least 40 personality inventories were evaluated for pilot selection between 1950 and 1976 and none of these were able to improve pilot selection procedures. In addition, the criterion problem in aviator selection is also thoroughly reviewed. The authors emphasize that a pass or fail in Undergraduate Pilot Training (UPT) is not always a satisfactory criterion and recommend that researchers focus their efforts on developing appropriate operational performance criteria. The authors state that the objective of present day aviator selection research has evolved and should have two main purposes: (1) to reduce attrition in UPT; and (2) to effectively predict pilot performance from 4 to 8 years into an officer's career.

They also recommend studying the predictive validity of non-paper-and-pencil measures. Specifically, those non paper-and-pencil measures that appear most worthy of investigation are selective/divided attention capabilities, stress and anxiety, measurement of motivation and perceptual-psychomotor skill assessment.

O'Connor, P. J. (1975). Testitis (excessive anxiety about flying checks). *Aviation, Space and Environmental Medicine, 46*, 1407-1409.

Article type: Descriptive
Keywords: Anxiety; Depression; Job Performance; Personality; Psychiatric

This article describes the "testitis" syndrome (excessive anxiety about flying proficiency checks). The health of civilian aircrew members is checked annually until age 40, at which time checks occur every six months. Aviation competency is checked two ways: (1) a 6-month competency check is usually carried out in the simulator; and (2) an annual route check occurs with a training Captain accompanying the aircrew on a commercial flight. Persistent failure by a pilot to pass either of these examinations will lead to the loss of his or her commercial license. Thus, this is often times a very stressful event for aircrew members. Most pilots are able to contain these "examination nerves," but for some the tension escalates to the point where it causes the pilot to fail his or her checkride. Common symptoms of testitis include a phobic anxiety state, a generalized anxiety state, depressive symptoms and/or hypochondriasis. The author reports that the prognosis for recovery of returning aircrew who develop this disorder is difficult to assess because often times the pilot resigns before a diagnosis can ever be given. For those who are diagnosed with this problem, the recovery rate was reported to be quite variable.

Pedersen, L. A., Allan, K. E., Laue, F. J., Johnson, J. R., & Siem, F. M. (1992). *Personality theory for aircrew selection and classification* (AL-TR-1992-0021). Brooks AFB, TX: Armstrong Laboratory, Human Systems Division.

Article type: Review
Keywords: Classification; Military; Personality; Pilot Selection; Psychological Status

This report reviews the personality literature for its utility in guiding research on aircrew selection and classification. The authors hoped that this review would help to guide future efforts to provide a personality component to supplement the physical, academic and aptitude requirements currently employed in aircrew selection. It was felt that this was needed because of the high training costs associated with attrition from Undergraduate Pilot Training (UPT), and to aid in the impending transition from single-track to multi-track Specialized UPT. Up to this point, the Basic Attributes Test (BAT) had included five different personality measures, none of which were valid as predictors of training performance. Nine personality theories were reviewed for this report: (1) 3-factor theory (Eysenck & Eysenck, 1985); (2) 5-factor theory (McCrae & Costa, 1985); (3) 16-factor theory (Cattell, 1972); (4) Circumplex theory (Wiggins & Broughton, 1985); (5) Socioanalytic theory (Hogan, 1983); (6) Psychological types theory (Jung, 1971;

Myers, 1980); (7) Temperament theory (Buss, 1988; Buss & Plomin, 1975); (8) Instrumentality-Expressivity theory (Bakan, 1966; Spence, 1983); and (9) Needs theory (Murray, 1938). The authors decided that there should be two main types of criteria to use in evaluating the theories: "general" criteria that address customary scientific issues and "operational" criteria that address practical application issues. The general conclusion reached was that only the 5-factor taxonomy of personality traits met all the criteria for being an effective framework from which to develop personality measures that could be used for selection and classification of pilots. The authors propose that development of an operational measure of personality traits should follow a four-step procedure: (1) identification of aircrew performance criteria; (2) identification of the most useful traits appropriate to the job performance criteria; (3) development and refinement of scales to measure those traits; and (4) assessment of scale reliability and validity.

Petrullo, L. (1951). Research on criteria of motivation and temperament conducted by the human resources research laboratory. In A. W. Melton (Chair), *Research planning conference on objective measurement of motivation and temperament* (Conference Report 51-3). San Antonio, TX: Human Resources Research Center, Air Training Command, Lackland Air Force Base.

Article type: Test Development and/or Evaluation
Keywords: Attitudes; Military; Motivation; Performance Ratings; Pilot Selection; Satisfaction

This paper describes research in which non-technical proficiency criteria were developed, which could then be used to validate motivation and temperament predictors. These measures primarily consisted of ratings collected from peers, supervisors and the ratee themselves on a variety of criteria. Scales were also developed to measure attitudes toward the job (in this case aircraft mechanic), job satisfaction, attitude toward supervisor, group interaction, cooperativeness, etc. In addition, several overall performance ratings were also developed, both for peers and supervisors.

Picano, J. J. (1991). Personality types among experienced military pilots. *Aviation, Space and Environmental Medicine, 62*, 517-520.

Article type: Group Comparison Study
Keywords: Classification; Military; Personality

The author notes that pilots have consistently been found to be more achievement-oriented, outgoing, active, competitive, and dominant; and less introspective, emotionally sensitive and self effacing than the general population. The purpose of this study was to identify and describe distinct personality subtypes within an experienced sample of military pilots. The author also examined the relationship between personality and type of aircraft assignments and other aviation/military variables in a sample of 170 U.S. Army helicopter pilots. Results for the pilots were also compared with a normative group of managers. The Occupational Personality

Questionnaire (OPQ) was administered to both groups. It is a self-report measure of personality, consisting of 372 items that form 31 scales related to three broad dimensions of personality: relationship, thinking and feeling. The pilot group tended to be more competitive, practical, introverted and oriented to concise, critical thinking than the managers. They were also less empathic, emotionally attuned, poised and socially confident. A cluster analysis was conducted to identify distinct personality subtypes. Three clusters were identified. Cluster 1 consisted of pilots who scored highest on Independent, Competitive and Decisive and they scored lowest on Modest, Democratic, Caring, Artistic and Social Desirability. Pilots in Cluster 2 scored highest on Modest and Worrying and lowest on Persuasive, Controlling, Outgoing, Socially Confident, Change-Oriented, Relaxed, Optimistic and Achieving. Cluster 3 pilots scored highest on Affiliative, Caring, Data Rational, Change-Oriented, Forward Planning, Detail Conscious, Relaxed and Optimistic and lowest on Worrying. A comparison of the three clusters showed that there were no differences in the pilots in terms of type of mission flown, years of aviation/military service, or number of flight hours. However, there were a disproportionately large number of instructor pilots among the Cluster 1 pilots.

Reinhardt, R. F. (1970). The outstanding jet pilot. *American Journal of Psychiatry, 127*, 732-736.

Article type: Group Comparison Study
Keywords: Background; EPPS; Military; MMPI; Personality; Psychiatric

This reports discusses descriptive data for 105 naval aviators selected by their commanding officers (in 1967 and 1968) as representing the upper tenth of aviators under their command. In addition, descriptive data for 70 aviator failures were also examined. All subjects completed several personality inventories and were interviewed by psychiatrists. Results from the interviews showed that two-thirds of the aviators' fathers had a military background with 85 percent of them serving in the Navy. Further, only nine percent of the outstanding aviators reported "father deprivation," as compared to 20 percent of the failure group. On the Maudsley Personality Inventory, the successful pilots' extroversion-introversion scores were identical with college norms, but their neuroticism-stability scores were significantly lower than the college norm. On the Edwards Personal Preference Schedule, the successful pilots described themselves as desiring success and not desiring introspection or giving and getting help. The failed pilots score significantly lower on the desire for success and the need for autonomy scales. On the Minnesota Multiphasic Personality Inventory, the successful pilots were described as energetic and having favorable views of themselves.

Retzlaff, P. D. D., & Gibertini, M. (1987). Air Force pilot personality: Hard data on the "right stuff." *Multivariate Behavioral Research, 22*, 383-399.

Article type: Group Comparison Study
Keywords: Classification; Military; Personality; PRF; Training

The authors address the question of whether successful pilots possess personality traits that distinguish them from both nonpilots and unsuccessful pilots. They framed this question from two different perspectives. The first was to ask whether a specific personality structure exists that represents successful pilots. The second was to ask whether specific personality traits can predict which pilots succeed on various criteria. This study examined the first question using a sample entering Undergraduate Pilot Training (UPT). The sample consisted of a total of 350 white males, with 249 in the development sample and 101 in the validation sample. Two different predictor measures were administered. The Personality Research Form (PRF; Form E) measures normal personality characteristics and has 352 true-false items organized into 21 scales. The Millon Clinical Multiaxial Inventory (MCMI) is a 20-scale instrument designed to measure DSM-III relevant dimensions. The MCMI assesses eight basic personality patterns, three pathological personality patterns and nine clinical syndromes. After the subjects completed the inventories, cluster analysis was used to cluster the development sample into several groups based on their scores, these clusters were then applied to the cross-validation sample. Three relatively distinct clusters were identified. Cluster 1 was high on the following scales: Histrionic, Narcissistic and Antisocial Compulsive, Affiliation, Aggression, Exhibition, Impulsivity and Play. They were low on Cognitive Structure and Order. Cluster 2 was moderate on the Histrionic and Narcissistic scales and low on the Defendence scale. This group scored high on the following scales: Compulsive Achievement, Affiliation, Endurance and Social Desirability. Finally, Cluster 3 was high on the Compulsive scale, but low on the Histrionic, Affiliation, Change, Dominance and Exhibition scales. The results replicated well in the cross-validation sample. Next, discriminant function analysis was used to classify pilots into one of the clusters. Results showed that they were able to classify 92 percent of the sample into their correct cluster. The first cluster comprised 21 percent of the sample and were those subjects most closely resembling stereotypical pilots who have the "right stuff." Cluster 2 contained 58 percent of the sample and the researchers speculated that these pilots might be best suited for tanker, transport and bomber aircraft. Cluster 3 comprised 21 percent of the sample and in terms of stereotypes had the "wrong stuff." The authors recommend longitudinal studies to determine whether different types of pilots are more successful at different points in their military career (and perhaps also suited to a specific type of aircraft).

Retzlaff, P. D., & Gibertini, M. (1988). Objective psychological testing of U.S. Air Force officers in pilot training. *Aviation, Space and Environmental Medicine, 59*, 661-663.

Article type: Group Comparison Study
Keywords: Military; MMPI; Personality; PRF; Psychiatric; Training

The authors reviewed the literature pertaining to the use of psychological testing with Air Force pilots and conclude that many of the instruments that have been used are not psychometrically sound and were used in an inappropriate manner. For example, they point out that the often used Minnesota Multiphasic Personality Inventory (MMPI) was developed over 40 years ago and normed on psychiatric patients. They advocate the use of three tests that they believe adequately measure a broad range of cognitive, personality and psychopathological domains. The first of these tests is the Multidimensional Aptitude Battery (MAB) which is a test of intellectual ability based largely on the WAIS-R. Verbal components include information comprehension, arithmetic, similarities and vocabulary. Performance measures include digit symbol coding, picture completion, spatial thinking, picture arrangement and object assembly. The Personality Research Form (PRF; Form E) measures normal personality characteristics and has 352 true-false items, organized into 21 scales. The following are some of the traits measured: achievement, aggression, autonomy, dominance, harm avoidance, impulsivity and social recognition. The Millon Clinical Multiaxial Inventory (MCMI) is a 20-scale instrument designed to measure dimensions associated with psychiatric diagnoses. The MCMI assesses eight basic personality patterns, three pathological personality patterns and nine clinical syndromes. The sample used in this study consisted of 350 white males entering Undergraduate Pilot Training (UPT). For the MAB, the subjects had an average full scale IQ of 120, with an average verbal IQ of 117 and an average performance score of 121. Most subjects scored above the MAB normative sample. For the PRF, the pilots scored higher than college students on affiliation, cognitive structure, dominance and social desirability. They scored lower than college students on abasement, autonomy, harm avoidance and understanding. The authors pointed out that this could be a result of differences in age, education, or other moderators between student pilots and college students. The MCMI pointed to histrionic and narcissistic patterns personality in the student pilots. The authors note that this is in line with the lay perception of the pilot as highly sociable and having strong self esteem.

Rimland, B., & Steinemann, J. H. (1960). The development of a fake-resistant test for NROTC selection. *Proceedings of the Tri-Service Conference on Selection Research* (pp. 100-110). Washington, DC: Office of Naval Research, Department of the Navy.

Article type: Test Development and/or Evaluation
Keywords: Military; Motivation; NROTC; Pilot Selection

This is a preliminary report discussing a project to develop a test (or tests) that might yield valid measures of both applicant motivation toward a naval career and potential for becoming an effective officer. In this study, the authors attempted to examine the prevalence and effectiveness of attempts to fake in a highly competitive situation. Subjects were 10,000 NROTC students who

took the tests under five experimentally induced response sets: Honest; Fake good; Fake, but be aware of a lie scale; Normal; and, Normal, but be aware of a lie scale. The authors tried several different approaches to develop a fake-resistant test. Items were presented in a forced choice format and were selected for operational use only if they were found to be resistant to faking. Form 3XA was one of the tests administered (based on Form 2 of the NROTC Regular Questionnaire) and was comprised of biographical information items, preference items and self-description items. Form 3XB was comprised of items from the Career Intensity Profile (CIP). The content was similar to that of Form 3XA. Form 3XC was modified version of a test used by the Army to predict ROTC leadership. It consisted of self-descriptive dyads. Subjects were also asked to indicate whether they "would fake" if they were taking the tests in an actual application setting. A self-rating attitude scale was also administered which consisted of six items from the Career Intensity Questionnaire (CIQ) and was a measure of career motivation (a marker variable). Peer ratings were also collected and each subject was asked to identify those four subjects in their class who they believed were the most highly career oriented and another four who would make the best officers. Fitness reports and actual retention were used as follow-up criteria. Generally, mean scores for those in the Fake condition were higher than means from any other condition and the means of those in the Fake, but be aware of the lie scale were next highest. The mean scores of the Honest, Normal and Normal but aware of a lie scale were all very similar, although in all cases these differences were not very large. There was also a tendency for those who indicated they would fake to have slightly lower criterion scores than the others, but the trend was rather weak. Further analyses were planned.

Sanders, M. G., & Hoffman, M. A. (1975). Personality aspects of involvement in pilot-error accidents. *Aviation, Space and Environmental Medicine, 46(2)*, 186-190.

Article type: Group Comparison Study
Keywords: 16 PF; Accidents; Military; Personality

The authors examined the utility of a decision-making task and various personality variables for discriminating between pilots who had been involved in an accident and those who had not. Subjects were 51 volunteer military aviators with ranks ranging from Chief Warrant Officer-2 to Lieutenant Colonel. Prior accident involvement was determined through an investigation of the United States Army Agency for Aviation Safety (USAAAVS) accident records. Each aviator listed as a causal factor in at least one aviation accident (either major, minor, or incident) was classified as pilot-error accident involved (PEAI), otherwise, they were classified as pilot-error accident free (PEAF). Several measures were administered to all of these aviators. The first was the 16 PF (Form A), a personality inventory, which consists of 16 primary factors and four secondary factors. The second inventory was the Mehrabian Achievement Scale which provides an indication of need for achievement or desire to attain success. A decision-making task was also administered which involved having the subject decide when to leave a light on or turn it off, based on a set of rules. Subjects' scores were the means of each of the various instruments taken over several different trials. The first analysis included the measures from the 16 PF and the N-Ach score from the Mehrabian. Three of the 21 scores discriminated between the PEAI and the PEAF aviators: (1) Group Dependent vs. Self Sufficient; (2) Practical vs. Imaginative; and

(3) Forthright vs. Shrewd. These three scores correctly classify 86 percent of the sample into either the PEAI or the PEAF groups. The second stepwise discriminant analysis included the scores from the decision-making task. Again, none of these scores successfully discriminated between the two groups of aviators. This study was cross-validated in a later paper (Sanders, Hoffmann, & Neese, 1976) and the results did not replicate.

Sanders, M. G., Hoffmann, M. A., & Neese, T. A. (1976). Cross-validation study of the personality aspects of involvement in pilot-error accidents. *Aviation, Space and Environmental Medicine, 47*(2), 177-179.

Article type: Group Comparison Study
Keywords: 16 PF; Accidents; Military; Personality

The authors conducted a cross-validation study in an attempt to replicate the results obtained by Sanders and Hoffman (1975). The sample consisted of 66 military aviators. Once again the 16 PF (Form A) was administered. Pilots were classified into one of two groups, either pilot-error accident involved (PEAI) or pilot-error accident free (PEAF) based on United States Army Agency for Aviation Safety (USAAAVS) accident records. The stepwise discriminant analysis did not discriminate between PEAI and PEAF aviators. A second stepwise discriminant analysis was performed using age, total military flight hours and years of flight status. None of these variables discriminated between PEAF and PEAI pilots. The authors conclude that individual differences in personality characteristics of aviators prevented the identification of personality traits associated with the PEAI and the PEAF groups.

Sells, S. B. (1955). Development of a personality test battery for psychiatric screening of flying personnel. *Aviation Medicine*, 35-45.

Article type: Test Development and/or Evaluation
Keywords: Accidents; Cognitive Ability; Military; Performance Ratings; Personality; Pilot Selection; Psychomotor Ability; Training Performance

The author characterizes the qualifications for jobs in combat aircrews as encompassing three areas: (1) physical qualifications; (2) abilities and aptitudes; and (3) personality factors. He notes that the USAF has developed thorough and effective standards for selection procedures for the first two areas, but not for the third area. This report summarized the progress to date on a large scale developmental project (started in 1949) to update the USAFs use of personality factors in selection. The present research included only pilots. Subjects were administered a wide range of paper and pencil, projective, performance and apparatus tests. Students entering flying school were not told that they could be eliminated based on the results of the test, but they were led to believe that the tests would be used for administrative decisions (i.e., selection). Validation of the tests were in progress (or planned) at the time of this article was written for three stages in the students careers: (1) training; (2) post-training and operational experience; and (3) combat performance (based on data collected in Korea). The results indicated that ratings made by

classmates and by instructors during training were superior to psychologists' evaluations in predicting combat performance. However, a significant positive correlation was found between performance ratings in combat and an absence of pathological behavior symptoms in the clinical reports made by a field survey team. Also, pilots with superior adjustment ratings in training and higher pilot stanines tended to have more accidents. At the time this article was written, all studies had been conducted using training level criteria, which relies on a pass/fail criterion. The author advocates using "purified" pass/fail measures. For example, one approach is to classify the population into more refined administrative categories according to their performance in training. Finally, the author stated that there are three broad, interrelated areas in which efforts should be concentrated: (1) specific motivational structure; (2) character integration; and (3) tolerance of frustration and anxiety.

Sells, S. B. (1956). Further developments on adaptability screening of flying personnel. *Aviation Medicine*, 440-451.

Article type: Test Development and/or Evaluation
Keywords: Accidents; Cognitive Ability; Military; Performance Ratings; Personality; Pilot Selection; Psychomotor Ability; Training Performance

This was the third progress report on the development of a personality battery to supplement present aircrew selection techniques. The goal of the program was to identify prospective aircrew trainees who were predisposed to difficulties in adapting to the rigors of military flying. Predictor measures included a wide variety of aptitude, personality, perceptual, attitude and psychomotor variables. Personnel information from Personnel File Form 66 and the 201 File concerning rate and extent of promotion, type of assignment, extent of command responsibility and flying duties involved were gathered to use as criteria. Psychologists made adjustment ratings, one of which was a purified pass/fail criterion. The high group included those who graduated from training and were well-adjusted. The low group consisted of pilots who failed training for reasons of poor motivation, excessive emotional reaction, or overt symptoms attributable to stress in the program. The following are the results obtained when correlating the different predictors with the various criteria. The Aviation Interest Key is comprised of 25 items covering the attitudes of parents toward the trainees participation in hazardous sports, a history of motion sickness, prior military flying experience, etc. It yielded correlations of .37 to .41 with the purified pass/fail criterion controlling for pilot stanine. Seven Minnesota Multiphasic Personality Inventory (MMPI) scales (Hs, Pd, Winne, Taylor, Seaquist, D and Hy) correlated from .10 to .40 (mean = .24) with the purified pass/fail criterion measure. The Pilot Opinionaire used an indirect polling approach to assess attitudes toward various aspects of military aviation (and also included an authoritarianism scale). Correlations were found between it and the pass/fail criterion that ranged from .28 for cadets to .11 for officers. The author also demonstrated that even though aptitude measures were much more highly correlated with the pass/fail criterion ($r = .53$), they correlated only .13 with information from Form 66 (which involves job performance). Whereas, the correlations of the personality variables remained relatively constant, albeit low ($r = .11$). At this point in time, only four screening tests have been validated against post-training criteria, the Personal Inventory (PI), the Cornell Index (CI), the Cornell Word Form (CWF) and the Sentence

Completion (SC) Factor scores. The correlations of these tests with various personnel file form (Form 66) information ranged from -.07 (for Interpersonal Attitudes) to .23 (for Self-Enhancement).

Sells, S. B., & Trites, D. K. (1957). Psychiatric screening of combat pilots: Correction of the record. *U.S. Armed Forces Medical Journal, VIII,* 1821-1824.

Article type: Validation
Keywords: Military; Pilot Selection; Psychiatric

The purpose of this article was to correct factual errors in an article by Sparks and Niess (1956). Specifically, these authors disputed Sparks and Niess' interpretations of ongoing work conducted by Sells and Trites (and others) concerning the adaptability screening of military flyers. They present three main concerns. First, Sparks and Niess concluded that the psychiatric screen was unable to effectively screen out failures or identify successes in combat flying. The authors dispute this conclusion and present evidence to the contrary. Second, Sparks and Niess concluded that it was essential that psychologists have a realistic understanding of combat, since training researchers tended to predict successes to fail and failures to succeed. The authors primary complaint with this statement is that it was based on a misinterpretation of a correlation between ratings of pilots in combat by their peers and officers and ratings by psychologists. For one set of correlations the psychologist had access to all rating information by the pilots' peers and supervisors and for the other they did not. Sparks and Niess seem to have missed this point. Finally, Sparks and Niess concluded that certain personality types were particularly suited to certain combat missions. The authors object to this because their results were based on relatively small numbers of subjects and they believed that Sparks and Niess' conclusions were too strong based on their data.

Shull, R. N., & Dolgin, D. L. (1989). Personality and flight training performance. *Proceedings of the Human Factors Society 33rd Annual Meeting, 2,* 891-895.

Article type: Validation
Keywords: Military; Personality; Pilot Selection; Training; Training Performance

In this paper, the relationships between a Risk Test, the Pilot Personality Questionnaire (PPQ), and a pass/fail training criterion and academic grades were examined. The authors also examined the relationships between the two predictor measures and the current U.S. Navy/Marine Corps aviation selection battery, the Academic Qualifications Test/Flight Aptitude Rating (AQT/FAR). Subjects were 407 student naval aviators (SNAs) and 182 student naval flight officers (SNFOs) who took the Risk Test and 110 SNAs and 114 SNFOs who took the PPQ. In the Risk Test, subjects are asked to choose which of ten squares is the reward square accumulating points as they go. The average number of responses (NR) and the corresponding average reaction time (RT) are measured for each trial. The PPQ is a self-administered personality inventory containing 112 multiple-choice items, which is a combination of 4 different established personality tests:

Locus of Control, Work and Family Orientation, Personality Attributes Questionnaire and a Social Desirability Scale. The AQT is a general cognitive ability measure and the FAR is comprised of the Mechanical Comprehension Test, the Spatial Apperception Test and the Biographical Inventory. Results showed that for the SNAs, the number of responses on the Risk Test correlated .13 ($p < .05$) with Undergraduate Pilot Training (UPT) pass/fail criterion and -.13 ($p < .05$) with academic grades. Reaction time correlated -.18 ($p < .05$) with UPT pass/fail. For the PPQ, the Social Desirability Scale correlated -.30 ($p < .05$) with UPT pass/fail and -.45 ($p < .05$) with academic grades. For the SNFOs, the number of responses on the gambling risk test correlated .28 ($p < .05$) with UPT pass/fail, while reaction time correlated -.45 ($p < .05$) with UPT pass/fail. Neither the number of responses nor the reaction time variables from the Risk Test were correlated with academic grades. For the PPQ, high self control correlated .40 ($p < .05$) with UPT pass/fail, but no other scales were correlated with this criterion. For the academic grades criterion, PPQ aggressiveness correlated .54 ($p < .05$), high competitiveness correlated .52 ($p < .05$) and submissiveness correlated -.55 ($p < .05$). The authors concluded that those trainees who exhibited more risk-taking behavior were more likely to complete flight training. However, the experienced pilots showed less risk-taking behavior than the SNAs. This may mean that although in training those that are successful are more likely to take risks, once they become pilots they demonstrate more cautious behavior (i.e., take more calculated risks where they can assess the likelihood of failing).

Siem, F. M. (1990). *Development of a selection model for fighter-qualified USAF pilot candidates*. Paper presented at the 1990 American Psychological Association Annual Convention, Boston, Massachusetts.

Article type: Validation
Keywords: AFOQT; BAT; Military; Personality; Pilot Selection; Response Latencies; Training Performance

The authors state that until recently, USAF pilots had been assigned to either fighter aircraft or tanker-transport aircraft based upon the recommendation of the advanced training recommendation board (ATRB). The purpose of this study was to develop a model to predict which USAF pilot candidates would be considered by the ATRB to be suitable for assignment to fighter aircraft using scores from the Air Force Officer Qualifying Test (AFOQT), the Basic Attributes Tests (BAT), college GPA and an index of candidates' preference for tanker-transport aircraft. A sample of 426 Officer Training School candidates were randomly assigned to one of two subsamples. For each subsample, multiple correlations were computed between the 42 variables and the ATRB recommendation outcome. These regression models were then each cross-validated in the other subsample. The full models for sample A and sample B were both significant ($R = .61, p < .001$; $R = .50, p < .05$, respectively). For each subsample, multiple regression was used to eliminate variables that did not contribute significantly to the prediction of ATRB outcome. For sample A, the final reduced model contained two predictors: AFOQT Instrument Comprehension and Response Time (RT) from the BAT Mental Rotation task ($R = .31, p < .01$). For sample B, the reduced model contained seven predictors: AFOQT Instrument Comprehension and Block Counting; the BAT percent correct and RT variability from the

Mental Rotation test and average tracking difficulty from the Time Sharing test; college GPA; and the desire to fly tanker-transport rating. For the cross-validity results, the model developed for sample A, had a multiple correlation of .22 with ATRB recommendation outcome in subsample B. The model developed in sample B, had a multiple correlation of .26 with ATRB recommendation outcome in sample A. Based on these results a final model was developed that produced a multiple correlation of .40 ($p < .001$).

Siem, F. M. (1990). *Predictive validity of an automated personality inventory for Air Force pilot selection* (AFHRL-TP-90-55). Brooks AFB, TX: AFHRL.

Article type: Validation
Keywords: AFOQT; BAT; Military; Personality; Pilot Selection; Response Latencies; Training Performance

The purpose of this study was to examine the validity of standardized measures of personality for predicting Undergraduate Pilot Training (UPT) outcome (pass/fail) in a sample of 506 UPT candidates. The candidates were, in part, selected into training on the basis of their Air Force Officer Qualifying Test (AFOQT) and the Basic Attributes Tests (BAT). Several additional inventories were administered. The Automated Aircrew Personality Profiler (AAPP) consisted of 202 items drawn from several personality inventories (Minnesota Multiphasic Personality Inventory, State-Trait Anxiety Inventory, Personal Orientation Inventory, Interpersonal Behavior Scale and the Jenkins Activity Survey). Data regarding item endorsement and response time were collected for each item. The items and response times were summed to form scores for five trait dimensions: Emotional Stability, Sociability, Extraversion, Competency and Cynicism. Latency (i.e., response time) scores were computed separately for endorsed and non-endorsed items. A factor analysis (with oblique rotation) was conducted and used to develop 5 personality scales from the AAPP: Hostility, Self-Confidence, Values Flexibility, Depression and Mania. Three of these scales were significantly correlated with UPT outcome, Hostility ($r = -.12, p < .05$), Self-Confidence ($r = .13, p < .01$) and Values Flexibility ($r = .12, p < .05$). The five personality scales did not contribute to the prediction of UPT outcome beyond the AFOQT and BAT alone. Even after the two tests measuring personality constructs (Self-Crediting Word Knowledge and Activities Interest Inventory) were removed from the BAT, the five AAPP scales still did not contribute unique variance to the prediction of UPT outcome.

Siem, F. M. (1991). *Predictive validity of response latencies from computer-administered personality tests.* Paper presented at the 33rd annual conference of the Military Testing Association, 28-31 October, 1991.

Article type: Validation
Keywords: AFOQT; BAT; Military; MMPI; Personality; Pilot Selection; Response Latencies; Training Performance

In this paper, past research (e.g., Holden & Fekken, 1987) is discussed showing that individuals who scored high on a personality trait manifested shorter response latencies than individuals who scored low on that trait on items that were "endorsed." The opposite was found for items that were "rejected." Some interpret these results in terms of a "self-schema," which is an integrated network of self knowledge. Five hundred and nine Air Force student pilots entering Undergraduate Pilot Training (UPT; 332 for which training outcome was available) participated in this study. The Automated Aircrew Personality Profiler (AAPP) was administered and includes 94 items from the Minnesota Multiphasic Personality Inventory (MMPI). Scores were computed on the following scales: Sociability (Psychoticism), Emotional Stability (Neuroticism), Extraversion, Competency (Inadequacy) and Cynicism. The criterion used was UPT training outcome (pass/fail). Response latencies were standardized within subjects to control for reading speed and by item to control for item length and vocabulary level. For each subject and each scale, a mean response time for endorsed items and one for rejected items was computed. Correlations between personality scores and UPT outcome were fairly weak. The only significant correlations with UPT were for Sociability ($r = .13, p < .01$) and Cynicism ($r = -.14, p < .01$). Correlations between response time scores with the criterion were only significant for endorsed items and then only for Extraversion and Cynicism. The author noted that the internal consistency for the sociability scale was only .34. Correlations between scale and response time variables for each trait were relatively small (ranging from .02 to .35). Correlations between the two response time measures (i.e., endorsed and rejected) for each trait were also small (ranging from -.01 to .-.26).

Siem, F. M. (1992). *Incremental validity of personality measures for predicting pilot training performance.* Paper presented at the annual convention of the American Psychological Association, Washington, DC, August, 1992.

Article type: Validation
Keywords: AFOQT; BAT; Military; Personality; Pilot Selection; Response Latencies; Training Performance

Past research (e.g., Siem, 1991) has indicated that personality measures based on inventory response latencies, as well as scale scores, have validity for predicting pilot training performance. The purpose of this study was to examine the incremental validities of these measures in predicting training performance. Specifically, two issues were addressed: (1) the factor structure of the scale scores and response latency scores were examined, with the expectation that the measures of the same trait would load on the same factor; and (2) the extent to which these

personality measures would contribute unique variance to the prediction of training outcome was examined. Three hundred thirty-two college graduates who had been selected for Air Force pilot training were included in the factor analyses, 277 of these students were subsequently included in the multiple regression analyses. Scores from three predictors were used in this study: (1) the Basic Attributes Tests (BAT) - accuracy and response time scores were collected for the tests and psychomotor tasks; (2) the Automated Aircrew Personality Profiler (AAPP) - data on item endorsement and response time were collected for each item; and (3) the Air Force Officer Qualifying Tests (AFOQT). Undergraduate Pilot Training (UPT) was dichotomously scored pass/fail. Results of the factor analysis related to the first hypothesis showed that one factor was defined by the scale scores for the Emotional Stability, Sociability and Cynicism dimensions. Two factors were defined by loadings both of a scale score and a latency score (Extraversion & Cynicism). The remaining two factors were defined by loadings of latency variables only for Cynicism and Emotional Stability. Results related to hypothesis 2 showed that the full model correlated .36 ($p < .01$) with UPT pass/fail, the model with the five AAPP variables removed resulted in a R-change of .06 ($p < .05$). Thus, the personality variables did add to the prediction of UPT outcome beyond the AFOQT and the BAT. However, further analyses showed that all but one of the five AAPP variables, namely, Extraversion, could be removed from the model without a decrease in validity.

Siem, F. M., Carretta, T. R., & Mercatante, T. A. (1988). *Personality, attitudes and pilot training performance: Preliminary analysis* (AFHRL-TP-87-62). Brooks AFB, TX: Air Force Human Resources Laboratory.

Article type: Validation
Keywords: Military; Personality; Pilot Selection; Response Latencies; Training Performance

The authors note that the cost for each Undergraduate Pilot Training (UPT) eliminee is estimated to be between $65,000 to $80,000. The purpose of this study was to examine the validity of five personality measures for predicting UPT outcome and advanced training recommendation board (ATRB) recommendation in a sample of 883 trainees in the USAF UPT program. Five different measures were administered: the Embedded Figures Test, Dot Estimation, Self-Crediting Word Knowledge, Activities Interest Inventory and Risk-Taking. Two criteria were used: UPT training pass or fail and ATRB recommendations for fighter or non-fighter assignments. A series of multiple regressions were conducted. For each regression, predictors were from the five personality measures, with the criterion being either UPT pass/fail or advanced training recommendation. The only test to significantly predict UPT pass/fail was the Self-Crediting Word Knowledge Test ($R = .14, p < .01$). For this test, the subjects who passed UPT tended to take longer to respond, had fewer correct responses and bet fewer points on their responses. None of the other four tests were correlated with UPT pass/fail and none of the five tests were correlated with ATRB recommendation.

Signori, D. I. (1949). The Arnprior experiment: A study of world war II pilot selection procedures in the RCAF and RAF. *Canadian Journal of Psychology, III*(3), 136-150.

Article type: Validation
Keywords: Attitudes; Background; Biodata; Military; Motivation; Pilot Selection; Psychomotor Ability; Training Performance

This article presents validity evidence for the pilot selection battery employed by the Royal Canadian Air Force (RCAF). This battery includes a variety of predictor variables ranging from cognitive ability to biodata to psychomotor coordination. Several different criterion measures were also used ranging from a training pass/fail criterion to various flight training and ground school ranks. Results showed that the RCAF Visual Link Test was the most valid predictor variable (zero-order validities ranging from .18 for initial training school rank to .57 for elementary flying training school rank). The validities for the Aircrew Information Sheet (biodata) ranged from .01 for elementary and service flight training school to .20 for initial training school rank. The validities for the Aircrew Interview Report Form (motivation and attitude appraisal) ranged from .06 for elementary flight training school pass/fail to .16 for initial training school rank. The authors conclude that the RCAF battery compares favorably with the 1943 U.S. Air Force selection battery.

Street, D. R., Jr., Helton, K. T., & Dolgin, D. L. (August, 1992). *The unique contribution of selected personality tests to the prediction of success in naval pilot training* (NAMRL-1374). Pensacola, FL: Naval Aerospace Medical Research Laboratory, Naval Air Station.

Article type: Validation
Keywords: Attitudes; Cognitive Ability; Military; Motivation; Personality; Pilot Selection; Psychomotor Ability; Training Performance

The authors report that aircrew selection research has typically focused on psychomotor and cognitive abilities, but evidence from flight training attrition studies suggests that many failures may be due to personality or motivational factors. This study examined the relationship between naval flight training performance and scores on the Aviation Qualification Test/Flight Aptitude Rating (AQT/FAR) and the automated Pilot Personality Questionnaire (PPQ). The AQT is a single test that measures general cognitive abilities. The FAR consists of three different tests: the Mechanical Comprehension Test (MCT), the Spatial Apperception Test (SAT) and the Biographical Inventory (BI). The PPQ contains 112 multiple choice items administered using a computer and a keyboard. The 112 items are drawn from four different instruments: (1) the Locus of Control scale (LOC); (2) the Work and Family Orientation scale (WOFO); (3) the Personality Attributes Questionnaire (PAQ); and (4) the Social Desirability Scale (SDS). These four inventories include 12 scales: (1) self-assertiveness; (2) interpersonal orientation; (3) aggressiveness; (4) hostility; (5) verbal aggressiveness; (6) submissiveness; (7) high-mastery motivation; (8) high work motivation; (9) competitiveness; (10) self control; (11) fatalism; and (12) high-social desirability. The researchers used discriminant analysis to identify a linear composite of the AQT/FAR and PPQ subtest variables that could be used to classify students into

pass/attrite flight training categories. They also examined the mean differences for all variables between those students who passed training and those who attrited. Results showed that the mean differences were significant for the AQT/FAR (and the MCT and SAT subtests of the FAR) and the PPQ competitiveness scale. The stepwise discriminant function identified the AQT/FAR and PPQ competitiveness scale as adding significant variance to the pass/fail prediction equation. The authors concluded that the PPQ competitiveness scale predict variance not otherwise accounted for by the AQT/FAR. The FAR and the PPQ competitiveness scale were the best predictors of overall flight training success.

Taylor, C. W., Murray, S. L., Ellison, R. L., & Majesty, M. S. (1971). *Development of motivation assessment techniques for Air Force officer training and education programs: Motivation for pilot training* (AFHRL-TR-71-21). Brooks Air Force Base, TX: Professional Education Division, Air Force Human Resources Laboratory.

Article type: Validation
Keywords: Attitudes; Background; Biodata; Military; Motivation; Personality; Pilot Selection; Training; Turnover/Attrition

The authors attempted to develop a motivational screening device that would reduce the number of voluntary eliminees from undergraduate pilot training (UPT) and to develop an instrument which could be used to measure changes in motivation as subjects progress through pilot training (and later on the job). Both an a priori and an empirical keying approach were used to develop measures intended to predict several different types of attrition. Two hundred twenty-four self initiated eliminees (SIE) from UPT identified their reasons for attriting and these were subsequently examined to develop subcategories (i.e., criterion measures) that could be used to classify eliminees (e.g., SIE, flying deficiency, medical elimination, etc.). The initial item pool for predictor battery development contained 2500 items, 899 of which were included in the first version of the battery. The predictor battery consisted of two biographical inventories (BIs) with approximately 300 items each and an Activities Index, which was a 300 item personality test. The battery was administered to all active students in the Officer training School (OTS), which was operated at Lackland AFB between January and November 1970. Of the a priori scales, only the BI creativity key demonstrated predictive validity with total attrition ($r = -.10, p < .05$). All other validities for a priori BI keys and for all of the other criteria were non-significant. Of the a priori personality scales, Abasement, Change, Energy, Science, Sensuality, Audacity and Motivation all demonstrated marginal negative validity for several of the attrition measures (generally on the order of -.10 to -.16). The validities for the Harm Avoidance scale ranged from .09 (*ns*) for elimination due to flying deficiency to .18 ($p < .01$) for UPT/OTS-SIE. Supplication also correlated positively with attrition, although the validity coefficients were generally small. The empirically derived scales resulted in moderately large validity coefficients (generally in the .20s and .30s) for almost all the attrition criteria, except for elimination due to flying deficiency. They also correlated significantly with self-rated lack of dedication to complete UPT (correlations ranged from .38 ($p < .01$) for flying deficiency to .72 ($p < .01$) for UPT/OTS-SIE). Generally, the a priori scales only accounted for a small proportion of attrition variance. Significantly more positive results were obtained using the empirically developed scales.

Terelak, J. F. (1991). Current research and trends in aviation psychology in Poland. *Aviation, Space and Environmental Medicine, 62*, 903-906.

Article type: Review
Keywords: Attitudes; Cognitive Ability; Military; Personality; Pilot Selection; Psychomotor Ability

This paper reviews current aviation psychology research in Poland. This research has focused on four different areas. First, the personality determinants of the occupational functioning of pilots have been studied, and this has covered three areas: the extent to which some temperaments may change as a psychological cost of working under conditions of extreme stress; the personality determinants of the occupational adjustment of pilots using the Scale of Attitudes Towards Aviation Military (SATAM); and, the personality determinants of the psychosomatic costs of this adjustment. The second focus has been on the perceptual efficiency mechanisms of task performance in flight simulators. This research has mostly focused on visual aspects of pilot performance. Third, the psychological criteria which predict success in pilot training have been examined. This research has reported moderate correlations (.22 to .54) between a measure of dual-tasking ability (i.e., visual and auditory) and flight training performance. Also, there have been some reports of success for using a variety of tests (e.g., Eysenck's MPI, Speilberger's STAI, etc.) to predict student pilots' progress in various stages of flight training. Finally, the computer assisted psychological research has been investigated. The authors recommend computerizing a battery of non-cognitive, cognitive and perceptual tests to be used in pilot selection.

Termohlen, J. (1984). Conceptual model of critical requirements for effective pilot performance in high performance aircraft. *In Human factors considerations in high performance aircraft, Conference Proceedings of the Aerospace Medical Panel Symposium*, Williamsburg, Virginia.

Article type: Job Analysis
Keywords: Attitudes, Military; Job Performance; Personality; Pilot Selection

The author conducted a critical incident interview study to update a conceptual model of the critical job requirements of the operational Air Force pilot. One hundred fifty-three subjects, ranging in rank from lieutenant to colonel, were interviewed. Six main trait categories emerged as particularly relevant for effective Air Force pilot performance: (1) action power and competence; (2) knowing one's limitations and not exceeding them; (3) dependability; (4) leadership; (5) self-discipline; and (6) social function and team work. During the interviews, the researcher also asked 17 experienced instructor pilots to weight the main categories according to their importance by distributing 100 points among them. These results showed that action power and competence were most heavily weighted (26 pts), followed by knowing one' limitations (22 pts), dependability (17 pts), leadership (14 pts), self-discipline (11 pts) and social function and team work (10 pts). One noteworthy aspect of this study is that it explicitly delineated information concerning job requirements necessary for effective pilot performance that previously were considered implicit or tacit knowledge.

pass/attrite flight training categories. They also examined the mean differences for all variables between those students who passed training and those who attrited. Results showed that the mean differences were significant for the AQT/FAR (and the MCT and SAT subtests of the FAR) and the PPQ competitiveness scale. The stepwise discriminant function identified the AQT/FAR and PPQ competitiveness scale as adding significant variance to the pass/fail prediction equation. The authors concluded that the PPQ competitiveness scale predict variance not otherwise accounted for by the AQT/FAR. The FAR and the PPQ competitiveness scale were the best predictors of overall flight training success.

Taylor, C. W., Murray, S. L., Ellison, R. L., & Majesty, M. S. (1971). *Development of motivation assessment techniques for Air Force officer training and education programs: Motivation for pilot training* (AFHRL-TR-71-21). Brooks Air Force Base, TX: Professional Education Division, Air Force Human Resources Laboratory.

Article type: Validation
Keywords: Attitudes; Background; Biodata; Military; Motivation; Personality; Pilot Selection; Training; Turnover/Attrition

The authors attempted to develop a motivational screening device that would reduce the number of voluntary eliminees from undergraduate pilot training (UPT) and to develop an instrument which could be used to measure changes in motivation as subjects progress through pilot training (and later on the job). Both an a priori and an empirical keying approach were used to develop measures intended to predict several different types of attrition. Two hundred twenty-four self initiated eliminees (SIE) from UPT identified their reasons for attriting and these were subsequently examined to develop subcategories (i.e., criterion measures) that could be used to classify eliminees (e.g., SIE, flying deficiency, medical elimination, etc.). The initial item pool for predictor battery development contained 2500 items, 899 of which were included in the first version of the battery. The predictor battery consisted of two biographical inventories (BIs) with approximately 300 items each and an Activities Index, which was a 300 item personality test. The battery was administered to all active students in the Officer training School (OTS), which was operated at Lackland AFB between January and November 1970. Of the a priori scales, only the BI creativity key demonstrated predictive validity with total attrition ($r = -.10$, $p < .05$). All other validities for a priori BI keys and for all of the other criteria were non-significant. Of the a priori personality scales, Abasement, Change, Energy, Science, Sensuality, Audacity and Motivation all demonstrated marginal negative validity for several of the attrition measures (generally on the order of -.10 to -.16). The validities for the Harm Avoidance scale ranged from .09 (*ns*) for elimination due to flying deficiency to .18 ($p < .01$) for UPT/OTS-SIE. Supplication also correlated positively with attrition, although the validity coefficients were generally small. The empirically derived scales resulted in moderately large validity coefficients (generally in the .20s and .30s) for almost all the attrition criteria, except for elimination due to flying deficiency. They also correlated significantly with self-rated lack of dedication to complete UPT (correlations ranged from .38 ($p < .01$) for flying deficiency to .72 ($p < .01$) for UPT/OTS-SIE). Generally, the a priori scales only accounted for a small proportion of attrition variance. Significantly more positive results were obtained using the empirically developed scales.

Terelak, J. F. (1991). Current research and trends in aviation psychology in Poland. *Aviation, Space and Environmental Medicine, 62*, 903-906.

Article type: Review
Keywords: Attitudes; Cognitive Ability; Military; Personality; Pilot Selection; Psychomotor Ability

This paper reviews current aviation psychology research in Poland. This research has focused on four different areas. First, the personality determinants of the occupational functioning of pilots have been studied, and this has covered three areas: the extent to which some temperaments may change as a psychological cost of working under conditions of extreme stress; the personality determinants of the occupational adjustment of pilots using the Scale of Attitudes Towards Aviation Military (SATAM); and, the personality determinants of the psychosomatic costs of this adjustment. The second focus has been on the perceptual efficiency mechanisms of task performance in flight simulators. This research has mostly focused on visual aspects of pilot performance. Third, the psychological criteria which predict success in pilot training have been examined. This research has reported moderate correlations (.22 to .54) between a measure of dual-tasking ability (i.e., visual and auditory) and flight training performance. Also, there have been some reports of success for using a variety of tests (e.g., Eysenck's MPI, Speilberger's STAI, etc.) to predict student pilots' progress in various stages of flight training. Finally, the computer assisted psychological research has been investigated. The authors recommend computerizing a battery of non-cognitive, cognitive and perceptual tests to be used in pilot selection.

Termohlen, J. (1984). Conceptual model of critical requirements for effective pilot performance in high performance aircraft. *In Human factors considerations in high performance aircraft, Conference Proceedings of the Aerospace Medical Panel Symposium*, Williamsburg, Virginia.

Article type: Job Analysis
Keywords: Attitudes, Military; Job Performance; Personality; Pilot Selection

The author conducted a critical incident interview study to update a conceptual model of the critical job requirements of the operational Air Force pilot. One hundred fifty-three subjects, ranging in rank from lieutenant to colonel, were interviewed. Six main trait categories emerged as particularly relevant for effective Air Force pilot performance: (1) action power and competence; (2) knowing one's limitations and not exceeding them; (3) dependability; (4) leadership; (5) self-discipline; and (6) social function and team work. During the interviews, the researcher also asked 17 experienced instructor pilots to weight the main categories according to their importance by distributing 100 points among them. These results showed that action power and competence were most heavily weighted (26 pts), followed by knowing one' limitations (22 pts), dependability (17 pts), leadership (14 pts), self-discipline (11 pts) and social function and team work (10 pts). One noteworthy aspect of this study is that it explicitly delineated information concerning job requirements necessary for effective pilot performance that previously were considered implicit or tacit knowledge.

Trites, D. K., & Sells, S. B. (1957). Combat performance: Measurement and prediction. *Journal of Applied Psychology, 41*, 121-130.

Article type: Validation
Keywords: Cognitive Ability; Military; Motivation; Performance Ratings; Personality; Pilot Selection; Training Performance

This research evaluated several measures of the combat effectiveness of USAF pilots, and the utility of data collected while the pilots were in flight training in predicting those criteria. Specifically, three research questions were asked: (1) Are ratings of combat performance by peers, supervisors and psychologists related to objective performance information? (2) To what extent are peer-supervisor ratings related to psychologist ratings (which are, in part, based upon information obtained from peers and supervisors)?; and (3) Are combat criteria predictable using pre-combat criteria of performance and adjustment? Sixty-five pilots who were in the Far Eastern Theater of Operations (Korea) and who had relatively complete training and overseas performance information participated. A variety of performance measures from both combat and training were used. The combat variables included peer-supervisor ratings of various combat-related criteria (e.g., competence in combat, fairness, courage, responsibility, etc.), psychologists ratings of symptomatology, and overall ratings of pilots' adjustment and performance in combat. Objective training measures included the pilot stanine (essentially flying aptitude), the officer quality stanine and age. In addition, peer-superior ratings of a variety of tactical and instructional measures were included, as well as some additional adjustment ratings by psychologists. The authors conclude that ratings of combat performance and adjustment made by peers, superiors and psychologists were related to one another and to objective performance criteria. They also demonstrated that combat criterion measures were predictable using pre-combat criteria of performance and psychological adjustment.

Turnbull, G. J. (1992). A review of military pilot selection. *Aviation, Space and Environmental Medicine, 63*, 825-830.

Article type: Review
Keywords: Cognitive Ability; Military; Personality; Pilot Selection; Training

The author reviews the various methods that have been used to select pilots over the years. This comprehensive review begins with a description of the "selection" method used by the Wright brothers in 1913 (i.e., they flipped a coin) to decide who would fly their plane and includes the present-day multivariate approach to selecting pilots. He describes the historical depiction of pilots as sportsmen with initiative and humor, as well as high-spirited, gregarious and lacking in imagination. The RAF developed the first basic aptitude tests used to select pilots during WWI, but these efforts receded after the war ended. However, by 1941 the high incidence of failure in RAF pilot training became a major problem and the focus on pilot selection once again increased. This cycle is contrasted against the remarkable success of the German Luftwaffe selection procedures. The Germans used a combination of psychological and aptitude testing and were very successful in selecting pilots. The author reviews the use of personality tests in general

and concludes that they have had some success in predicting pilot training performance, but that their susceptibility to faking is a major failing. The use of "objective" tests (e.g., the Rod and Frame test) is also reviewed. He concludes that while there is no one reliable way to select pilots, there is considerable evidence suggesting that the measurement of the "right stuff" is feasible.

Ursano, R. J. (1980). Stress and adaptation: The interaction of the pilot personality and disease. *Aviation, Space and Environmental Medicine, 51*(11), 1245-1249.

Article type: Review
Keywords: Military; Psychiatric; Psychological Status

The author notes that there is a growing body of literature recognizing that one's health is a reflection of a complex interaction between one's body, one's psychological system and one's sociocultural environment. Several different perspectives on the interaction between mind, body and environment are reviewed, showing that many psychological stressors can cause physiological illness (and vice versa). Thus, viewing illness from a static biological perspective is often times inappropriate. The author presents case studies of three pilots, in which psychological problems resulted in physiological conditions which, in turn, affected flying performance. He emphasizes the importance of understanding the temporal nature of disease and in utilizing a comprehensive model that accounts for environmental and psychological determinants of disease, as well as the conventional biological determinants.

Værnes, R. J., Myhre, G., Aas, H., Homnes, T., Hansen, I., & Tonder, O. (1991). Relationships between stress, psychological factors, health and immune levels among military aviators. *Work & Stress, 5*(1), 5-16.

Article type: Single Group Study
Keywords: Civilian; CRM; Leadership; Military; Personality; Psychiatric; Psychological Status; Stress

The authors examine whether immunoglobin levels are related to work stress and personality factors in a sample of 64 healthy aviators (31 fighter pilots and 33 transport pilots), and whether these factors are in turn related to level of health complaint. Two different types of measures were administered. The first set assessed psychological characteristics of the subjects and consisted of the Life Style Index (LSI) and the State-Trait Anxiety Inventory (STAI). The Life Style Index (LSI) has 92 items measuring various psychological defensive strategies. The State-Trait Anxiety Inventory (STAI) has two scales with 20 items each. The "state anxiety" scale measures transitory emotional states that vary in intensity over time. The "trait anxiety" scale measures relatively stable "anxiety proneness." The second set of inventories measured subjects' work environments. Cooper's Job Stress Questionnaire is comprised of 22 items that measure feelings of stress in the work situation. Autonomy was also measured using a four item scale and job involvement was measured using an eight item scale. Ursin's health inventory was also administered, via a structured interview to examine the presence of 30 different health symptoms

during the past month and the past three years. Various immunoglobulins were also measured. Results showed that immune levels and perceived health complaints were related to work stress and personality factors. Subjects who scored high on projection, compensation and regression reported more work related stress. Projection was also related to dissatisfaction with colleagues and organization. Degree of health complaints was low compared to other groups (e.g., fire brigade workers), however, consumption of alcohol was higher than other groups. The authors conclude that people who use defensive strategies tend to demonstrate inadequate performance followed by high autonomic activation when performing tasks which require swift and accurate cognitive performance under conditions of stress. These subjects tended not to cope well during training and developed psychosomatic problems over the long-term. The authors recommend implementing training for cockpit management, leadership and general teamwork that includes feedback on emotional behavior.

Viteles, M. S. (1945). The aircraft pilot: 5 years of research and a summary of outcomes. *Psychological Bulletin, 42*, 489-526.

Article type: Review
Keywords: Attitudes; Background; Biodata; Civilian; Cognitive Ability; Motivation; Personality; Pilot Selection; Psychomotor Ability

This article describes a massive research project undertaken by the Civil Aeronautics Authority (CAA) in 1939 "to make young men and young women air-minded and to prepare the present generation of young people to fly the private and commercial planes of the future" (p. 489). The article primarily consists of detailed descriptions of a wide range of selection tests, including tests of physical abilities, general cognitive ability, personality, biodata, attitudes, somatotyping, mechanical comprehension, desire-to-fly, educational accomplishments, perceptual/psychomotor ability, spatial ability and job samples. No validity data are presented.

Voas, R. B. (1959). Vocational interests of naval aviation cadets: Final results. *Journal of Applied Psychology, 43*, 70-73.

Article type: Group Comparison Study
Keywords: Military; Training; Training Performance; Turnover/Attrition

This study attempts to determine whether interest patterns, as measured by the Kuder Preference Record (KPR), differ between students who successfully complete flight training and those who do not. The KPR is a standard interest inventory that consists of nine scales: Mechanical, Computational, Scientific, Persuasive, Artistic, Literacy, Musical, Social Service and Clerical. This was a follow-up study for work completed by Rosenberg and Izard (1954), which found that students who failed flight training scored lower on the Mechanical and Scientific scales and higher on the Persuasive, Literacy and Musical scales than did those students who passed flight training. Six hundred and five aviation cadets participated: 465 who successfully completed training, 74 who withdrew at their own request and 66 who failed at some point during flight

training or were eliminated for medical/physical reasons. The author constructed an empirically keyed Voluntary Withdrawal (VW) scale based on results from Rosenberg and Izard. For those scales that differentiated successful cadets from unsuccessful cadets, a score of "1" was assigned to those items that unsuccessful cadets endorsed and a score of "0" was assigned to those items that successful cadets endorsed. A high score on the VW scale indicated a pattern of interests similar to cadets who had failed flight training. Results of this study showed that of the nine KPR scales, significant differences between successful and unsuccessful cadets were only found for the Mechanical and Scientific scales. In addition, the cadets who voluntarily withdrew from training scored significantly higher on the VW scale than cadets who succeeded, although there was quite a bit of shrinkage in the cross validity sample ($r = .56$ compared to $r = .17$). When scores on the Mechanical Comprehension Test were held constant, the differences between the successful and unsuccessful cadets on the VW scale became nonsignificant. Thus, the validity of the VW scale was apparently based mostly on its relationship to mechanical ability.

Voas, R. B., Bair, J. T., & Ambler, R. K. (1956). Relationship between behavior in a stress situation and later separation from flight training with expressed anxiety toward flying. *Psychological Reports*, 2, 393-397.

Article type: Group Comparison Study
Keywords: Anxiety; Military; Pilot Selection; Stress; Training Performance; Turnover/Attrition

This study examines the effects of stress tolerance on a purified success in training criterion. Subjects were 1540 flight trainees who were participating in decompression chamber classes. Stress tolerance was operationalized as whether or not subjects replaced their air mask during a chamber ride or experienced ear blocks. Flight training criteria included poor performance leading to failure in training or voluntary withdrawal, and an indication of "anxiety toward flying" (assessed in an unstructured exit interview). Subjects were classified into one of three groups at the end of training: (1) "S" group were those who successfully completed training; (2) "P" group were those who failed (this group was subsequently divided into "PA" - poor performance with anxiety and "PN" - poor performance without anxiety); and (3) "M" group made up of all subjects who did not complete training for various other reasons. Results indicated that those subjects who demonstrated anxiety in the chamber ride were more likely to demonstrate anxiety during flight training, that is, the S group demonstrated considerably less anxiety than the P group during the chamber ride. In addition, the PN group demonstrated less anxiety than did the PA group. Also, the PN group did not differ from either the S or M groups in level of anxiety.

Walters, L. C., Miller, M. R., & Ree, M. J. (1993). Structured interviews for pilot selection: No incremental validity. *International Journal of Aviation Psychology, 3*(1), 25-38.

Article type: Validation
Keywords: AFOQT; Background; BAT; Cognitive Ability; Leadership; Motivation; Personality; Pilot Selection; Psychomotor Ability; Training Performance

The authors examined the incremental validity of a structured interview over and above that of aptitude tests alone for selecting pilots. The structured interview (SI) yielded several ratings of subject attributes: Educational Background, Self-Confidence/Leadership, Flying Motivation, success in undergraduate pilot training [UPT] and success in flying different classes of aircraft. The sample consisted of 223 subjects who had just started advanced USAF pilot training. All were commissioned through the Reserve Officer Training Corps or Officer Training School. Subjects had already been preselected on the basis of their Air Force Officer Qualifying Test (AFOQT) scores. Aptitude in this study was measured by the AFOQT and the Basic Attributes Tests (BAT). In addition, a self-report measure of the number of flying hours was also collected. The criterion was a dichotomous pass/fail outcome for a 53-week pilot training course (T-38; 79% of subjects passed training). Correlation and regression analyses were used to evaluate the incremental validity of the SI. All of the SI ratings except for the Self-Confidence/Leadership and Flying Motivation were significantly related to the pass/fail criterion (fully corrected significant validities ranged from .31 for Educational Background to .29 for both Success in Training and Success in Flying different classes of aircraft). However, the regression analyses also showed that the SI did not add any incremental validity over that provided by the AFOQT and the BAT. In addition, some of the partial regression coefficients for the SI ratings were weighted negatively (i.e., Self-Confidence, Motivation and UPT success). The authors suggest that more appreciable incremental validity may be found for the SI if it's relationship to actual job performance were examined (e.g., after the trainees had flown operationally for several years).

Wetzler, H. P., Ursano, R. J., & Cruess, D. F. (1983). Psychological well-being in United States Air Force fliers. *The Journal of Nervous and Mental Disease, 171*, 342-347.

Article type: Group Comparison Study
Keywords: Classification; Military; Psychological Status

The authors note that military fliers must meet medical requirements for acceptance in training (and for continuation), but these standards are mostly physical standards and largely ignore other determinants of health [e.g., psychological well-being (PWB)]. In this study, PWB was characterized as a two-dimensional construct, namely Positive Feelings (PFs) and Negative Feelings (NFs). This study examined the PWB of USAF fliers, USAF non-flying officers and a civilian group using an adaptation of Bradburn's PWB scale. Data were collected as part of a Health Survey in a cross-section of USAF personnel (N = 6675). The Health Survey consisted of 78 diverse, multiple-choice items. In addition to demographic items, the survey collected information regarding Air Force job classification preferences, health practices, dietary attitudes, coronary prone behavior and eight PWB items. Results showed that the Air Force fliers had

uniformly better PWB scores than both the Air Force nonfliers and the civilian group. The authors did not find a correlation between education and PWB, which conflicts with earlier findings. Active pilots showed significantly greater PWB than any other groups (active navigators, inactive fliers and nonfliers). Pilots and navigators had different patterns in the two components of PWB (PF and NF). Navigators NF patterns resembled those of pilots and PF resembled those of nonfliers. The authors conclude that there are differences between PWB of fliers and nonfliers that are not due to age or educational level and that PWB could be an important predictor of long-term flying performance.

III. Situational Judgment Test Reviews

Executive Review

One particularly promising approach to measuring individual differences in the interpersonal and personality areas is the situational judgment test. Situational judgment tests present respondents with a series of job-relevant situations and ask them to indicate which of several alternative actions would be most effective and which would be least effective in each situation. These tests are based on the premise that there are important and often subtle differences between the behavior of effective and ineffective persons as they respond to problems or dilemmas confronted in the course of carrying out their job responsibilities, and that such differences are reflected in their responses to similar situations presented in written form.

Several researchers (e.g., Motowidlo, Russell, Carter, & Dunnette, 1990; Wagner & Sternberg, 1985) have speculated that situational judgment tests actually tap "practical intelligence" or "street smarts" -- the kind of generalized practical knowledge that is central to success in dealing with the problems and dilemmas occurring in daily life and in a wide range of job circumstances. Such "practical intelligence" related to interpersonal problems is likely to be extremely important for success as a pilot, both while in the air and while carrying out other duties on the ground. While there is not a great deal of research available on situational judgment tests, researchers have shown that situational judgment tests can be valid predictors of job performance and that these tests show the expected patterns of correlations with a variety of different job performance measures (e.g., Forehand & Guetzkow, 1961; Hanson & Borman, 1990; Motowidlo et al., 1990; Phillips, 1992; Tenopyr, 1969). Because these tests do not rely on applicants' self reports, the serious problems with response distortion that plague traditional personality measures can be avoided. Situational judgment tests are also typically drawn directly from job content and are thus much more palatable to candidates than most other types of selection instruments.

We believe that a situational judgment test is especially promising for use in selecting Air Force pilots and will be of considerable interest to airlines and many other industries for several reasons. First, as noted by Motowidlo, Dunnette and Carter (1990), situational judgment tests can be used to capture the predictive potential of behavioral samples while avoiding the excessive costs associated with more labor intensive simulations such as assessment centers. Unlike assessment center exercises, a situational judgment test can be efficiently administered to large groups of applicants simultaneously. Second, correlations between situational judgment tests and measures of general ability and academic achievement are generally low to moderate (e.g., Motowidlo et al., 1990). Thus, in all likelihood, such a test would not only be a valid predictor of performance, but would also account for incremental variance in performance over and above that accounted for by the AFOQT and the BAT. Third, situational judgment tests tend to have little adverse impact against protected groups (e.g., Motowidlo et al., 1988; Motowidlo et al., 1990; Motowidlo & Carter, 1990). Research has demonstrated only small, nonsignificant differences between the mean scores of black and white test takers on situational judgment tests. Cognitive ability tests, on the other hand, typically yield ethnic and racial subgroup differences ranging up to a full standard deviation. Only slight differences between the mean situational judgment test

scores of men and women have been found, with the mean scores of women slightly higher than those of men.

Annotated References

Bruce, M. M., & Learner, D. B. (1958). A supervisory practices test. *Personnel Psychology, 11*, 207-216.

Article type: Test Development
Keywords: Cognitive Ability; Civilian; Performance; Situational Judgment; Supervision

This article describes the development and evaluation of the Supervisory Practices Test, a situational judgment test designed to aid in the selection of managerial and supervisory personnel. The initial experimental form containing 100 questions was administered to several samples of nonsupervisors, managers and executives. A total of fifty items were retained based on the adequacy of the distracters and the extent to which they discriminated between supervisors and non-supervisors. The fifty item form was administered to large samples of supervisors and nonsupervisors and a key was developed based on both the extent to which alternatives differentiated between supervisors and non-supervisors and the degree to which the majority of *both* groups chose the alternative. In yet another set of large samples, supervisors and nonsupervisors had significantly different mean scores on this test. Test-retest and internal consistency reliabilities were in the 80s. Several validation studies were conducted using *very* small samples which showed moderate correlations between scores on the Supervisory Practices Test and supervisor rankings and subordinate attitudes. In yet another large sample the correlation of the new test with a test of cognitive ability was .27 and the correlation with a similar test, *How Supervise?* (File, 1945), was .56.

Dalessio, A. T. (May, 1992). *Predicting insurance agent turnover using a video-based situational judgment test.* Paper presented at the seventh Annual Conference of the Society for Industrial and Organizational Psychology, Inc., Montreal.

Article type: Test Development
Keywords: Civilian; Situational Judgment; Turnover

This article describes the development and *longitudinal* validation of a video-based situational judgment test designed to predict insurance agent turnover. Fifteen existing video scenarios were used to present the sales situations and a single multiple choice question was developed for each of these situations. Response options were developed by asking sales persons to generate actions or responses that would be effective, moderately effective and ineffective for addressing the client objections raised in each scenario. Then, a large sample of insurance agents and their managers rated the effectiveness of each response option and these ratings were used to pare down the number of response options for each question to five. The validation study involved administering this test to almost 700 newly hired insurance agents and asking them which

response they would be most likely to take and which they would be least likely to take. They divided this sample into thirds for cross validation and developed empirical keys to predict turnover. Cross-validities of these empirical keys were low (about .12) but significantly different from zero. They also factor analyzed the response alternative level scores and develop seven dimension scores based on these results. The adjusted multiple correlation between the seven dimension scores and turnover was also .12. They do not report validity results for the rational key (described previously).

Decker, R. (1956). An item analysis of *How Supervise?* using both internal and external criteria. *Journal of Applied Psychology, 40*, 406-411.

Article type: Validation
Keywords: Attitudes, Civilian; Performance; Performance Ratings; Civilian

A literature review indicated that *How Supervise?* is widely accepted as a measure of supervisory knowledge, but relatively little research has examined its effectiveness as a selection tool. This study investigated the concurrent validity of *How Supervise?* as a measure of supervisory ability. Item difficulty, item validity and item-total correlations were also examined. Subjects were 208 supervisors from a large manufacturing organization (all male). *How Supervise?* (Form M) consists of 100 items and taps knowledge/attitudes from three broad categories: Supervisory Practices, Company Policies and Supervisor Opinions. The performance criterion was an overall score from a 60-item rating instrument (split-half reliability = .89). The correlation between *How Supervise?* total score and overall performance was only .11 (*ns*). Item difficulties ranged from .01 to .64 (median = .10). Item-total correlations ranged from .00 to .84 (median = .40). Item validities ranged from -.26 to .70 (median .07). A subset of 25 items from *How Supervise?* had validities significant at the .01 level. Correlations of a subscore consisting of these 25 items correlated .35 ($p < .01$) with overall performance (not cross-validated). Supervisory practices of supervisors that were rated as "less successful" are presented. The supervisory practices identified tend to be those that are more autocratic, authoritarian and confrontational in nature. Author concludes that *How Supervise?* may be more useful for selecting out those individuals who unsuited for supervisory positions, rather than selecting in those who would be effective supervisors.

DuBois, P. H., & Watson, R. I. (1950). The selection of patrolmen. *Journal of Applied Psychology, 34*, 90-95.

Article type: Test Development and/or Evaluation
Keywords: Cognitive Ability; Civilian; Performance; Situational Judgment; Training Performance

This article describes the development and validation of a face-valid test for selecting patrolmen (e.g., police photos and excerpts from training manuals were included in some sections). A total of 90 multiple choice items were developed to measure memory, spelling, reading arithmetic and

"general information and judgment related to police work." A sample of 129 patrolmen participated in a longitudinal validation study. The Army General Classification Test and the Cornell Word Form 2 (designed to measure neurotic tendencies) had been used to select these patrolmen. Criteria included both training and job performance. Correlations between the judgment section and final training (i.e., Academy) grades were .32 ($p < .01$) and .37 ($p < .01$) for the two training classes included in this study. Validities for a wide variety of cognitive ability tests against training grades ranged from .18 (*ns*) to .55 ($p < .01$). Scores on the experimental test correlated .36 ($p < .01$) with scores on an achievement test and these scores were not significantly correlated with the two job performance measures included: marksmanship and service ratings made by supervisors on 12 different traits (e.g., attitude, loyalty, report writing, judgment, etc.) after 10 weeks on duty.

File, Q. W. (1945). The measurement of supervisory quality in industry. *Journal of Applied Psychology, 29*, 323-337.

Article type: Test Development and/or Evaluation
Keywords: Attitudes; Civilian; Performance; Performance Ratings; Situational Judgment; Supervision

Describes the construction and validation of File's *How Supervise?*. This instrument was designed to measure factors important to effective supervision that are generally important across most supervisory positions. Several principal assumptions underlie the development of this test: (1) the ability to supervise others requires skills that are general in nature and are not necessarily specific to any one given organization and that this effectiveness is largely dependent on one's ability to deal with human relations; (2) the lack of ability to deal with others is the chief cause of supervisory failures; (3) knowledge of how to be an effective supervisor can be measured by a paper-and-pencil instrument and that; (4) questions for such an instrument can be obtained through direct contact of supervisors on the job (i.e., SMEs) and by reviewing literature. Items for *How Supervise?* were selected from three sources: Industrial supervisors/personnel managers, contacts with labor leaders and publications concerning industrial supervision. Two groups of experts were utilized to develop the scoring key, both groups were asked to review/critique the items and to choose which answer they thought was correct. The test was administered to 577 supervisors from a variety of organizations. Ratings from multiple sources (generally, four ratings of each supervisor) for each of these supervisors were obtained using *The Purdue Rating Scale for Supervisors*. Analyses of *The Purdue Rating Scale for Supervisors* suggested that the ratings were of questionable validity (e.g., extremely high levels of halo), thus the ratings were used as a secondary criterion of the test. Generally, the author did not demonstrate any criterion-related validity for *How Supervise?*. The only validity evidence presented is essentially content-validity evidence in that there was high interrater agreement between the judgments of the two groups of experts concerning whether the items from *How Supervise?* were related to effective supervisory performance and if the wording of the items was unambiguous.

File, Q. W., & Remmers, H. H. (1946). Studies in supervisory evaluation. *Journal of Applied Psychology, 30*, 421-425.

Article type: Group Comparison Study
Keywords: Attitudes; Civilian; Performance; Situational Judgment

Presents preliminary validity evidence for using *How Supervise?* as a selection device. One form of *How Supervise?* was given to a group of 46 "successful" supervisors and a group of 14 non-supervisors (these subjects were by-passed for supervisory positions because of judged lack of ability). Eighty percent of the successful supervisors and only 15 percent of the unsuccessful supervisors scored above the 50th percentile on *How Supervise?*. In a different study, the Supervisor of Testing from a manufacturing company reported a split-half reliability of .85 for *How Supervise?* in a sample of 50 subjects. In this same company, an extreme groups design was again used to compare the performance of Superior n = 20) and Inferior (n = 20) supervisors on *How Supervise?*. Once again, the Superior group performed highly significantly better than the Inferior group. A third study compared performance on *How Supervise?* between supervisors classified into one of five groups. Ranging from supervisors in which the company had complete confidence to handle all types of supervisory problems to those who could only handle the most limited supervisory responsibilities. Again, the more effective the supervisor, the better their score on *How Supervise?*.

Forehand G. A., & Guetzkow, H. (1961). The administrative judgment test as related to descriptions of executive judgment behaviors. *Journal of Applied Psychology, 45*, 257-261.

Article type: Validation
Keywords: Cognitive Ability; Civilian; Performance; Performance Ratings; Situational Judgment; Supervision

This article describes a concurrent validation study of the Administrative Judgment Test (AJT), which is a situational judgment test that "attempts to measure broad understanding of the processes of administration . . . whether government or private" (p. 257). The AJT is a 55-item, multiple choice test. The sample used in the validation research included over 100 government administrators. Supervisor and peer ratings were collected using forty scales, 39 tapping decision-making capabilities and styles and one global rating. These were grouped into seven cluster combinations based on cluster analyses conducted in previous research. For each administrator in the sample, ratings were collected from one supervisor and one "peer" or coworker. Ratings were statistically adjusted to remove "halo," and all analyses were conducted once using the unadjusted ratings and once using the adjusted ratings. Sixty-seven percent of the 39 individual unadjusted supervisor ratings correlated significantly with the AJT while only 28 percent of the peer ratings correlated significantly. Only a few of the ratings adjusted for halo correlated significantly with the AJT. Results for the cluster combinations showed that the discernment, policy applying, analytic and cautious clusters had the highest correlations with AJT, whereas self-confidence, bureaucratic and policy execution had the lowest correlations with AJT scores. The correlation between the AJT and a composite of all 39 ratings including both

peers and supervisor ratings was .53 ($p < .01$) and the correlation between grade level and this composite was .32 ($p < .01$). This is very comparable with the results of previous AJT research (Mandell, 1959). The Thurstone test of Mental Alertness did not correlate significantly with the composite rating or grade level, but correlated .58 ($p < .01$) with the AJT. The authors conclude that the large number of significant supervisor rating correlations support the notion that the AJT's coverage is broad and that the lower number of significant peer ratings suggest a hierarchical perspective.

Hanson, M. A., & Borman, W. C. (1989, April). Development of a situational judgment test to be used as a job performance measure for first line supervisors. In W. C. Borman (Chair), *Evaluating "Practical IQ": Measurement issues and research applications in personnel selection and performance assessment.* Symposium conducted at the 4th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Article type: Test Development and/or Evaluation
Keywords: Military; Performance; Performance Ratings; Situational Judgment; Supervision

This article describes the development of a situational judgment test (SJT) that was used as a measure of supervisory job knowledge as part of the Army's Project A. The authors report that SJTs have frequently been developed to predict job performance (e.g., managers, supervisors), but very rarely does information about such tests appear in the literature. Also, this paper is the first instance of researchers using a SJT as a criterion measure, as opposed to a predictor measure. The test was developed to measure supervisory knowledge for first-line supervisors serving in the U. S. Army. Development of the test involved asking non-commissioned officers generate a large number of difficult, but realistic situations that Army first-line supervisors faced on their jobs, as well as a wide variety of possible response options for each situation. This process resulted in a collection of 300 situations, of these, 180 were chosen, to be included in the next step of the development process, which involved collecting information regarding the effectiveness of these responses from both novices and experts. Situations and response options were then selected based on these effectiveness ratings from experts and novices and based on the content of the items. For each of these 35 items, the respondents are required to read through a situation and three to five possible response options and then select the most and the least effective response alternative. Several different scoring procedures are presented by the authors. For example, scores involving only the choices of the most effective response, the choices of the least effective response and a linear combination of the two are all reported. They also reported developing a set of scores that may be considered measures of supervisory "style." Validation results are presented in Hanson and Borman (1990).

Hanson, M. A., & Borman, W. C. (1990, November). *A situational judgment test of supervisory knowledge in the U. S. Army.* Paper presented at the 32nd Annual Conference of the Military Testing Association, Orange Beach, AL.

Article type: Validation
Keywords: Military; Performance; Performance Ratings; Situational Judgment; Supervision

This paper describes the development, field test and preliminary construct validation of a situational judgment test (SJT) designed to measure supervisory skill for non-commissioned officers (NCOs) in the U. S. Army. In contrast with most work on SJTs, this test was a *criterion* measure, not a predictor measure. The SJT was developed by asking groups of soldiers similar to the target NCOs to describe a large number of difficult, but realistic situations that Army first-line supervisors face on their jobs. Next, a large number of viable responses to each of these situations were gathered and ratings of the effectiveness of each of these response options were collected from both experts and novices. These ratings were used to select situations and response options to be included in the SJT. For each item, respondents read through a situation and three to five possible response options and then chose the choices of the most and the least effective response alternative. Several different scoring procedures were devised, but probably the most useful and informative one involved a combination of both the most and least effectiveness response options (i.e., the M-L score). Of all scoring procedures used in this study, the M-L score was both the most reliable (KR-20 = .75) and the most valid. Intercorrelations were computed between the various SJT scores and other criterion measures (Sample sizes ranged from 873 to 909, across nine different military occupational specialties). For example, correlations with performance ratings ranged from a low of .10 ($p < .05$) with a Military Bearing rating composite to a high of .22 ($p < .05$) with a Leading/Supervising rating composite. The SJT also correlated .40 ($p < .05$) with scores on a written technical Job Knowledge test, .22 ($p < .05$) with Promotion Rate and .20 ($p < .05$) with a composite score from three different supervisory role play exercises. The authors conclude that the SJT is a construct valid measure of supervisory job knowledge.

Hodak, G. W. (1992). *A study of tacit knowledge in navy senior enlisted leaders.* Unpublished doctoral dissertation, University of Central Florida, Orlando.

Article type: Test Development and/or Evaluation
Keywords: Cognitive Ability; AFQT; ASVAB; Attitudes; Biodata; Military; Motivation; Performance Ratings; Personality; Tacit Knowledge

The purpose of this study was to identify and analyze the tacit knowledge skills of Navy senior enlisted personnel. Hodak also examined individual differences and tacit knowledge (e.g., sex, billet). In addition, the relationships between tacit knowledge and cognitive ability as measured by the Armed Services Vocational Aptitude Battery (ASVAB) and the Basic Test Battery (BTB) were examined. Tacit knowledge is defined as knowledge that is informal, somewhat disorganized and relatively inaccessible. It is knowledge that generally is not taught in

classrooms or other formal educational settings. Wagner and Sternberg suggest that three kinds of tacit knowledge are applied by successful managers. Tacit knowledge about managing self, managing career and managing others. The tacit knowledge questionnaire from this study consisted of eight different sections: demographics, background, general abilities, motivation, personality traits/personal preferences, strategies/individual traits, personal iterations within the organization and other. Items relating to the three types of tacit knowledge (managing self, career and others) were spread throughout the categories in the questionnaire. Analyses were conducted both by tacit knowledge type and by category. Generally, he concluded that this questionnaire did indeed measure tacit knowledge relevant for Navy senior enlisted leaders, however, there were no significant relationships found between scores on the tacit knowledge questionnaire and grade, billet, community, or sex. There were also very few significant relationships between scores on the tacit knowledge measure and cognitive ability as measured by the ASVAB and the BTB. He also did not demonstrate a relationship between tacit knowledge and success as a senior enlisted leader. Although his sample consisted of a very restricted group of successful Navy enlisted personnel, thus, range restriction probably played a large role in attenuating the results.

Jagmin, N., Wagner, R. K., & Sternberg, R. J. (1989, April). The development of a generalized measure of tacit knowledge for managers. In W. C. Borman (Chair), *Evaluating "Practical IQ": Measurement issues and research applications in personnel selection and performance assessment.* Symposium conducted at the 4th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta.

Article type: Test Development and/or Evaluation
Keywords: Civilian; Cognitive Ability; Performance Ratings; Personality; Satisfaction; Tacit Knowledge

This study describes the development and validation of a measure of tacit knowledge in a sample of managers. The authors defined tacit knowledge as a "fundamental ingredient in the operation of practical intelligence," it is considered practical know-how or common sense strategies that are not usually taught. It's acquisition is considered a necessary component for the operation of intelligence. This paper examined the relationships between tacit knowledge and common measures of both personality (e.g., Myers-Briggs Type Indicator, Fundamental Interpersonal Relations Orientation-Behavior, the Hidden Figures Test, the Kirton Adaptation Innovation Inventory and the Managerial Job Satisfaction Questionnaire) and intelligence (IQ; the Shipley Institute of Living Scale). In addition, they also examined the relationship between tacit knowledge and ratings of performance in a small group simulation (Behavioral Assessment Data [BAD]). Tacit knowledge had the single highest correlation with the BAD ratings ($r = -.61, p < .001$; because of the way the tacit knowledge measure was developed, lower scores are better), compared with a correlation of .38 ($p < .001$) between the IQ measure and the BAD ratings. The tacit knowledge measure also correlated only -.14 (*ns*) with IQ. In addition, the tacit knowledge measure also predicted additional variance in the BAD ratings over and above that predicted by *all* of the other variables included in the study. The authors concluded that tacit knowledge was a viably measurable construct that could be used to predict organizationally-relevant criteria.

Karn, H. W. (1949). Performance on the File-Remmers Test, *How Supervise?* before and after a course in psychology. *Journal of Applied Psychology, 33*, 534-549.

Article type: Validation
Keywords: Supervision

This study examined whether the test *How Supervise?* had universal validity for selecting supervisory personnel. More specifically, the author attempted to determine the impact of a psychology course upon students' understanding of supervisory skills and practices as measured by the File-Remmers test. Two groups of students completed *How Supervise?* at the beginning and at the end of the semester. The first group consisted of 108 students who were enrolled in a required course in psychology (training group), while the second group (N=104) was enrolled in a required course in English (control group). The training group was taught to apply psychological principles to adjustment problems. Real case studies were given to them to solve. It was thought these would be similar to human relations problems that would be found in industry. Results were similar whether both forms were analyzed or just one form. They showed that there were insignificant differences between the two groups on the initial test. The control group scores increased by an insignificant amount on the second test, while the training group scores increased by a significant amount on the second test. Since the psychology class is supposed to teach skills which are helpful in the management of human relations, this study provides evidence of validity for the File-Remmers test as a means of measuring one aspect of supervisory success.

Maloney, P. W. (1952). Reading ease scores for File's *How Supervise? Journal of Applied Psychology, 36*, 225-227.

Article type: Validation
Keywords: Supervision

The author notes that any test used to select supervisors should be easy to read. In this study, the application of Flesch's reading ease formula was applied to the directions and items of *How Supervise?*. The mean readability of both forms, including the directions, is of high school graduate difficulty. Thus, if reading ability is not to be a factor in the selection of supervisors, then they must have a reading ability at least that high. At the time of this writing there was much variability in educational levels of industrial supervisors, over 60 percent did not have four years of high school education. The author concluded that results and interpretations from *How Supervise?* were contaminated by reading ability and recommended re-writing the test using simpler language.

Mandell, M. M. (1950). The administrative judgment test. *Journal of Applied Psychology, 34,* 145-147.

Article type: Validation
Keywords: Cognitive Ability; Civilian; Performance; Performance Ratings; Situational Judgment; Supervision

The Administrative Judgment Test (AJT) is a situational judgment test that was developed by the United States Civil Service Commission for selection into executive or administrative jobs. Questions include problems in relationships between headquarters and field offices, between research and operating personnel, problems in the timing of programs and the organization of the office of an administrator. It is longer than many other situational judgment tests, with between 80 and 100 items. This article describes several concurrent validation studies involving the AJT. Mandell examined relationships between scores on this test, job performance ratings by peers and supervisors and pay grade in four relatively small samples (sample sizes ranging from 20 to 63). Correlations with the ratings ranged from .49 to .68 for the AJT (median = .51) and from .13 to .64 for mental ability (median = .41). Correlations with grade ranged from .28 to .56 for the AJT with a median of .52 and from .21 to .38 for mental ability with a median of .26. The AJT had moderately high correlations with scores on the mental ability tests (in the 50s and 60s), but the mental ability tests had lower correlations with the criterion measures. The split half reliability of the AJT was .94.

Mandell, M. M. (1956). The validity information exchange. *Personnel Psychology, 9,* 105.

Article type: Validation
Keywords: Civilian; Performance; Performance Ratings; Situational Judgment; Supervision

Mandell reports validities from four additional studies (sample sizes ranged from 51 to 196) using a 70-item version of the Administrative Judgment Test (AJT; Mandell, 1950). Two of the studies examined AJT correlations with salary and reported obtained validities of .48 and .54. Another study reports a biserial correlation of .32 between AJT and number of promotions received in a four-year period. The final study reports a tetrachoric correlation of .62 ($p < .01$) between AJT and job performance ratings made by supervisors.

Millard, K. A. (1952). Is *How Supervise?* an intelligence test? *Journal of Applied Psychology, 36,* 221-224.

Article type: Validation
Keywords: Cognitive Ability; Civilian; Situational Judgment; Supervision

Millard found a nonsignificant correlation of .22 between *How Supervise?* and intelligence test (the Adaptability Test) for higher level supervisors (office) but very high correlations for lower

level or first line supervisors: the correlation was .71 ($p < .01$) for factory supervisors and .69 ($p < .01$) for those supervising newspaper carriers and dealers. These correlations are not significantly different from the alternate form reliability for the test.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedures: The low fidelity simulation. *Journal of Applied Psychology, 75*, 640-647.

Article type: Test Development and/or Evaluation
Keywords: Cognitive Ability; Civilian; Performance; Situational Judgment; Training Performance

This article describes the development and concurrent validation of a situational judgment test (referred to as a "low-fidelity simulation") for selecting entry-level managers in the telecommunications industry. Job analysis results identified three managerial dimensions that were common to a variety of managerial jobs -- communication, problem-solving and interpersonal skills -- and the test development focused on the latter two areas. Researchers collected 1,200 critical incidents from incumbents and their supervisors and used these to generate 64 task situations. Then, 150 incumbent mangers were asked to write in a few sentences how they would handle each problem situation and these were used to prepare 5 to 7 general strategies for handling each situation. The effectiveness of each of these strategies (i.e., response options) was evaluated by about 35 very experienced senior managers. These ratings were used to drop situations with a lack of agreement or a lack of differentiation among response alternative and also to identify one "best" and one "worst" response for each situation. When taking the resulting test, respondents were asked to indicate what they would most likely do and what they would least likely do in each situation. These researchers conducted validity research for this test using several samples of managers. Correlations between scores on the new test and job performance ratings made by respondents' managers using behaviorally anchored rating scales were generally in the low to mid 30s. Correlations between test scores and tenure were not significantly different from zero. In an applicant sample the correlation between test scores and GPA was .30. Many of the correlations with assessment center scores and interview dimensions that targeted the same managerial skills as the experimental test were also significantly different from zero. Race and sex differences in test scores were small and generally did not reach statistical significance. They offer three possible explanations for this test's validity: (1) behavioral consistency (i.e., applicants recalled past behavior when answering); (2) intentions and goal setting (i.e., applicants are more likely to do what they say they will do); (3) or the test measures "practical intelligence" which in turn is related to job performance.

Mowry, H. W. (1957). A measure of supervisory quality. *Journal of Applied Psychology, 41,* 405-408.

Article type: Test Development and/or Evaluation
Keywords: Cognitive Ability; Performance; Personality; Situational Judgment; Supervision;

This article describes research to develop and evaluate a situational test of supervisory "quality." Mowry distinguishes between supervisory knowledge and supervisory insight and suggests that insight is expected to be related to ability to supervise while knowledge may or may not be. He argues that a test of supervisory *insight* should present "problems similar to the way they would occur in a real-life situation" (p. 405), including some structure, some latitude and attractive alternatives, each of which would be acceptable under certain conditions. One-hundred and fifty situations were selected from case study training materials. These were administered to 200 supervisors in several companies. A key was apparently developed based on the author's own judgment and 50 items were selected based in internal consistency, item-total correlations and item difficulty. He dichotomized some very small groups based on job performance ratings and found significant differences in scores; relationships were largest for ratings of "ability to handle people." In several additional small samples Mowry obtained the following correlations between his "Supervisor's Problems" test: Otis (.33), mental maturity (.39) and F-scale (-.49). The latter correlation indicates that more democratic (less authoritarian) supervisors obtained significantly higher scores. The split half reliability was .81. Later versions of this test (e.g., Tenopyr, 1969) are referred to as the Leadership Evaluation and Development Scale (LEADS).

Myers, C., & Davids, K. (1993). Tacit skill and performance at work. *Applied Psychology: An International Review, 42,* 117-137.

Article type: Review
Keywords: Tacit Knowledge

For years, psychologists have neglected the impact of tacit knowledge on performance. This paper reviews some of the recent applied and experimental work on how tacit knowledge effects performance in work contexts. Leplat's (1990) definition states that "tacit knowledge is exemplified by skilled activity which is *procedurally informal* and/or *inarticulable*" (p. 123). The first part is most frequently examined via field studies. It is comprised mostly of information that is gained through experience as one works on the job. It has a heuristic quality and is untaught. Laboratory studies have more frequently examined the inarticulable aspect of tacit knowledge. Practical experience must be used to gain knowledge, but it is not easily explained to others. Cognitive psychologists have provided much of the empirical work focusing on expert systems development. Implicit knowledge may develop under circumstances in which there are many variables in a complex task. On the other hand, explicit knowledge may develop when key variables are made more salient. Others have looked at how well such knowledge can be transferred to novel situations. A few examples are given of studies showing how tacit knowledge is developed in working environments and how it contributed to performance. Management is

beginning to recognize the benefits of the knowledge of the worker and is increasing the autonomy and responsibility of the front-line worker.

National Computer Systems, Inc. & Richardson, Bellows, Henry & Co. (1981-1983). *Technical Reports Supervisory Profile Record.* Authors.

Article type: Test Development and/or Evaluation
Keywords: Civilian; Performance; Performance Ratings; Situational Judgment; Supervision

This series of three technical reports describes an extensive line of research to develop and validate a test for identifying those with highest potential for success as first-line supervisors. This test, the Supervisory Profile Record (SPR) included both situational judgment and biodata type items and all results are presented for the test as a whole. The developmental and validation samples included supervisors from a wide variety of organizations and locations. This provided a great deal of statistical power and information about validity generalization. All of these samples were predominantly white males. Job analysis results confirmed that there is substantial overlap in the duties of first-line supervisors and especially in the abilities required across organizations, functions and locations. SPR items came from existing item pools and supervisory practices articles, workbooks and texts. They were intended to obtain respondents "views" in areas such as employee motivation and communication, personnel training and development, people and production problem resolution, discipline application and general supervisory style and practice. They also took steps to minimize reading difficulty and to ensure that the ability to respond to the judgment items would not be dependent on having had supervisory experience. A total of 99 judgment and 128 biodata items were developed. For the judgment items, respondents were asked to select both the best response and the second best response. Eighty judgment and seventy biodata items were included in later research; and the final version of the test contained forty one biodata and fifty nine judgment items. The reading levels of the various sections ranged from grade 10 to grade 14. Scoring involved developing and item-level empirical key by comparing the responses of successful supervisors with those who were not. They report validity results for many samples and each sample studied contained over 2000 supervisors, resulting in a total of over 30,000 supervisors. Concurrent validities across a large number of samples ranged from .22 to .40 for duty ratings and from .33 to .48 for the ability ratings. Validities for duty performance adjusted for unreliability in the ratings range from .36 to .44. Test-retest reliabilities for two samples of about 100 each were about .85.

Phillips, J. F. (1992). Predicting sales skills. *Journal of Business and Psychology, 7,* 151-160.

Article type: Test Development and/or Evaluation
Keywords: Civilian; Performance; Performance Ratings; Situational Judgment

This article describes research to develop and evaluate a situational judgment test to predict sales skills for a telephone sales job. Across all job functions, selling was rated the highest in impor-

tance and second highest in time spent. Job analysis results indicated that selling includes five main elements: establishing rapport, fact finding, sales presentation, overcoming objections and closing the sale. "Trained industrial psychologists" wrote 64 situational judgment test items and response alternatives to cover these five areas evenly. Then, 20 job incumbents and supervisors were asked to determine the correct response to each item. A total of 51 items were included in the final test. A concurrent validation study was conducted with over 200 job incumbents. This sample was restricted in range of cognitive ability because they had been selected, in part, based on their scores on a written cognitive ability test. The criterion was job performance ratings made by supervisors using behavior-base scales that had been developed based on the job analysis. These items were used to construct two rating composites: "job duties" and "selling." A composite of these two variables was labeled "job overall." Correlations between test scores and these three rating composites were .24, .18 and .16 respectively. Women scored significantly higher than men.

Rosen, N. A. (1961). *How Supervise?* -- 1943-1960. *Personnel Psychology, 14*, 87-99.

Article type: Review
Keywords: Cognitive Ability; Civilian; Performance; Performance Ratings; Situational Judgment; Supervision; Training

This article reviews research on a situational judgment test called "*How Supervise?*" This test was "designed to measure knowledge of human relations principles and practices in supervising employees and attitudes relating to such principles and practices" (p. 87). When this article was published over 750,000 copies of the test had been administered. Rosen reviews about half a dozen criterion-related validity studies for *How Supervise?* and concludes that the results are mixed but the published literature reveals more positive than negative evidence of validity for predicting supervisory success. One of these studies (McCormick & Middaugh, 1956) did not find significant validity for the original key, but they developed a new key empirically using job performance ratings and found that this empirical key cross validated well for supervisors at the same level as those used in key development and somewhat less well for supervisors at different organizational levels. This suggests that tailor made keys for *How Supervise?* may need to be developed in each organization. *How Supervise?* has been used as part of *many* training programs and the positive reception by most recipients is seen by these authors as indicating high face validity. Several studies have shown that How Supervise scores improve after managerial or supervisory training. Several researchers have also found that *How Supervise?* scores improve with training in psychology. Some research has shown that this test is correlated with tests of cognitive ability and with college grade point average. Research on the relationship between *How Supervise?* scores and educational level is mixed.

Tenopyr, M. L. (1969). The comparative validity of selected leadership scales relative to success in production management. *Personnel Psychology, 22*, 77-85.

Article type: Validation
Keywords: Cognitive Ability; Civilian; Performance; Job Performance; Situational Judgment; Supervision

Tenopyr (1969) conducted a concurrent study to compare the validity of a situational judgment test called the Leadership Evaluation and Development Scale (LEADS; Mowry, 1964), a measure of supervisory attitudes called the Leadership Opinion Questionnaire (LOQ) and a vocabulary test. The version of LEADS she used was formed by shortening Mowry's 50 item Supervisor's Problems test to 44 items based on item-total and item-criterion correlations. All of these measures were administered to a sample of 113 male production managers. The criteria were several measures of managerial success. Ratings of these managers' performance were collected from their supervisors using a scaled checklist. Labor relations staff also rated these supervisors' handling of employee relations matters. These managers' LEADS scores correlated .36 with salary corrected for age and length of service and .25 with performance ratings (handling of employee relations matters) by labor relations staff. LEADS also had a moderately high correlation with a test of verbal comprehension (.49), but the verbal comprehension test had lower correlations with corrected salary (.29) and the ratings by labor relations staff (.08). LEADS correlated .37 with the LOQ consideration score and , -.36 with LOQ structure score. The two LOQ scores did *not* correlate significantly with corrected salary and only the structure score was significantly (negatively) correlated with the ratings made by labor relations staff. None of the predictors correlated significantly with the ratings made by supervisors, age or seniority. Tenopyr concludes that although LEADS is structured like an ability test, it is only moderately correlated with verbal comprehension and is to some extent related to attitudinal variables as well.

Wagner, R. K. (1987). Tacit knowledge in everyday intelligent behavior. *Journal of Personality and Social Psychology, 52*, 1236-1247.

Article type: Test Development and/or Evaluation
Keywords: Cognitive Ability; Civilian; Situational Judgment, Supervision; Tacit Knowledge; Training

This article further investigates the scope and structure of tacit knowledge (Wagner & Sternberg, 1985). Tacit knowledge is defined as knowledge that is usually not openly expressed, stated or taught and the measures of tacit knowledge he developed are situational judgment type tests. In addition to varying in content (related to managing self, others, or career) Wagner proposed that situations in which tacit knowledge may be applied also vary in terms of context (local versus global) and orientation (idealistic versus pragmatic). He developed 12 work-related situations for academic psychologists that sampled all of the above dimension levels in a perfectly crossed fashion. These were administered to over 200 subjects with differing levels of training and experience in the field of academic psychology. In addition to asking respondents to rate the importance of each possible response (from extremely unimportant to extremely important), he

also asked them to rate the quality of each from extremely bad to extremely good. They were also asked how good and important each of these "should be" in one's ideal academic world. Results showed that faculty had significantly greater congruence between their actual and ideal responses than did students. Tacit knowledge was computed as the sum of the squared deviations from a "prototype" rating derived using 11 "experts" (faculty members from the highest quality schools). Results showed almost identical patterns of correlations with the "criterion reference measures" as did the earlier studies (Wagner & Sternberg, 1985). Correlations for the ideal ratings were similar to those for the actual ratings, but generally smaller. Results of exploratory and confirmatory factor analyses suggest that tacit knowledge can be explained by a single general factor and many specific factors (i.e., there was only one factor). Wagner also found that undergraduates' tacit knowledge scores for psychology and business correlated in the 50s and they take this as support for the generality of tacit knowledge. Unlike the earlier study, verbal ability was significantly correlated with the measure of psychology tacit knowledge, but not business.

Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology, 49*, 436-458.

Article type: Test Development and/or Evaluation
Keywords: Cognitive Ability; Civilian; Situational Judgment, Supervision; Tacit Knowledge; Training

This article describes the development and validation of several situational judgment type tests designed to measure "tacit knowledge." Tacit knowledge is defined as knowledge that is usually not openly expressed, stated or taught. They divide tacit knowledge into three categories: that related to managing (1) self; (2) others; and (3) career. Wagner and Sternberg developed two tacit knowledge measures: one for business managers and one for academic professors. Highly effective individuals from each of these fields were interviewed and to describe *typical* work related situations and the responses to them. They then "identified" alternative responses and identified key responses statistically through the use of item discrimination, those responses that differentiated between those with more and less experience and formal training in the career pursuit of interest. Each "work-related" situation had between 6 and 20 response items. They obtained ratings of the importance of each response item from both novices and experts and retained those items that correlated significantly with experience. For professors, total tacit knowledge score was correlated with publications, conferences attended, level of school and percentage of time spent in research, but *not* with academic rank or year of Ph.D. "It is what we learn from experience, rather than experience, per se, that seems to matter" (p. 445). For graduate students, tacit knowledge correlated with years completed, level of school, research projects and publications, but not with teaching experience. They used a sample of only 20 students to compute the correlation between tacit knowledge and the verbal reasoning subtest of the Differential Aptitude Tests did not obtain a significant correlation. There appeared to be a serious ceiling effect for verbal reasoning (with a mean of over 46 out of 50 possible points). For the manager group, scores on this tacit knowledge scale were correlated with level of company, years of schooling beyond high school and salary, but not with years of management experience, level

of title, or number of employees supervised. Of the subscales, only that related managing career had any significant correlations. Again the correlation with verbal ability was not significant, but again there was a large ceiling effect. They use a sample of only 22 bank managers to cross-validated the test developed for managers. Significant correlations were obtained between total tacit knowledge and average percent of salary increase and ratings of generating new business and implementing bank policy.

Weitz, J. & Nuckols, R. (1953). A validation study of "*How Supervise?*". *Journal of Applied Psychology, 37,* 7-8.

Article type: Validation
Keywords: Cognitive Ability; Civilian; Performance; Situational Judgment; Supervision

Weitz and Nuckols (1953) administered a revised version of *How Supervise?* containing 100 items to about 80 district managers in a life insurance company. Criteria were sales volume, agent turnover and a measure of business quality. The only correlations that even bordered on significant were with 4-year turnover and these correlations were just barely significant at the $p < .05$ level. The correlation between *How Supervise?* scores and educational level was .31 ($p < .01$). They conclude that *How Supervise?* is not valid for these managers.

Wickert, F. R. (1952). Relation between How Supervise? intelligence and education for a group of supervisory candidates in industry. Journal of Applied Psychology, XXXVI, 301-303.

Article type: Validation
Keywords: Cognitive Ability; Civilian; Situational Judgment; Supervision

Wickert conducted research concerning the extent to which *How Supervise?* measures intelligence. This involved correlating verbal intelligence, *How Supervise?* and years of education for a sample of about 100 candidates for supervisory positions. These correlations ranged from .54 to .66. When the group was dichotomized by educational level, the correlation between *How Supervise?* and verbal intelligence was .65 for those with 11th grade or less and .20 for those with 12th grade or more. This is in agreement with the results of previous research by Millard (1952) and they interpret it to mean that *How Supervise?* is of doubtful validity as a measure of supervisory ability for lower level personnel.

IV. Crew Resource Management Reviews

Executive Review

Another relatively recent line of research that points to the importance of interpersonal skills and personality dimensions in effective pilot performance is research on crew (or cockpit) resource management (CRM). This research was motivated by analyses of the causes of aircraft accidents which showed that the majority of these accidents involved crew errors (Foushee, 1984). Further analysis indicated that aspects of the crews' interpersonal interactions, such as breakdowns in coordination and communication, most frequently played a causal role in these accidents (e.g., Cooper, White, & Lauber 1980). Research in the CRM area provides evidence of relationships between measures of crew communication and crew performance (e.g., Foushee & Manos, 1981; Orasanu, 1990) and highlights the importance of variables such as information transfer, effective decision-making and planning in effective crew management.

The concept of crew resource management (CRM) has become widely accepted and is now being incorporated into the flight training programs of most airlines (Helmreich, Wiener, & Kanki, 1993). Development of CRM programs began primarily in the early 1980s, and initially focused on changing pilots' attitudes toward crew involvement in the cockpit. Recently, researchers and practitioners have begun to realize that CRM really includes more people than just those on the flight deck. In fact, the acronym CRM has evolved to stand for *crew* resource management as opposed to just *cockpit* resource management. The training was first offered only to flight deck crewmembers, but programs have since been expanded to include a host of others who are all responsible for providing safe air travel. Some airlines have joint CRM programs in place that include both flight deck crewmembers and flight attendants (e.g., America West Airlines, United Airlines). It has also been suggested that CRM training be expanded to include persons from other relevant domains (e.g., Air Traffic Controllers). The modern day CRM training courses typically occur over several days and provide experiential learning in interpersonal communication, decision-making, conflict resolution, stress management and leadership (Helmreich & Wilhelm, 1991). Although belief in the effectiveness of CRM training is widespread throughout the industry, there is little empirical evidence available to support the position that CRM training has long-term behavioral effects on crewmembers; most of the evidence supporting the positive impact of CRM training has focused on its effects on trainees' attitudes toward crew management. Only recently have researchers begun to examine whether CRM training is effective with respect to changing behaviors, both in the simulator and during actual flights. Some researchers have also found that a small subset of CRM trainees actually react *negatively* to the training (i.e., a "boomerang effect") and have now begun to explore the reasons behind why this occurs.

While CRM researchers typically take a training approach to improving crew performance, their research points to the importance of the sorts of interpersonal skills and personality dimensions that can be measured using a situational judgment test for effective performance as a member of a crew (e.g., as an aircraft commander or Captain of an aircraft). In fact, some research in this area has actually demonstrated relationships between measures of relevant personality traits, attitudes and performance, both flight simulator and actual performance (e.g., Helmreich,

Foushee, Benson, & Russini, 1986; Chidester, Kanki, Foushee, Dickinson, & Bowles, 1990). The current aviation environment, both military and civilian, reflects a growing emphasis on CRM and a growing concern with the interpersonal skills necessary for effective aircrew management.

Annotated References

Adams, R. J., & Payne, B. (1992). Administrative risk management for helicopter operators. *The International Journal of Aviation Psychology, 2(1),* 39-52.

Article type: Training Development and/or Evaluation
Keywords: Accidents; Personality; Training; Communication; Coordination; Decision-Making

This article describes a proposed risk management program that could be used to train personnel who work with air helicopter medical operations. Past personality research has shown a tendency for helicopter pilots involved in accidents to be mature, experienced professionals who make poor decisions from a safety perspective. These pilots tend to be low in conformity and productivity and in expressed need to control others. This program targets all individuals who are involved in the operation of air helicopter emergency response teams (e.g., pilots, medical personnel, hospital administrators). This risk assessment program includes aspects of mission planning over which management has direct control. It also includes other elements such as aircraft selection and maintenance, training, crew scheduling and getting hospital administrators to recognize the risks involved in helicopter medical missions. The chief pilot or flight manager is viewed as mostly responsible for publishing an operating philosophy, in addition to carrying the ultimate responsibility for "making the program work." Finally, these individuals are also responsible for monitoring the success and failures of the program and adjusting accordingly. The program director's responsibilities on the medical side are analogous to those of the chief pilot on the flying side. The article also discusses specific training requirements that are characteristic of successful risk management programs. Roles of pilots, non-flying personnel, medical personnel and management are discussed in detail.

Alkov, R. A (1991). U.S. Navy aircrew coordination training - a progress report. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 368-371).

Article type: Training Development and/or Evaluation
Keywords: CRM; Military; Training

The Navy has developed an Aircrew Coordination Training (ACT) program and, according to this author, it improves upon commercial crew resource management (CRM) programs because it includes segments on pilot judgment, aircrew decision-making, loss of situation awareness, policies and regulations, command authority, workload performance, use of available resources, operating strategies, and communication skills. The training was initially introduced in three

different Naval aircraft communities: helicopters, attack bombers and multi-placed fighters. Aircrew mishap rates were examined for each of these communities after training. Initially, the helicopter crews had an increase in their mishap rate, but this quickly declined. The reason cited for the initial increase was that this aircraft community was spread all over the country and this made it difficult to train the pilots. The attack bomber community, on the other hand, was centrally located and the training was easily administered. This group quickly demonstrated decreases in error rates. The fighter community did not accept the training initially and instead conducted their own seminar. Their error rate subsequently increased and at this point they were forced to institute the ACT program. Soon after the institution of the ACT program, their mishap rate also decreased.

Alkov, R. A., Borowsky, M. S., Williamson, D. W., & Yacavone, D. W. (1992, August). The effect of trans-cockpit authority gradient on Navy/Marine helicopter mishaps. *Aviation, Space and Environmental Medicine, 63*(8), 659-661.

Article type: Group Comparison Study
Keywords: Accidents; Communication; Leadership

This study examined the hypothesis that "there is an optimum gradient to allow an effective interface between aviators." Post-hoc aircraft mishap studies have suggested that some communication failures (e.g., desire to avoid conflict, deference to authority, arrogant/domineering style of Captain) may result from distances between the rank of the pilot and the first officer. The authors analyzed aircrew error mishaps that occurred between 1980 and 1990. They hypothesized that pilots of equal rank and those with the greatest disparity in rank would have higher mishap rates than other rank pairings. Mishaps were categorized into three pilot/co-pilot pairings (same rank pilot/co-pilot; one rank difference; two rank difference). They did not find any statistically significant differences, however, equal ranked pairs had the fewest mishaps (contrary to expectations), followed by one rank difference and then two rank differences. Authors suggest that the Naval Safety Center's Aircrew Coordination Training Program addressed such issues.

Baker, D. P, Bauman, M., & Zalesny, M. D. (1991). Development of aircrew coordination exercises to facilitate training transfer. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 314-319).

Article type: Training Development and/or Evaluation
Keywords: Attitudes; CRM; Military; Training

The Navy was in the process of developing an Aircrew Coordination Training (ACT) program for the Marine Corps helicopter crews at the time during this article was written. This study examined the reactions of trainees to two exercises: (1) the pre-flight brief; and (2) assertiveness exercises. The trainees were 41 CH-46 pilots who attended a two-day ACT program. Results for the pre-flight brief showed that, overall, pilots thought the exercise was worthwhile and would

have an impact on how they reacted to briefings in the future. Rankings of the value of the assertiveness role play placed it in the top two out of four exercises.

Beach, B. E. (1979). Line-oriented flight training. In G. E. Cooper, M. D. White, & J. L. Lauber (Eds.), *Resource management on the flight deck, Proceedings of a NASA/Industry Workshop* (pp. 107-118). San Francisco, CA: National Aeronautics and Space Administration.

Article type: Training Development and/or Evaluation
Keywords: LOFT; Training; Training Program Design; Simulation

The author describes the line-oriented flight training (LOFT) program that was in place at Eastern Airlines in 1979. The presenter notes that Eastern had a similar program in place in the late 1950's, but the simulators used in that program did not have full visual or motion capabilities. The program at this time consisted of six scenarios per each type of aircraft. Instructors were generally not allowed to deviate from the simulator script and the simulation continued until the aircraft "reaches the ground" (one way or another). Simulations are essentially "full flight" in that they start at a crew briefing and end in the post-flight debrief. Eastern has used LOFT for remedial training (e.g., crews have been taken off of the line if they demonstrate a problem of one kind or another). The LOFT format is then used to duplicate the problem that the crew had; they are allowed to work through the scenario and then receive instruction on a better way of responding. The article notes that LOFT does not provide training while the flight is in progress, but only before the flight or after the flight in debriefing.

Berlin, J. I., Gruber, E. V., Holmes, C. W., Jensen, P. K., Lau, J. R., Mills, J. W., & O'Kane, J. M. (1982). *Pilot judgment training and evaluation* (DOT/FAA/CT-82/56-1). Daytona Beach, FL: Embry-Riddle Aeronautical University. (NTIS No. AD-A117 508).

Article type: Validation
Key words: Decision-Making; Training; Training Evaluation

There has been little research on the nature of pilot judgment and whether or not it can be improved through training. This study had three purposes, to: (1) develop optimized pilot judgment training and evaluation course materials; (2) demonstrate the effectiveness of the program in a small-scale experiment; and (3) develop and test, on a small scale, guidelines and procedures for the objective measurement of pilot judgment in both ground and flight training environments. Complete instructions for implementing the program are contained in an Instructor's Manual. The two theories underlying the implementation process are behavior modification and facilitation. There were three groups in this study. The experimental group was given an academic pretest, ground and flight judgment training, an academic posttest and an observation flight at the completion of the training program (26 subjects). Two control groups were included. One was labeled the flight control group because these people only received the observation flight at the end of the study (24 subjects). The second control group was the academic control group, which received the written pretest and posttest (25 subjects). The pretest and posttest consisted of the

following five sections: judgment concepts - nonjargon, scenario analysis, scenario analysis - factor importance and certainty rating, action ways, judgment concepts - specific terminology. During the observation flight, subjects were placed in twenty judgment situations and their behavior was rated. Differences between the experimental group and the academic control group were significant for four sections of the pretest and posttest. Performance on the observation flight also showed significant differences between the experimental group and the flight control group.

Birnbach, R. A., & Longridge, T. M. (1993). The regulatory perspective. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 263-281). San Diego, CA: Academic Press, Inc.

Article type: Review
Keywords: CRM

Commercial aviation began to develop after the passing of the Air Mail Act of 1925. During the following year, there were many well-publicized fatal accidents and these caused public uproar. Consequently, Congress passed the Air Commerce Act of 1926; this gave responsibility and regulatory authority for the aviation industry to the Aeronautics Branch of the Department of Commerce. Today, the Federal Aviation Administration (FAA) has the dual task of both regulating safety in aviation and also promoting the growth of the industry.

Parts 121 and 135 of the Federal Aviation Regulations (FAR) have had great impact on commercial and general aviation. Part 121 applies to the large air carriers, while Part 135 applies to the commuter airlines. The requirements under Part 135 tend to be less stringent, so these pilots fly with a smaller margin of error. All airlines must have their training programs approved by an FAA Principal Operations Inspector (POI). There are some differences in training programs, for example, the number of required hours and the content of training programs differ between Parts 121 and 135. The amount of initial operating experience (IOE) required also varies for pilots and co-pilots if covered under Part 135. The extent to which simulators are incorporated as a part of training also varies between the two parts. Part 121 operators have relied on simulators much more than the smaller carriers. These differences must be kept in mind when designing crew resource management (CRM) training programs, and standardized measurement conditions need to be implemented for Part 121 and Part 135 operators. Also, the measurement process needs to be calibrated, such that all safety inspectors and company employed check airmen are trained on a recurrent basis. In order to determine the reliability and validity of CRM programs, research has begun as part of the Advanced Qualification Program (AQP). As technology changes, so does the role of the pilot in the cockpit. Instead of being the controller, the pilot is now much more of a systems manager. The demands of the modern cockpit fall much more under the domain of CRM than in the traditional cockpit. Unfortunately the carriers most in need of CRM training are the ones least able to afford it, the Part 135 carriers. The FAA is trying to make more opportunities available for this section of the industry, and an AQP program may be developed in the future specifically for them. This program gives the

industry more flexibility in the equipment which is used to provide training. The establishment of several databases to track the integration of CRM is also being conducted through AQP.

Bolman, L. (1979). Aviation accidents and the "theory of the situation." In G. E. Cooper, M. D. White, & J. L. Lauber (Eds.), *Resource management on the flight deck, Proceedings of a NASA/Industry Workshop* (pp. 31-58). San Francisco, CA: National Aeronautics and Space Administration.

Article type: Theoretical
Keywords: Decision-Making; Situational Judgment; Teamwork; Training Program Design

This presentation suggests that flight crews are never fully aware of the situation they are in during a flight; rather, they develop "theories of the situation" (TOS). Each TOS is a set of beliefs, goals and behaviors that provide a coherent picture of what is occurring around them and what action is appropriate. Under routine situations, these theories are very much in sync with the real world, however, in situations that are more complex and difficult the chances for error in these TOSs increases. This paper addresses how these TOSs develop, which background factors influence the TOS that any given pilot is likely to use, the extent to which education/training can reduce erroneous TOSs and the situational factors that govern whether a pilot is likely to notice errors in his or her TOS. It also identifies several factors that influence whether a faulty TOS will be detected: (1) the degree to which the TOS that the pilot uses has been built and tested in situations of confusion, crisis, etc.; (2) the abilities of the crewmembers to combine skills (i.e., teamwork); (3) the management style of the Captain; and (4) the degree to which the role structure in the cockpit is understood and mutually shared. The author recommends that pilots receive training in how to recognize their idiosyncratic patterns for learning, relating to others and for managing the flight deck, but without de-emphasizing technical training.

Borowsky, M. S. (1981, May). Dual accident pilots in Naval aircraft. *Aviation, Space and Environmental Medicine*, *52*(5), 310-311.

Article type: Group Comparison Study
Keywords: Accidents; Safety

Data continues to show that over 50 percent of personnel and aircraft loss in naval aviation is due to pilot error. The author compared the "accident-proneness" of pilots who had an accident in their first 1000 flying hours and pilots who had no accidents in their first 1000 hours. Using a Chi-Square test, the author then looked at the subsequent 10,000 hours of flying time for two groups. He found no significant differences between the two groups, but there was a trend toward more accidents in the future for those pilots who had accidents during their first 1000 hours of flying.

Bortolussi, M. R., & Vidulich, M. A. (1991). An evaluation of strategic behaviors in a high fidelity simulated flight task: Comparing primary performance to a figure of merit. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 1101-1106). Columbus, OH: Ohio State University.

Article type: Single Group Study
Keywords: Civilian; Performance; Simulation

This study utilized a standardized approach to develop Figures of Merit (FOMs; also known as measures of effectiveness or measures of merit). The FOMs are essentially composites of differentially weighted primary performance measures. The authors evaluated whether the FOMs performed better as a performance criteria as compared to unweighted primary performance measures in two full mission scenarios. Six pilots (average flight time was 1825 hours) participated in the full mission simulations. The FOM was developed on the basis of a detailed task analysis, the total FOM consisted of nine differentially weighted primary performance measures. The nine measures were Rudder SD (smoothness), Elevator SD, Aileron SD, Altitude SD, Mean Altitude, Heading SD and Mean Heading. In addition three subFOMs were developed for altitude: Altitude FOM, Heading and Airspeed. Two full mission instrument flight rule scenarios were developed for the study, one was considerably more difficult than the other. A counterbalanced within-subjects design was used. Each pilot's data were included in a 2 X 10 repeated measures analysis of variance (ANOVA). Only four of the nine primary performance measures (Aileron SD, Elevator SD, Altitude SD and Mean Airspeed) showed a main effect due to scenario. There were no significant main effects for scenario for the Total (weighted) FOM, indicating that the FOM was not sensitive to the differences in difficulty between the two scenarios. However, all three subFOMs demonstrated a main effect for scenario type.

Bowers, C. A., & Salas, E. (1991). The assessment of coordination demand for helicopter flight requirements. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 308-313).

Article type: Training Development and/or Evaluation
Keywords: Coordination; CRM; Military; Training; Training Evaluation; Training Program Design

Crew resource management (CRM) programs are being developed to address the coordination needs of the military. The authors list several areas that are in need of research. These programs teach coordination skills, but it is unclear how these skills should be applied to tasks. Also, the relationship between coordination skills and workload is not known, for example, the effectiveness of coordination behavior may depend on the task situation.

Brecke, F. H. (1982). Instructional design for aircrew judgment training. *Aviation Space and Environmental Medicine, 53*, 951-957.

Article type: Theoretical
Keywords: Decision-Making; Situational Judgment; Training

Brecke briefly reviews previous research on judgment training, focusing on Jensen and Benel's (1977) work for the FAA. These authors had suggested that judgment consists of both a cognitive and an affective component; they also believe judgment can be trained if an instructional systems design (ISD) is used (including situational training techniques); and judgment can be evaluated if evaluations are based on criteria established by behavioral objectives and if the training situation is carefully structured. Brecke presents a conceptual framework for thinking about the problem of judgment and a discussion of the applications of this conceptual framework to training design and research. The author recommends the use of "situational tests" for both selection and training purposes. For selection, the author suggests designing carefully graduated situations varying in cognitive complexity, uncertainty, time availability and stress. Current military aircrew training (circa 1982) is characterized by an emphasis on correct completion of prescribed procedures, compliance with rules and specific flight techniques. For training, the author suggests training in judgment, with training under conditions of uncertainty, especially if such training follows an orderly step-by-step course, systematically developing a trainee's self-reliance and initiative. Finally, he recommends the systematic design of scenarios where variables of uncertainty, cognitive complexity and time availability are manipulated.

Brown, D. D. (1987). Introduction to Military Airlift Command (MAC) CRM training. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 132-134). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: CRM; Military; Training; Training Program Design

According to the author, the Air Force has been implementing training programs similar to crew resource management (CRM) programs for many years. He described training he received in order to obtain co-pilot status and for instructor upgrade training. He also described a class taught in the early sixties called "pilot judgment" training, in which previous accidents were reviewed and discussed. He interprets the recent (in 1987) Air Force co-sponsoring of a workshop devoted solely to CRM as due to the increasing recognition that while the Air Force was forward-looking in their philosophies regarding training, they lagged in their use of educational technology.

Brown, C. E., Boff, K. R., & Swierenga, S. J. (1991). Cockpit resource management: A social psychological perspective. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 398-403).

Article type: Theoretical
Keywords: Communication; Decision-Making

The authors review previous research on communication and decision-making. Research conducted in the area of social cognition may be helpful for crew resource management (CRM). For example, the fundamental attribution error states that people tend to attribute others' actions to dispositional factors, rather than situational factors. Whereas, the reverse is true when considering explanations underlying their own actions. Other biases (e.g., the ultimate attribution error; self-fulfilling prophecies) may subsequently perpetuate and affirm the misattribution. Glass cockpits have changed the role of pilots in the cockpit. Communication is potentially reduced when the automated systems decrease the amount of information crewmembers can determine about one another's actions. Social psychology is sensitive to these types of issues and has much to contribute for understanding them and providing useful solutions.

Bruce, K. D., & Jensen, D. (1987). Cockpit resource management training at People Express: An overview and summary. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 50-51). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: Civilian; CRM; Training; Training Program Design

All curriculum development at People Express, as well as simulator and flight instruction is conducted by current and qualified line-flying crewmembers. The academic portion of this program consists of twelve study units. In each case, simple effective methods to improve performance are provided. The program consists of three basic parts. First, a self-study academic course is provided that consists of a printed text (workbook format), used interactively with two audio cassette tapes. Second, workshop seminars are conducted in which the pilots meet for one day and discuss text material, watch videotapes and discuss and analyze an NTSB accident report. Third, a line-oriented flight training (LOFT; full-mission simulation) session is completed. The scenarios are designed to maximize the crews' opportunities to apply and practice crew resource management (CRM) methods and behaviors.

Butler, R. (1987). Pan American World Airways flight training -- A new direction flight operations resource management. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 61-67). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: CRM; LOFT; Training; Training Program Design

According to the author, Pan Am decided to buy a complete crew resource management (CRM) program rather than develop one of their own. At Pan Am, each cockpit crewmember must participate in a three day seminar (this is a one-time requirement) during which they must introspect on their own personal management style. In the first year of annual training, each cockpit crewmember participates in one day of ground school, followed by a LOFT exercise which also contains a self-critique (intended to measure the effectiveness of the training). When an annual recurrent flight training program is in place, it will consist of two days of ground school, one day of LOFT and one day for a proficiency check.

Butler, R. E. (1991). Lessons from cross-fleet/cross-airline observations: Evaluating the impact of CRM/LOFT training. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 326-331).

Article type: Training Development and/or Evaluation
Keywords: CRM; LOFT; Training; Training Evaluation

The author analyzes the responses of several different airlines to measures that have been developed to evaluate the effectiveness of various crew resource management (CRM) programs. One purpose was to discover if there was any standardization of programs across fleets and airlines. Measures that were used included the line-oriented flight training (LOFT) survey and the Line/LOS Checklist. Descriptions of these measures are given by the author and can also be found in other references. Most crews' performance ratings were average, with very few receiving very high or very low ratings. One disturbing finding was significant differences between the different airlines on ratings of overall technical proficiency and overall crew effectiveness. There were also differences with respect to the content of the LOFT scenarios from one airline to another. LOFT instructors still tended to focus on technical skills in the debriefing and did not utilize the videotape of crew interactions to its fullest extent. Variability was also found between evaluations of different observers for the same crew. This highlights the need for rater training and the need to calibrate the performance ratings.

Butler, R. E. (1993). LOFT: Full-mission simulation as crew resource management training. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 231-259). San Diego, CA: Academic Press, Inc.

Article type: Review
Keywords: Communication; CRM; LOFT; Training Evaluation

This chapter discusses uses and evaluations of line-oriented flight training (LOFT) programs. Crew resource management (CRM) LOFT training is supposed to foster coordination among crewmembers. Good scenarios should be realistic and provide an environment for the crew to make choices which will lead to effective crew performance. Even though accident rates have decreased in recent years with the advent of advanced technology, there are still many problems to be overcome. Issues of safety are not based solely on the equipment or the pilot's proficiency, but also on CRM skills. Training programs are starting to encompass these skills, but the process has been slow. Regulations have changed over the years, making LOFT training program approved required by the FAA. As of 1987, only three airlines had fully integrated training with a recurrent CRM LOFT program in place for all crews. Finally, the author also discusses ways to implement and improve LOFT and cites several key components as required for developing an effective LOFT program: (1) task and goal analysis; (2) baseline CRM training; (3) LOFT evaluation skills training for checking and instructional personnel; and (4) a method for tracking and validating the training process.

Byrnes, R. E., & Black, R. (1993). Developing and implementing CRM programs: The Delta experience. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 421-443). San Diego, CA: Academic Press, Inc.-

Article type: Review
Keywords: CMAQ; CRM; Training Evaluation; Training Program Design

This chapter describes the development of Delta Airlines' crew resource management (CRM) program. The authors believe that the knowledge gained at Delta can be generalized to almost all other organizations no matter the size or type. One outcome of CRM training has been increased safety, but researchers have also found that crewmembers exhibited increased job satisfaction and increased self-esteem. This chapter addresses nine issues relevant for the success of a CRM program, including: (1) corporate cultural influences; (2) methodology and research; (3) curriculum development; (4) module development; (5) preparation of course material; (6) facilitate selection and training; (7) check airmen and management training; (8) classroom facilities; (9) line-oriented flight training and (10) impact of automation on CRM training.

Cannon-Bowers, J. A., Stout, R., & Salas, E. (1994). The role of shared mental models in developing shared situational awareness. In D. V. Day (Chair), *Performance schemas: Issues of development, similarity and change*. Symposium conducted at the ninth annual conference of the Society for Industrial and Organizational Psychology, Nashville, TN.

Article type: Theoretical
Keywords: Situational Judgment; Performance; Teamwork; Training Program Design

These authors present a model explaining team performance in terms of shared mental models. Teams are defined as "a distinguishable set of two or more people who interact, dynamically, interdependently and adaptively toward a common and valued goal/objective/mission, who have each been assigned specific roles or functions to perform and who have a limited life-span of membership." The authors posit that the extent to which teams perform effectively may be due, in part, to the extent to which they are operating with a shared mental model. The authors discuss three types of mental models that can be shared by members of a team. Declarative models contain factual information about the concepts and elements in a domain and their interrelationships. Procedural mental models store information about the steps that much be accomplished in order to reach some goal or end state. Strategic mental models consist of information that is the basis for problem-solving and is considered a product of declarative and procedural knowledge. They suggest that it is only when team members share strategic knowledge that consistent and effective team performance is realized. Different methods of training for shared mental models that arise from this perspective include direct information presentation, positional knowledge training, demonstration modeling, computer animation and cross-training. They also advocate using communication training, team leader training and team de-briefing training to foster shared mental model development.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Converse, S. A. (1991). Toward an integration of training theory and technique. *Human Factors*, *33*(3), 281-292.

Article type: Theoretical
Keywords: Training; Training Evaluation; Training Program Design

The article presents a conceptual framework to guide programs of research. The authors argue that the field of training will not be able to advance until the scientist and practitioner "come together." They offer this conceptual framework as an approach to more closely link training theory with training practice. The framework is organized around three research questions: (1) what should be trained; (2) how should training be designed; and (3) how should training be evaluated. They posit linkages between each training theory and its corresponding training technique as well as linkages between each of the cells containing training-related theories. They also expect that links will emerge between each of the training-related techniques; however, the literature has not progressed to the point where this could be included in the current conceptualization of the framework. The remainder of the article is spent reviewing literature that supports each of the currently proposed links in the framework. They argue that the training field

is currently at a crossroads in which linking training practice to training theory holds the greatest potential for advancing both of these very important aspects of training.

Carroll, J. E. (1979). Flight manager and check-airmen training. In G. E. Cooper, M. D. White, & J. L. Lauber (Eds.), *Resource management on the flight deck, Proceedings of a NASA/Industry Workshop* (pp. 119-132). San Francisco, CA: National Aeronautics and Space Administration.

Article type: Anecdotal/Descriptive
Keywords: Training

Around 1976-77, United began to inform field personnel regarding an increase in the rate of incidents, and to discuss what could be done to respond to this problem. The author discusses two resource management training programs at United. One is a specialized program for flight management personnel and the other is for crewmembers in command or resource management. Prior to this time, training for flight management was non-existent. A brief description of the program is given by the author. At the time of this writing, the command training was still in its planning stages and was being reviewed by line personnel. Some of the reactions to the proposed training were also given by the author.

Carroll, J. E., & Taggart, W. R. (1987). Cockpit resource management: A tool for improved flight safety (United Airlines CRM training). In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 40-46). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: CRM; Training; Training Program Design

According to the authors, a crew resource management (CRM) training program should be looked upon neither as a threat to the pilot's sense of security nor something to be viewed as simply satisfying legal liability issues. Any training provided must be of high quality, accepted by the population and repeated often enough to become an accepted ingredient in the day-to-day operation. There were two issues that were addressed first in developing the United Airlines program: (1) content of what was to be learned; and (2) learning methodology utilized to gain needed behavioral changes. Grid theory was chosen as the basis for teamwork dynamics. Five elements were identified as important for learning cockpit management: Inquiry, advocacy, conflict resolution, critique and decision-making. Phase I of the training program is a home study course which provides a framework for the program. Phase II is conducted in a seminar environment where the crewmembers analyze how they reacted in the cockpit and how different behavioral styles affected outcomes. The recurrent training program consists of a refresher on the theory and principles in CRM, a LOFT exercise which is videotaped, reviewed and discussed with peers.

Cavanagh, D. E., & Williams, K. R. (1987). The application of CRM to military operations. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 135-144). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: CRM; training program; military; civilian

There are many similarities between civilian crews and military crews. There are also, however, some significant differences which need to be recognized and assessed. The general differences fall into six areas: military rank, purpose, crew qualifications, crew lifestyle, labor relations and other miscellaneous differences. Problems with rank may occur when the Captain is junior in rank to the first officer or if the navigator is senior in rank to the aircraft commander. Also, pilots and navigators are commissioned officers while the engineers and loadmasters are non-commissioned officers. Thus, an inexperienced enlisted crewmember may be hesitant in communicating with officers. The motivations to fly are different in the military, and military airlift command (MAC) crews are expected to fly anywhere, anytime, with cargo or passengers, often to unfamiliar destinations with no local ground support. The levels of experience of a flight crew in the military tends to be much less than those of civilian airlines due to the fact that many pilots and flight engineers have military experience prior to working in the civilian sector. In addition, many airline flight engineers have been trained as pilots and are eventually assigned accordingly. Military crews have a much wider variety of ground duties in connection with a flight. There is no formal union within the military. A greater proportion of the training takes place in actual flight, instead of simulators. According to the authors, these general differences may be classified into two categories: operational or cultural. Operational differences mean that military planes may be different from civilian planes and their missions may differ as well. Thus military training programs need to be revised to deal with these specific differences.

Chappell, S. L. (1991). Training and cockpit design to promote expert performance. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 133-138).

Article type: Training Development and/or Evaluation
Keywords: Training; Training Program Design

This paper examined how expert pilots perform practiced tasks. In some domains, expert performance has been labeled "pattern recognition." This performance is said to be skill-based. Experience with a task causes patterns in the environment to be linked to action sequences. After this occurs, when the situation triggers a response, the person reacts automatically, without conscious attention. This type of performance may be desirable in certain cockpit situations. Training for these tasks may be approached in different ways. First, subtasks could be practiced separately and then incorporated into the entire task. Pilots can then be trained to monitor for anomalies. Finally, expert knowledge can be studied to teach the student how the expert views a situation. The cockpit should be designed so that the spatial layout of information will help the pilot recognize familiar patterns. Events must also be present which will trigger the skill-based

behaviors. On the other hand, sometimes barriers need to be in place to prevent a pilot from performing a well-learned behavior in an inappropriate situation.

Chidester, T. R. (1993). Critical issues for CRM training and research. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 315-336). San Diego, CA: Academic Press, Inc.

Article type: Review
Keywords: CRM; LOFT; Regulation; Training

This chapter was written with three goals, to: (1) document recommendations by the FAA and by researchers regarding the evaluation of crew resource management (CRM) training effectiveness; (2) describe the constraints airlines face in implementing CRM training programs; and (3) highlight research issues that need the greatest attention. The FAA developed the first of a series of advisory circulars (FAA, 1989), which recommended that training programs be broken into three phases: (1) an awareness phase, (2) a practice and feedback phase and (3) a reinforcement phase. NASA provided recommendations in two areas: General CRM program implementation issues and specific problems in line-oriented flight training (LOFT) implementation. The author discusses twelve characteristics which are associated with successful CRM programs; six deal with initial implementation and six focus on post training program implementation and more long-term issues. The author also discusses four issues important when implementing LOFT programs. First, when designing scenarios, a set of objectives should be developed that the scenario directly addresses. These objectives should also be conveyed to the instructors. Second, organizations are faced with many different types of constraints when trying to implement training programs, one of the biggest being an economic constraint. Third, there are also logistical and regulatory constraints that should be considered. Finally, there are still many issues that have not received an adequate amount of research attention. For example, how to select crewmembers in the future will become a more difficult process because the pool of applicants will have lesser amounts of experience. Finally, research needs to move beyond the cockpit. Policies and procedures should be developed for facilitating communication between the cockpit and other work groups. As of this writing, this has not been studied. According to the author the unit of analysis should shift from the flight crew to include the cabin crew. At this time, accidents have provided a great deal of information for training. Accidents which did *not* occur could also provide valuable insight. These sources are still untapped, CRM modules could very easily be built around these success stories.

Chidester, T. R., Kanki, B. G., Foushee, H. C., Dickinson, C. L., & Bowles, S. V. (1990, April). Personality factors in flight operations: *Volume I. Leader characteristics and crew performance in a full-mission air transport simulation* (NASA Technical Memorandum 102259). Moffet Field, CA: National Aeronautics and Space Administration.

Article type: Group Comparison Study
Keywords: Personality; Simulation

This technical report describes the results of a full-scale simulation study conducted at NASA Ames Research Center designed to assess the impact of individual personality on crew performance. Commercial aircrews (23 3-person crews; $N = 69$) flew 5 flight segments over a two-day period. Prior to flying the simulator, the 23 pilots completed a battery of personality instruments composed of the Expanded Personal Attributes Questionnaire, the Work and Family Orientation Questionnaire and the Achievement Striving and Impatience/Irritability scales (derived from the Jenkins Activity Survey). Scoring resulted in 10 scale-scores: instrumentality, expressivity, negative instrumentality, verbal aggressiveness, negative communion, work, mastery, competitiveness, achievement striving and impatience/irritability. Pilots were classified into one of three clusters: (1) IE+ (above median on 3 of following dimensions -- instrumentality, expressivity, mastery, work); (2) I- (above median on negative instrumentality and verbal aggressiveness, below median on expressivity); (3) EC- (above median on negative communion and below median on 3 of following dimensions -- instrumentality, achievement striving, mastery, work, impatience/irritability). Criteria included observer ratings, error analyses and aircraft handling (collected by the simulator's computer). IE+ Captained crews were consistently effective and made few errors. EC- Captained crews were consistently less effective and made more errors. I- Captained crews were less effective on the first day, but were equal to the best on the second day, suggesting familiarity facilitates performance among I- led crews. The authors believe that results from the current 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.

Chidester, T. R., Kanki, B. G., & Helmreich, R. L. (1989). Performance evaluation in full-mission simulation: Methodological advances and research challenges. In R. S. Jensen (Ed), *Proceedings of the Fifth International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.

Article type: Training Development and/or Evaluation
Keywords: Communication; Coordination; Performance; Performance Ratings; Simulation

This article describes the "full-mission flight simulation" that was developed by Chidester, Kanki, Helmreich and colleagues. These simulations were designed to "mirror" as much as possible an actual flight for an aircrew. Generally, the simulations consist of several legs and participants are provided with detailed descriptions of weather conditions, maintenance

information, air traffic control (ATC) information, etc. The scenarios are designed to "maximize generalizability" and to require cooperation and coordination between the crewmembers. Three basic types of criterion data are collected: (1) expert observations of individual and crew performance (e.g., communication, manual control); (2) crew communication as measured via audiotape and videotape; and (3) aircraft handling parameters (which are measured via the simulator computer). Error analyses are conducted using expert observers who keep a record of all errors observed during the simulation. In addition, videotape records are analyzed by at least two condition-blind experimenters, then viewed by the expert observer who has the option of eliminating one of the errors based on his or her records from observing the simulation. Three areas of further study are noted: (1) comparing simulator performance using two different versions of the same aircraft (one glass cockpit and one electromechanical); (2) documenting the *process* through which decisions are made in the cockpit; and (3) comparing CRM training effectiveness across aircraft fleets and organizations. This simulation research methodology is described as "a strategy emphasizing careful design of full-mission flight scenarios to capture real-world performance processes, recruitment of subjects with credible operational experience and assessment of the crew performance via converging sources of data."

Christian, D., & Morgan, A. (1987). Crew coordination concepts: Continental Airlines CRM training. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 68-74). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: Communication; CRM; Decision-Making; Leadership; Training Program Design

The authors note that in order to be an effective crewmember, each person needs to understand his or her own basic behavioral style and the behaviors that are required to perform effectively. Continental's program is based on the Grid Model, which has the task listed on the X-axis and the relationship on the Y-axis. Each quadrant in the grid is labeled: Quadrant 1 - nurturing (hi-relationship, lo-task), Quadrant 2 - autonomous (lo-relationship, lo-task), Quadrant 3 - aggressive (lo-relationship, hi-task), Quadrant 4 - assertive (hi-relationship, hi-task). In a workshop, a personal style instrument, the Strength Deployment Inventory (SDI), is used to help each person understand how they relate to and communicate with others. A review of ASRS reports indicated that there was a need for pilot development in five key areas: (1) *Communication* - this is accomplished by developing skills in advocacy and inquiry, which are ways to present information when it is needed; (2) *Leadership* - need well-developed skills in task and relationship. Leadership behavior is best determined by the situation; (3) *Role*; (4) *Decision-making* - responsibility making decisions is not only on the Captain; and (5) *Resources* - all crewmembers must fully utilize all of the resources available to them.

Clark, R. E., Nielsen, R. A., & Wood, R. L. (1991). The interactive effects of cockpit resource management, domestic stress and information processing in commercial aviation. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 2* (pp. 776-781).

Article type: Training Development and/or Evaluation
Keywords: Stress; CRM; judgment

In this study, the authors examined the interactions between domestic stress, information processing ability and crew resource management (CRM) training. One hundred thirty-five subjects each completed a survey designed (by the authors) to measure aviation and domestic stress, information processing and interaction patterns. Most respondents stated that their stress level had stayed the same or increased in the last 90 days and in the last year and most of the participants also reported that when they were under stress, they experienced a decrease in their information processing ability. The CRM training program was said to increase their information processing ability under stressful conditions, however, the majority of respondents stated that the CRM training could help them to overcome some, but not all, of the negative effects of stressors. It is interesting to note that 71 percent of the participants indicated that CRM would also have positive benefits in other aspects of their lives.

Clothier, C. (1991). Behavioral interactions across various aircraft types: Results of systematic observations of line operations and simulations. *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 332-337). Columbus: Ohio State University.

Article type: Group Comparison Study
Keywords: CRM; LOFT

This was a review conducted in one major domestic airline, analyzing crew behavior with the NASA/UT Line/LOS checklist, before and after implementation of a crew resource management (CRM) training program. The second version of the checklist was used which includes fourteen behaviors considered critical to flight crew interaction. All pilots were trained in 1989 and then had recurrent training in 1990. Observations of crew interactions took place both on the line and in line-oriented flight training (LOFT) sessions. Results showed that crew behavior was rated higher as they progressed through the different levels of training. Greatest improvement was seen in the Briefing and Crew Self-Critique categories and least improvement was gained in Task Concern and Technical Proficiency. Trained crews' overall performance was compared to untrained crews in line and LOFT settings. The trained crews showed improved performance in almost all categories. The author also found that crews interacted differently in different aircraft types. In LOFT, crews in advanced technology cockpits outperformed the crews in standard cockpits, but the opposite was found for performance on the line. Other results showed that two-person crews tended to outperform three-person crews regardless of cockpit type in both line and LOFT settings. Based on these results, the author suggests that CRM training has improved the effectiveness of the crews. One reason given to explain why those in advanced cockpits show higher performance during LOFT is because crews have more time to gather and process

information instead of being as involved in actually flying the plane. Two person crews may be more effective due to the fact that communication can be simpler. Finally, younger pilots may be more receptive to CRM concepts because it is part of their training to become safe pilots and they have not been as much a part of the authoritarian atmosphere in the cockpit.

Command/Leadership/Resource Management: *Home study course and Seminar Materials Manual for United Airlines.* (1985). Denver, CO: United Airlines.

Article type: Training Manual
Keywords: Communication; Coordination; Leadership; Personality; Training

This packet of training materials ("Home Study Course"; "Seminar Materials Manual") is intended to be used by United Airlines pilots prior to attending the United Airlines C/L/R seminar. Both sets of materials focus on the notion that pilots must not only be skilled aviators, but skilled managers as well; emphasizing the fact that accidents are increasingly due to the lack of proper C/L/R management. The "*Home Study Course*" is composed of four major parts: (1) fictional stories illustrating aircrew events and problems; (2) an introduction to Blake & Mouton's *Managerial Grid* applied to C/L/R; (3) examination of accident and incident reports as they relate to effective resource management; and (4) references and additional readings. Chapter 1 uses in-flight scenarios to make critical points about C/L/R and to stimulate thinking. Chapter 2 introduces the C/L/R Grid as a way to recognize styles in interpersonal relationships and how individual and group team member behavior influences cockpit effectiveness. Chapter 3 once again relies on the C/L/R Grid to initiate resource management discussions about communication, coordination, conflict resolution and critique. Chapter 4 does the same with leadership issues such as problem definition, inquiry, advocacy and decision-making. Chapter 5 focuses on command styles, the Captain's authority and the Captain's role in crew training and development and the interaction of the Captain's command style and the styles of crewmembers. Chapter 6 discusses perceptions versus reality, the importance of testing your assumptions (including advocacy and inquiry) and managing cockpit roles (role ambiguity, role conflict, role overload). Finally, Chapter 7 provides a set of readings from military aviation, NASA, NTSB and an airplane manufacturer about the importance of C/L/R. The "*Seminar Materials Manual*" provides the pilot about to participate in the C/L/R seminar with prework and seminar reading materials so that participants will be better prepared. These materials are aimed at helping the participants reach five target goals: (1) learning the C/L/R Grid as a framework; (2) gaining insight into one's behavioral style; (3) setting standards for advocacy based on openness and candor; (4) the effective use of Captain's authority and crewmember leadership; and (5) teamwork.

Conley, S., Cano, Y., Bryant, D. (1991). Coordination strategies of crew management. *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 260-265). Columbus: Ohio State University.

Article type: Group Comparison Study
Keywords: Civilian; Communication; Coordination; Simulation; Performance Ratings; Training

These researchers compared coordination strategies between two three-person flight teams in a B-727 simulator. Specifically, they examined the degree to which communication patterns were consistent with two modes of coordination: (1) standardization; and (2) planning/mutual adjustment. These coordination techniques were inferred from crew communication across two flight phases (routine and non-routine). Crews participated in a full-mission flight simulator study in which they were required to fly five legs. Cockpit transcripts, videotapes and narrative records from the last leg of the simulation were analyzed. Observer ratings of errors and effectiveness were used to select one high performing and one low performing crew to be compared and contrasted in this study. The more effective crew consistently communicated more frequently, across flight phases, than the less effective crew. Also, many more statements classified as monitoring/evaluating occurred in the effective crew as compared to the ineffective crew. These results suggested that the less effective crew relied more upon standardization during the simulation, whereas the more effective crew relied more upon planning/mutual adjustment. These findings were also more pronounced for the routine flight phase than they were for the non-routine flight phase.

Cook, E. (1987). The regulatory horizon. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 178-187). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: Civilian; CRM; Training Program Design

The author discusses two topics dealing with how the Federal Aviation Administration (FAA) views crew resource management training. First, the stance of the FAA in 1987 was neutral toward training and they did not know enough about the programs to predict what the regulatory activity would be in the future. The second topic has to do with the fact that some airlines have received exemptions to allow their pilots to go through recurrent training on an annual basis. In 1987, the FAA was awaiting data that purported to measure how successful crew resource management training had been to be used in making further regulatory decisions.

Costley, J. (1991). Pilot reaction to ultra-long-haul flying. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 372-376).

Article type: Single Group Study
Keywords: CRM; Stress; Training; Training Program Design

Aviation has changed over the years, especially with respect to the length of flights that can now be accomplished. In this paper, the effects of long-haul flights are examined from a resource management perspective. Interviews were conducted with 34 pilots who were regularly flying flights of 12-15 hours duration. This type of operation evoked either very positive (32.3%) or very negative reactions (67.7%) from the pilots. Sleep seemed to be the key variable, since those who did not like this type of flight were the ones who could not rest while away from the controls. The issue was whether pilots could be trained to get more effective rest on the aircraft and cope with the situation outside the aircraft. Some research has found that controlling the diet during all phases can help. The author presents an outline which should be followed when preparing to sleep on the aircraft. Results with this technique have been good for those who before could not fall asleep. The crewmember can help by using self-management techniques, which could very easily be taught as part of CRM training. Stress management should also be incorporated into CRM programs.

Costley, J., Johnson, D., & Lawson, D. (1989). A comparison of cockpit communication B737-B757. *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 413-418). Columbus: Ohio State University.

Article type: Group Comparison Study
Keywords: Attitude; Civilian; Communication; Coordination

Fears are expressed that the coming of automated (EFIS) flight decks might change the nature of communication and interaction between flight crew and consequently, decrease the safety of flights. This study compared inter-pilot communication on the flight decks of three types of aircraft: the 737-200 (conventionally instrumented flight deck; used as the "baseline"); the 737-300, with 'glass' flight instrumentation; and the 757 with full 'glass' cockpit. Communication was observed and recorded by an observer present on the flight deck for 19 flights (17 day; 2 night). Flight length ranged from about 1 hour to more than 4 hours. Only data that were collected from the climb, cruise and descent (i.e., gear up to gear down) were presented. Including all categories of communication (both operational and non-operational), the two 737s did not appear to differ, but the rate of communication for the 757 was substantially less than both of the other planes. The authors also examined the data in terms of a ratio of commanding to giving and seeking (i.e., 'push-pull' ratio), that was used as a rough measure of authoritarianism. They found that the flight decks of the more highly automated aircraft tend to be more authoritarian. Two general conclusions from this study were that, as the level of automation on the flight deck increases, there was a trend toward less communication and a higher incidence of authoritarian behaviors.

Diehl, A. E. (1992, November). *The effectiveness of civil and military cockpit management training programs.* Paper presented at the 45th International Air Safety Seminar, Long Beach, CA.

Article type: Review
Keywords: Coordination; Leadership; Situational Awareness

Only in the last decade have formal training programs evolved which are designed to reduce the primary causes of contemporary mishaps -- cockpit management problems. These programs address crew coordination, situational awareness and judgment issues. The author notes that crew resource management (CRM) is the term used more commercially, while aircrew coordination training (ACT) is favored by the military. Diehl also mentions aeronautical decision-making (ADM) and judgment training courses (both of these tend to focus on enhancing attitudes and cognitive abilities of individual pilots) have been successfully incorporated into the general aviation arena. The author presents statistics (for 1987-1989) illustrating that the most frequent type of aircrew error is tied to aircrew coordination issues. During the 1970's, the aviation industry (public and private sector) began to develop CRM training programs and the article suggests that today, most progressive airlines have CRM programs. Less is known about programs designed to improve the decisional judgments of the individual pilot. The author believes that distinctions between CRM and ADM training are disappearing, with a common focus on five primary areas: (1) attention management; (2) crew management; (3) stress management; (4) attitude management; and (5) risk management. In terms of measuring training effectiveness, the author's review of the literature suggests that most of the studies have been in the area of *attitude improvement* (most using the CMAQ); some also use the LLC (Line/Line Operational Checklist). In general, the author notes that Helmreich's work suggests improvement in attitudes after CRM training and more frequent use of positive CRM-type behaviors. In addition, a number of studies focus on: (1) *controlled experiments to reduce error rates* (ADM programs have been extensively tested in this way, using simulations, with statistically significant differences between trained and untrained groups); (2) *case studies* (personal accounts reinforce value of CRM); and (3) *measuring reduced accident rate.* In general, across users (e.g., Bell Helicopter, Petroleum Helicopter, US Navy and USAF Military Airlift Command), there is a reduction in error rates over a number of years; this is attributed, in part, to CRM/ADM training). The author summarizes this review by saying that today there exist many diversified applications of CRM, including USAF "single-seat" fighters, FAA air ambulance program administrators, USAF ATCS for telemetry or flight tests, USN combat information center personnel and Continental Airlines maintenance personnel.

Diehl, A. E. (1993, December). Crew "recourse" management. *Flying Safety*, 20-21.

Article type: Training Development and/or Evaluation
Keywords: Accidents; CRM; Communication; Decision-Making; Situational Judgment; Training Program Design

This article describes the evolution of crew resource management (CRM) training concepts and provides the direction US Air Force CRM programs are likely to take in the future. United Airlines started the first CRM training program after a DC-8 crashed in Oregon because the Captain was "deeply involved" with a landing gear-unsafe light and ignored repeated "hints" from his crew that fuel was low. Soon after this time, most other airlines' CRM programs also began. Currently, most of these first-generation civilian and military CRM programs are undergoing major revisions. The Air Force is targeting CRM-related problems like inadequate briefings, imprecise communications and unchallenged breaches of judgment. The groundwork for development of an Air Force-wide CRM program was outlined at a series of meetings conducted in the past year. One expectation is that there will be guidelines for specific skills that need to be taught, in addition to a layered training program in which participants will progress through several levels of training. These are expected to include introductory, airframe-specific and mission-specific training, as well as special courses for instructors, evaluators and supervisors.

Driskell, J. E., & Adams, R. J. (August, 1992). *Crew resource management: An introductory handbook.* Washington, DC: U.S. Department of Transportation, Federal Aviation Administration, Research and Development Service.

Article type: Review
Keywords: Communication; Coordination; Training

Driskell and Adams describe their objectives as to foster an understanding of the background and philosophy of crew resource management and to provide an overview of the development, implementation and evaluation of crew resource management (CRM) training. As such this handbook serves as a supplement to FAA Advisory Circular 120-51 (CRM). The authors provide a brief history of CRM and define CRM as the effective utilization of all available resources -- hardware, software and personnel -- to achieve safe, efficient flight operations and delineate CRM principles. These principles include: (1) effective performance depends on both technical proficiency and interpersonal proficiency; (2) a primary focus of CRM is effective team coordination (the team encompasses flight crew, dispatchers, ATCSs, maintenance); (3) CRM focuses on crewmembers' attitudes and behaviors; (4) effective CRM involves the entire flight crew; (5) the acquisition of CRM skills requires active participation of all crewmembers; and (6) CRM training should be blended into the total training curriculum. The authors spend much of the handbook describing the CRM skills necessary for effective performance. These skills are grouped into three clusters: (1) communication and decision-making (assertiveness, communications, decision-making, conflict resolution); (2) team building and maintenance (leadership, team management); and (3) workload management and situational awareness

(mission planning, stress management, workload distribution). Finally, the authors discuss the steps involved in CRM training, including: (1) development (needs assessment, setting performance objectives and preparing a training plan); (2) implementation (seek participation, demonstrate program support, communicate, follow-up); and (3) evaluation (measurement of pre-training performance, trainee reactions, learning, performance, organizational outcomes).

Estridge, W. W., & Mansfield, J. L. (1979). Upgrade and interpersonal skills training at American Airlines. In G. E. Cooper, M. D. White, & J. L. Lauber (Eds.), *Resource management on the flight deck, Proceedings of a NASA/Industry Workshop* (pp. 87-96). San Francisco, CA: National Aeronautics and Space Administration.

Article type: Anecdotal/Descriptive
Keywords: Training Program Design

The first author describes the former Captain upgrade program at American Airlines and notes complaints received about the program. Consequently, new objectives and a new upgrade program were developed. A series of presentations are given to the new Captains as well as a few supporting activities outside of the company. A brief description of the new program, based on transactional analysis, is given. The second author presents segments of the actual training program.

Federal Aviation Administration. (1993). *Crew resource management* (Advisory Circular 120-51A). Washington, DC: Author.

Article type: Review
Keywords: Communication; Coordination

The purpose of the circular is to present federal aviation administration (FAA) guidelines for developing, implementing, reinforcing and assessing crewmembers essential for flight safety. The pamphlet suggests that 60-80 percent of all air carrier incidents/accidents involve human errors and notes that many problems encountered by flightcrews have little to do with technical operation of multiperson cockpit. Rather, the problems are associated with poor group decision-making, ineffective communication, inadequate leadership and poor task/resource management. The circular suggests that NASA/FAA research has found that crew resource management (CRM) trained crews operate more effectively with non-routine situations. CRM encompasses dispatchers, cabin crewmembers, maintenance personnel and air traffic control specialists (ATCSs). This document emphasizes that in order to be effective CRM concepts must be permanently integrated into all aspects of training and operations. The circular describes critical components of effective CRM training (i.e., initial indoctrination/awareness, recurrent practice and feedback and continuing reinforcement), suggested curriculum topics (i.e., communications processes and decision behavior and team building and maintenance) and training program assessment.

Fiedler, M. T. (1987). CRM training in the 1550th combat crew training wing. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 145-147). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: CRM; Military; Training Program Design; Training Evaluation

According to the author, this unit's Aircrew Coordination Training (ACT) has become part of a five-day simulator systems refresher course. On the first day, crews receive eight hours of aircrew coordination training academics. On day two, each crew has a mission-oriented simulation training (MOST) period. Days three through five provide crews with both systems academics and systems refreshers in the simulator. During day one, crews state what they believe to be the most common crew interaction problems. Past accidents are reviewed. The following elements of crew coordination are discussed: inquiry, advocacy, conflict resolution, decision-making and critique. Communication is discussed reinforcing the different leadership and followership styles. A MOST exercise is videotaped and discussed when completed. Evaluation of the program has found that 97 percent of the participants have rated it as effective.

Foster, G. C., & Garvey, M. C. (1979). Left seat command or leadership? Flight leadership training and research at North Central Airlines. In G. E. Cooper, M. D. White, & J. L. Lauber (Eds.), *Resource management on the flight deck, Proceedings of a NASA/Industry Workshop* (pp. 133-147). San Francisco, CA: National Aeronautics and Space Administration.

Article type: Anecdotal/Descriptive
Keywords: Communication; Motivation; Training Program Design

The first author gives a brief overview of management personnel training at North Central Airlines, which is based on the Blake and Mouton managerial grid. The company then decided to expand this concept to the cockpit. The first program developed was a two-day seminar and after review by a group of check pilots and airline pilots association (ALPA) representatives, it was revised to be more prescriptive for specific situations and now encompasses three days. The second author gives a more detailed explanation of the training program, including an overview of the management grid concept. Captains are asked to describe behaviors, reactions and consequences for the different styles of management (examples are given). The central role of communication, motivation and expectations are also discussed. Captains are to use the framework to improve their judgments and predictions. The program also includes presentations by maintenance, dispatch, flight attendants and air traffic control to give Captains a clearer understanding of the job performed by those people whom he or she will be interacting with.

Foushee, H. C. (1982). The role of communications, socio-psychological and personality factors in the maintenance of crew coordination. *Aviation, Space and Environmental Medicine, 53*(11), 1062-1066.

Article type: Review
Keywords: Communication; Coordination; Leadership

The author cites research, personal observation and testimonials in emphasizing the importance of crew coordination on the flightdeck. He points to small group research as applicable to the transport aircraft situation and believes a number of social and personality variables are relevant as well. Foushee stresses two variables as central to person-situation type of analysis: (1) goal orientation (or instrumentality) -- persons high on this characteristic are performance oriented, decisive and get the job done; and (2) group orientation (expressivity) -- persons high on this characteristic are expressive, sensitive to others feelings, warm in interpersonal relationships and communicative. Traditionally, airlines selected for goal oriented types, while largely ignoring group oriented types. It is now becoming clear that cockpit communications are crucial determinants of information transfer and crew coordination. Research has also shown that cockpit communication is related to group cohesion, attitudes toward work and complacency. He notes that some airlines are developing CRM training programs that provide feedback to pilots on certain personality dimensions and their role in crew coordination. Such training does not seek to change personalities, but assumes that an increased awareness of group dynamics will cause many to think before reacting in ways detrimental to group cohesion and performance.

Foushee, H. C. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. *American Psychologist, 39*(8), 885-893.

Article type: Theoretical
Keywords: Personality; Simulation

Foushee notes that despite efforts by human factors experts and others to improve cockpit designs, still (in 1982) about 65 percent of all aircraft accidents fall into the human error category. He stresses that more emphasis needs to be placed on crew performance and the factors that affect crew coordination. The author's orientation is that the cockpit crew is a highly structured small group and as such a number of socio-psychological, personality and group process variables are relevant to crew effectiveness. He notes that group performance is "understudied" because it is: (1) difficult to assimilate the sheer number of variables; and (2) easier to control examination of individual performance (especially in a laboratory). The author reviews input, process, outcome variables in the small group performance situation faced by aircrews. He examines links between personality or interpersonal styles of leadership as input variables; and operational errors, incidents and accidents as outcome variables. Foushee then summarizes some of his previous research on group-vs.-goal orientation and information/ communication patterns. He describes on-going airline crew coordination programs, including those that focus on simulation performance videotape feedback; seminars that focus on role of interpersonal styles, personality assessment, feedback, role playing, case studies and

interpersonal encounters; LOFT training; and the need to re-orient selection to include leadership/team functioning issues.

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

Article type: Review
Keywords: Communication; Coordination; Teamwork

This article examines the group performance process as it impacts the effectiveness of airline crews in multipilot aircraft. The authors note that the increasing reliability and performance of modern day aircraft has all but eliminated incidents and accidents that are due directly to mechanical failure. Rather, the vast majority of aircraft accidents are due to human error. They also note that one attempt to reduce the impact of human error, the advent of multipiloted aircraft to increase role redundancy, has failed to provide an adequate safeguard in several very salient cases. They present a framework in which group performance factors can be classified into three rather broad categories: input variables, process variables and outcome variables. Input variables are personal characteristics that are brought to the situation, both individual and group, as well as environmental. Process variables refer to the dynamics of group interaction. Outcome variables refer to how well the group performs specific tasks. They note that some of the least understood and least studied, but most important factors, are those related to process. The most easily measured and straightforward of these is the communication process. Relatively few research studies have examined the effects of process variables on performance; instead most researchers have examined a direct link between input variables and outcome variables. Thus, it is not surprising that few consistent findings have resulted from this research, since it is very likely that process variables moderate the relationship between input and outcome variables. Next, literature is reviewed showing that crews who communicate more effectively tend to perform better and to make fewer errors than those who do not. They also show that the type of communication (i.e., group oriented vs. task-oriented) plays an important role in how well an aircrew performs. Past research is also reviewed which suggests that the best way to effect meaningful change in this type of small group performance is to concentrate on the input variables. Three general approaches to bringing about this change are: altering group norms, increasing member effort and coordination (this also appears to be related to process variables) and changing group composition.

Foushee, H. C., Lauber, J. K., Baetge, M. M., & Acomb, D. B. (1986). *Crew factors in flight operations: III. The operational significance of exposure to short-haul air transport operations.* (NASA Technical Memorandum #88322). Moffett Field, CA: NASA Ames Research Center.

Article type: Group Comparison Study
Keywords: CRM; LOFT; Personality

This project focused on the effects of fatigue-related problems in short-haul flight operations. Also, individual attributes were examined to see if some people had adaptive strategies which allowed them to cope more effectively. Studies on the effects of fatigue have been conducted mostly in the laboratory, with mixed results. Simple tasks show more deleterious effects over time than more complex tasks. Studies using cockpit-like tasks have found inconsistent results. Studies have also looked at the decrement in psychomotor flying skills, but no one has explored the effects of fatigue on higher-level decision-making skills, which have been cited as one of the main causes of more recent aircraft accidents. For the current study, crews were placed into one of two experimental conditions: (1) "Post-duty" condition - these pilots had just completed a three day trip (9 crews); and (2) "Pre-duty" condition - these pilots had just completed a minimum of two days off-duty (11 crews). Prior to the simulation flight, all crews were provided with the same information they would receive prior to any actual flight. A flight planning phase by the crew followed. The crew flew the aircraft according to standard operating procedures and were debriefed at the completion of the study. All pilots completed an extensive background questionnaire, a measure of their perceptions of fatigue (after the simulation) and a 26-item adjective checklist. The following crew performance measures were used: expert observer ratings, subjective assessments of workload, aircraft handling data, error analyses (both real-time and videotape) and crew communication patterns. No demographic differences were found between pilots in the two conditions. Fatigue data showed that "post-duty" crewmembers reported less sleep, more "tiredness," and more negative mood states than did "pre-duty" crewmembers. Overall post-duty crewmembers had significantly higher ratings in several phases of flight. Captains in the pre-duty condition reported that they had exerted significantly more mental effort. The aircraft handling data was only analyzed for one flight segment. Results showed that post-duty crews performed closer to the desired levels than pre-duty crews. Number of errors committed by crews in both conditions were not significantly different, but post-duty crews had means lower than pre-duty crews. A "crew familiarity" factor was cited as the reason for these results. Analyses of all the communications data found that post-duty crews and crews that had flown together engaged in more task-related communication and less non-task related communication. As task demands increased, so did communication behaviors. According to this study, it might appear that it does not matter how long a duty-cycle is because it had little to no effect on flight safety. It may not be a good idea, however, to assign permanent crews because this may lead to a decrement in performance in the long run. It is unknown at this time how much familiarity is necessary to increase crew coordination.

Foushee, H. C., & Manos, K. L. (1981). Information transfer within the cockpit: Problems in intracockpit communications. In C.E. Billings & E. S. Cheaney (Eds.), *Information transfer problems in the aviation system* (NASA TP-1875). Moffett Field, CA: NASA-Ames Research Center.

Article type: Group Comparison Study
Keywords: Civilian; Communication; Coordination; Teamwork

The authors note that over 70 percent of all accidents reported to the Aviation Safety Reporting System (ASRS) cite human error as the major cause. In addition, a common difficulty appears to be failure of the information transfer process. This study examined communication patterns among cockpit crewmembers and its impact on flying performance. Researchers content analyzed the voice recordings from 12 B-747 full mission flight simulations flown by fully qualified crews. From these recordings, an attempt was made to classify each statement or phrase into one of 10 categories. The ten categories and the percentage of all communications that were placed in each category are as follows: crewmember observation (44%), inquiries (25%), commands (12%), acknowledgments (11%), tension release (2.5%), agreement (1.6%), frustration or anger (1%), response uncertainty (0.8%), embarrassment (0.5%) and pushes (0.5%). Overall, there was a tendency for crews who did not communicate effectively to not perform as well as those that did communicate effectively. There were significant negative correlations between systems operational errors and both crewmembers observations (r = -.51) and acknowledgments (r = -.61). Acknowledgments were also negatively related to total errors (r = -.68). In addition, commands were associated with a lower incidence of flying errors (r = -.64). Thus, intra-cockpit communications were significantly related to overall flight performance of these B-747 crews. The authors note that both the lack of communication and the *style* of communication also had substantial impact on crew performance.

Franz, T. M., Prince, C., Cannon-Bowers, J. A., & Salas, E. (1990). The identification of aircrew coordination skills. *Proceedings of the 12th Annual Department of Defense Symposium* (pp. 97-101). Colorado Springs, CO: U.S. Air Force Academy.

Article type: Review
Keywords: Coordination; Communication; Teamwork; Training Program Design

The authors note that many resources have been directed toward developing aircrew coordination training programs, but much less effort has been directed towards developing ways to evaluate these programs. This article presents an outline of methodology used to develop a listing of behaviors that could be used to evaluate a skill-based aircrew coordination training program for rotary wing crews. Thirty-seven behaviors were identified from a literature review and another 18 behaviors were identified through interviews with pilots (no examples of critical behaviors are provided). These behaviors were then classified by job experts into one of seven categories: mission analysis, assertiveness, adaptability/flexibility, situational awareness, decision-making, leadership and communication. Survey data indicated that pilots (N = 137) considered the

aircrew coordination behaviors to be critical, frequently occurring and important to train. However, they were not rated as very difficult to perform.

Freemon, C., & Simmon, D. A. (1991). Taxonomy of crew resource management: Information processing domain. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 391-397).

Article type: Theoretical
Keywords: Accidents; CRM

This article presents a taxonomy for analyzing the flow of information in the cockpit. It is broken into five phases: (1) problem recognition - what is the problem; (2) setting of objectives - what should be done about it; (3) planning - how should it be done; (4) implementation - implementing the decision; (5) critique - did it work. When the area of a breakdown is identified, then the skills and strategies to deal with the problem can be taught to the pilots. Thus, each organization can focus its training specifically to its crews' needs. Four years of accidents were analyzed using the taxonomy. From this investigation, a series of recommendations are generated by the author for improving safety.

Funk, K. (1991). Cockpit task management: Preliminary definitions, theory and recommendations. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 222-226).

Article type: Review
Keywords: Decision-Making; Training

This paper presents a model of how a crew can integrate the multiple behaviors which are required to fly an airplane. Crews are typically called upon to perform multiple tasks simultaneously. A preliminary framework termed called Cockpit Task Management (CTM) is presented and it's implications for crew resource management (CRM) are described. CTM describes a procedure crews may follow for task management. An error taxonomy has also been developed. The key elements are task initiation, task interruption, task resumption and task termination. He recommends that CTM should be incorporated into pilot training and implemented in improved cockpit designs.

Ginnett, R. C. (1990). Airline cockpit crew. In J. R. Hackman (Ed.), *Groups that work (and those that don't).* San Francisco: Jossey-Bass.

Article type: Anecdotal/Descriptive
Keywords: Coordination; Teamwork

This paper is an anecdotal description of an airline crew (Captain, first officer, flight engineer and 4 attendants) as they work as a team during an 85 hour period. One main point was that in this team, as with many types of organizational teams, very little time was devoted to forming and building the team. Like almost all airline crews, they were forced to come together and begin performing effectively very quickly. This underscores the importance of those early moments in that transition from a group of strangers to an integrated crew. Four behaviors exhibited by the Captain were seen as critical to the formation of this team: (1) explicitly discussing tasks that required coordination between attendants and flight crew; (2) defining and expanding crew boundaries; (3) explicitly setting norms for crew behavior; and (4) appropriately managing the dynamics surrounding the authority that the Captain, by definition, brings to the situation. The Captain is characterized as providing the "shell" in which the rest of the team operates. This shell is, in part, provided by the organization, but it is up to the Captain to expand the shell and provide the crew with a larger definition of tasks, boundaries and norms under which they need to work.

Ginnett, R. C. (1993). Crews as groups: Their formation and their leadership. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 71-98). San Diego, CA: Academic Press, Inc.

Article type: Review
Keywords: CRM; Teams; Leadership; Group Process

In America, individual-level achievement is emphasized moreso than group-level achievement. As the author and others discovered, this is also very true of training for pilots. This chapter advocates that the time has come for the aviation community to learn about groups. Research has shown that when people work together on exercises, the performance of the group frequently exceeds the performance of individuals. There are certain conditions, however, which are associated with groups which need to be better understood, including group boundaries, role conflict, group norms, status within the group and group dynamics. Each of these concepts is discussed in some detail by the author. The author next turns to ways that the group can gain leverage and more effectively perform. Past models of group performance have looked at input-throughput-output. The problem is that many training programs have tried to change the output by modifying the throughput. But if the inputs are poor from the start, then it doesn't matter what happens during the process. One factor which can have a big impact on crew performance is the authority dynamic. In the past, the single authoritative command has been dominant. History, regulations and individual characteristics all tend to be forces which have pushed the Captain to be authoritative. In some instances, this type of authority is good, but not always. The manner in which highly effective Captains engaged the crew during the briefings involved assuming

legitimate authority in three ways. The Captains conducted the briefing in some logical manner, used language specific to flying and were comfortable in their role as leader. These Captains go on to involve the crew in the responsibility for the flight. The crew is shown that the Captain is willing to face his or her own vulnerabilities. When these Captains conduct briefs, they do so in an interactive, not authoritarian manner. Highly effective Captains also encouraged other crewmembers to contribute (e.g., via questioning/suggesting) to the planning process.

Goguen, J., Linde, C., & Murphy, M. (1986). *Crew communication as a factor in aviation accidents.* (NASA Technical Report 88254). Moffett Field, CA: NASA-Ames Research Center.

Article type: Single Group Study
Keywords: Communication; Coordination

The authors posit that crew communication problems are a major source of air transport accidents. Researchers examined the voice cockpit recordings from eight commercial aviation accidents. These 1,725 speech acts from the eight accidents were coded according to twelve categories (most of which were drawn from existing linguistic theories): speaker, addressee, discourse type, new or old topic, topic success or failure, draft order, ratification, speech act type, mitigation level, crew recognized emergency, crew recognized problem and operational relevance. Two of the eight recordings were used to formulate hypotheses and the remaining six were the test group. The results suggest that crew communication is a complex phenomenon and effective and efficient crew communication is crucial to mission success. The authors propose extending this methodology from the detailed analysis of cockpit discourse to non-verbal communications and other factors in the aircraft system.

Green, R. G. (1985, July). Stress and accidents. Fourth annual scientific symposium of the United Kingdom Association of Aviation Medical Examiners: Stress in aviation. *Aviation, Space and Environmental Medicine, 56*(7), 638-641.

Article type: Review
Keywords: Accidents; Stress

The author examines whether the different types of stress pilots were exposed to increased the probability of errors and accidents. Three types of stress are discussed: (1) environmental stress (noise, heat, sleep deprivation, etc.); (2) acute reactive stress ("fight or flight syndrome"); and (3) life stress (recent life events, or other traumatic occurrences). Green suggests that accidents caused by environmental stress are rather infrequent and are due most often to sleep deprivation or fatigue. Acute reactive stress (how pilots react in an emergencies) typically results in effective responding (experimental evidence is weak, however). Simulator training can be helpful in preparing appropriate pilot responses to emergencies. The author suggests that a life-stressed pilot may have his or her mental capacity partially devoted to life problems and events, thereby increasing the chances for errors and accidents.

Gregorich, S. E. (1991). What makes a good LOFT scenario? Issues in advancing current knowledge of scenario design. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 2* (pp. 981-987).

Article type: Training Development and/or Evaluation
Keywords: LOFT; Training Evaluation; Training Program Design

Line oriented flight training (LOFT) has been recognized as a tremendous training tool. Many different scenarios have been designed, but the relative strengths and weaknesses of these scenarios have not been investigated. The author discusses three questions which need to be addressed: (1) What combination of objective task demands were built into the scenario to obtain the intended training objectives? (2) To what extent were the intended objective task demands subjectively realized by crews in training? and (3) What was the perceived training value of the scenario? There are several groups who possess information which could be used to address these questions: LOFT scenario designers, LOFT instructors and LOFT students. The author presents the following suggestions for answering the suggested questions. For measuring the objective task demands, a rating system (e.g., the NASA-TLX) could be used which assesses workload of different task demands. After completing a LOFT session, students could fill out the LOFT Survey as well as a modified version of the NASA-TLX. This information could be integrated and compared to see how successful the scenario was as a training devise. This approach could provide answers to questions such as, "did the crew realize the task demands that were intended for the scenario?" "What was the combination of objective task demands and subjective workload which were rated highest by students and instructors for training value?" Consideration should also be given in these assessments for the fact that instructors may vary the scenario depending on the needs of the crew. These results could be used to make comparisons across fleets and organizations.

Gregorich, S. E., Helmreich, R. L., & Wilhelm, J. A. (1990). The structure of cockpit management attitudes. *Journal of Applied Psychology, 75*(6), 682-690.

Article type: Theoretical
Keywords: Attitudes; Coordination

These authors note that personality characteristics are an appropriate basis for pilot selection, but being resistant to change, they are not a fruitful target of training. Attitudes, in contrast, are less resistant to change and may be altered through appropriate intervention. They believe that the Cockpit Management Attitudes Questionnaire (CMAQ) is an objective index of crew resource management (CRM) attitudes. The objectives of the current study were to: (1) identify a consistent internal structure of CMAQ; and (2) examine attitudes across three organizations and a pre-and-post CRM training situation. The authors discuss two primary *exogenous* influences on CRM attitudes -- organization effects (i.e., policies & procedures and informal norms); and (2) history effects (i.e., media coverage, personal study, or general infiltration of concepts into the aviation community). They argue that this history effect has been the main reason researchers have been unable to find a stable structure of the CMAQ; that is, attitude convergence about

CRM has caused some items to lose their informational value (or discriminatory power). In this study, they collected attitudinal data (via the CMAQ) from three samples with sample sizes of 380, 3836 and 725 flight crewmembers. Prior to factor analysis, the items were weighted by their informational value within each organization and timeframe. Gregorich et al. identified a 4-factor solution, that fit the data from each of the samples, respectively: (1) Communication & Coordination (interpersonal awareness, communication, crew coordination); (2) Command Responsibility (inappropriateness of dominant Captain); (3) Recognition of Stressor Effects (imperviousness to stress); and (4) No consistent set of items across samples but generally related to avoidance of interpersonal conflict. The authors also note that a pre/post-test with sample #3 found that CRM training had a significant impact on each of the first three factors. The authors suggest that attitude change indirectly implies performance improvement, but note that data to assess the relationship is scarce thus far.

Gregorich, S. E., & Wilhelm, J. A. (1993). Crew resource management training assessment. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 173-198). San Diego, CA: Academic Press, Inc.

Article type: Review article
Keywords: Attitudes; CRM; Training Evaluation

Crew resource management (CRM) training programs have been in place for the last decade, but evaluations of these programs are still in the early stages. Evaluation in this context refers to all aspects of the training program; not just the evaluation of a pilot's performance for certification. There has been some assessment of training programs in the past and these have met with some success, but additional training evaluation is necessary for three main reasons: (1) the training takes place in a dynamic environment, changes are subsequently incorporated in the various programs and the impact of these changes needs to be assessed; (2) training programs can still be improved upon and assessment provides a strong basis on which to design these improved programs; (3) a single training program can have different effects on individuals and organizations and these issues need to be considered. There are primarily two modes of assessment. The first is evaluation of training program outcomes. This concerns how well a training program achieved its intended results. These can be assessed at any time, but if a delayed assessment is used then information concerning how well training transferred can be obtained. Assessment should occur at all levels of the organization: student, crew, session, fleet and organization. The second mode is the evaluation of training program elements. This focuses on the training materials themselves and how they can be improved. It concentrates on the "how?" and "why?" of training program outcomes. It is difficult to develop absolute standards to compare performance against. Thus, the effects of CRM training are compared against prior levels or a comparison group. Questionnaires have been used extensively by many different organizations for assessment. The authors go on to provide a framework for assessment criteria. There are four potential targets: (1) outcomes; (2) curriculum; (3) instruction; and (4) individual and organizational characteristics. Information about these targets will, in turn, come from three different sources: (1) students; (2) instructors; and (3) auxiliary evaluators. All these different types of criteria need to be linked in some manner and studied more systematically. First an

initial assessment of student and organizational characteristics should identify the areas to be addressed by the program. Next, reactions and evaluations of the program by students and evaluators can provide useful feedback. Although, to be truly useful, this information needs to be linked to outcome data. Once relationships are established between them, then changes and recommendations will be more highly validated. This can be done using a cross-sectional design. Evaluation of training outcomes should be conducted using a longitudinal design. Since training outcomes are affected by individual and organizational characteristics, these need to be linked to training program outcomes if the relationships are to be studied. Furthermore, without this type of information, it will be difficult to know if the information learned in the program transfers to line operations.

Grob, N. (1979). Captains' training at Swissair. In G. E. Cooper, M. D. White, & J. L. Lauber (Eds.), *Resource management on the flight deck, Proceedings of a NASA/Industry Workshop* (pp. 97-106). San Francisco, CA: National Aeronautics and Space Administration.

Article type: Anecdotal/Descriptive
Keywords: CRM; Training

At Swissair, a person flies about twelve years as a co-pilot before upgrading and this upgrade training concentrates on three main areas: theoretical knowledge, flying skill or ability and management. About 24 Captains go through the course every year and of these, 15 percent typically fail. Unsuccessful pilots are given a second chance after one year. After that, if they don't pass, there are no more chances. The most frequently cited reasons for failure include: poor flying aptitudes, mental inflexibility and meager leadership qualities. An outline of the Captain's course is provided in the paper, this is then followed by brief descriptions of each of the different phases in the program.

Haakonson, N. H. (1980, September). Investigation of life change as a contributing factor in aircraft accidents: A prospectus. *Aviation, Space and Environmental Medicine*, *51*(9, Section (2), 981-988.

Article type: Theoretical
Keywords: Accidents; Anxiety; Depression; Stress

The author posits that many accidents are caused by an imbalance between performance ability and performance demand; specifically, this deficit is seen as being caused by some stressor (in this case the stressor is recent life change). The author notes that 47 percent of all human failures fall into the category of "unprofessional attitude or behavior" and sees recent life change as playing a major role in these failures. The paper presents a model whereby distress is caused by prolonged exposure to some stress acting on a person. If the person is unable to deal effectively with this stress, distress is the result and eventually an accident is likely to occur. The reason why this author chose life change as one possible source of this stress is because much research has accumulated dealing with this issue. He specifically advocates using the Holmes and Rahe (1974)

Recent Life Changes questionnaire (RLCQ) to measure "accident proneness" that may be due to stress from life changes. The RLCQ is a 55-item questionnaire that asks subjects to indicate how recently they have experienced changes with respect to several different aspects of their life (i.e., health, work, home/family, personal/social, financial). The author also reviews some past research that examines the link between personality/attitudes and aircraft accidents.

Hackman, J. R. (1987). Group-level issues in the design and training of cockpit crews. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 23-39). San Francisco, CA.

Article type: Theoretical
Keywords: CRM; Training

The author advocates a change of mindset, focusing on the crew not as a group of individual pilots, but as a team, a whole performing unit. He discusses three questions which help to determine crew effectiveness: (1) Does the performance of the crew fully meet (or exceed) the standards and expectations of others who have a legitimate stake in how the crew performs? (2) As crewmembers gain experience with one another over time, do they become increasingly expert in working as a team? (3) Does the experience of being in the crew contribute positively to the personal learning and satisfaction of each individual member? The author also describes four critical times in the crew life cycle. (1) *Pre-arrival or the "Shell"* which is before the crew meets. It consists of things that are accepted as givens by crewmembers or that they assume to be true without questioning. (2) *Team Creation*, where people now occupy the shell in a real working environment. This is an important area for crew resource management (CRM) training, because a briefing can have lasting effects on the crew. (3) *Task execution* -- CRM training will not show up here in the short run because in high stress situations, people will revert to well-learned habits. Only after the new responses become the dominant responses, will the training have a more overarching effect. Captains should take advantage of down times to fine tune and build the team. (4) *Team Termination*, which could provide a chance for the team to explore what they have learned in their time together.

Hackman, J. R. (1993). Teams, leaders and organizations: New directions for crew-oriented flight training. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 47-69). San Diego, CA: Academic Press.

Article type: Review
Keywords: CRM; Leadership; Pilot Selection; Training

This chapter is based on research conducted with intact crews from the time they first meet until the time the team is disbanded. The chapter is structured around three facts of cockpit crews: (1) cockpit crews are teams; (2) the Captain is the team leader; and (3) the crew is very enmeshed within the organizational, technological and regulatory contexts in which they operate. Crews are affected by how well the pilot can function as a team member. Not everyone has the skills that

are necessary to effectively work in a team, therefore, organizations should select out those individuals who can not effectively work in a team. Finally, airlines need to pay more attention to how they schedule crews. Time is needed for a crew to develop a rapport with one another and an organization needs to provide crews with the time that is necessary for this to occur. Hackman believes that past programs in crew resource management (CRM) training have focused on changing the attitudes and behavior of individuals, instead of focusing on how to train people to work more effectively in teams. According to the author, the most important components of training include significant amounts of hand-on practice, feedback and reinforcement in using team skills. Currently, crew scheduling decisions are not made in a manner that is conducive to facilitating teamwork. It takes a crew time to learn how to more efficiently and effectively work together. There are several things that can be accomplished in training to give Captains the skills they need to be more effective leaders. Captains can be shown that there are many times that can be used to build the team, not just those times of emergency. Furthermore, Captains can be taught the skills needed to take advantage of those down times. They can learn how to effectively build a team and how to keep it going. Use of line-oriented flight training (LOFT) can provide a safe environment for Captains to practice these skills. Hackman states that there are two things organizations can do to increase the probability that these skills will actually be used on the line. The first is the direction senior management communicates to the Captains about what is expected of them and their crews. Second, is the amount of latitude Captains are given in order to achieve those directions. Crews are also affected by the organizational context in which they must work. If the organizational policies do not support what is learned in CRM training, then the training will be for naught. According to the author, there are two factors which can either support or inhibit what is taught in CRM training, the: (1) consequences of performance; and (2) information and material resources needed to do their work.

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

Article type: Review
Keywords: Civilian; Communication; Coordination; Training Program Development

The authors discuss conceptual and methodological issues that arise when examining the performance of work groups within an organizational context (with a focus on air transport crews). They note that there is an assumption that technically proficient aircrew personnel will also form effective teams. For this article, the word *team* is meant to describe: (1) *real* groups (intact social systems); (2) groups that have one or more tasks to perform; and (3) groups that operate within an organizational context. An airline crew typically consists of a Captain, first officer (co-pilot) and second officer (flight engineer), as well as flight attendants. Individual crewmembers "bid" for sets of flights (trips), thus, any one crew composition is typically short-lived (usually about one month). The actual tasks performed by the crew revolve around five general purposes: (1) planning and decision-making; (2) manipulating the flight controls; (3) completing paperwork; (4) monitoring and adjusting mechanical and electrical systems; and (5) communicating. The authors note that currently, crew assessment is only made through the use of

proficiency checks and line checks and these methods typically evaluate an individual's proficiency. Little emphasis is placed on performance of *the crew as a whole*. Several issues that are problematic to the collection of "objective" performance measures are noted (e.g., hard data [accidents] typically become available very rarely and are often very incomplete). As an alternative they recommend utilizing "process criteria," this typically takes the form of check airmen's observations and evaluations of the performance *process* of the crew as they are working. They also make the observation that many events that affect crew competence occur outside of the cockpit (e.g., at dinner during overnight trips) and that these noncockpit periods of time also need to be factored in to assessment methodologies.

Halliday, J. T., Biegalski, C. S., & Inzana, A. (1987). CRM training in the 349th military airlift wing. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 132-134). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: CRM; Military; Training; Training Program Evaluation

These authors have developed their own crew resource management (CRM) program, termed, Aircrew Resource Management (ARM), to emphasize the use of all resources of the aircraft, especially the loadmasters. It is a seminar-based program run by two facilitators and reinforced by LOFT sessions. The Phase I seminar consists of a typical ten-member C-5 crew - three pilots, three flight engineers and four loadmasters. It is a nine-hour training day with no breaks (fatigue is part of the course design). Three learning objectives are emphasized: (1) synergy, the concept; (2) the synergy graph - a common language; and (3) the synergy formula - the heart of their program. The authors feel that they can attribute about 12 "saves" as a result of this program. The synergy formula is an inflight tool for problem-solving and decision-making in the cockpit. Most times in the cockpit, problems require group decision-making. Participants learn the formula and use it in role-play exercises. The formula is given to the students on a three by five index card which they keep at the completion of the program. The synergy formula is comprised of five different components: (1) questioning; (2) promoting; (3) decision; (4) how're we doin?; and (5) repeat. A preliminary evaluation of the program has been completed. The authors have surveyed over 250 crewmembers and found that students show a positive attitude towards the program. Those who have not gone through the training, but who worked with those who did, were also surveyed. Of these individuals, 80 percent felt that they had observed better coordination and flightdeck atmosphere from those who had gone through training.

Halpin, A. W. (1954). The leadership behavior and combat performance of airplane commanders. *Journal of Abnormal and Social Psychology, 49*, 19-22.

Article type: Validation
Keywords: Combat/Fighter Pilots; Leadership; Military; Performance Ratings; Satisfaction

This study examined the relationship between crew perceptions of leadership ability of their aircraft commander (AC); ratings of ACs performance made by supervisors and crew's satisfaction with leadership ability of their AC. Subjects were 52 B-29 bomber crews. AC leader behaviors were measured using the Leader Behavior Description Questionnaire (LBDQ; adapted for use by the military). Factor analysis was used to develop an empirical key for the LBDQ. For this study, "factor scores" on Initiating Structure-In-Interaction and Consideration were used as measures of leader behaviors. The ACs were also rated by squadron and wing commanders with respect to seven criteria of combat performance. Crew satisfaction was measured by asking crewmembers to indicate which crewmembers from among their crew they would choose to serve with if they had their choice. Criterion data were collected under both training and actual combat conditions. For both situations, there was a trend toward negative correlations between superiors' ratings and scores on the Consideration dimension and positive correlations between superiors' ratings and scores on the Initiating Structure dimension (both for zero-order validities and partial correlations). Conversely, for the crew indicators of satisfaction with leader behaviors, scores on the Initiating Structure dimension were negatively correlated with crew satisfaction and scores on the Consideration dimension were positively correlated with crew satisfaction.

Harper, C. R., Kidera, G. J., & Cullen, J. F. (1971). Study of simulated airline pilot incapacitation: Phase II, subtle or partial loss of function. *Aerospace Medicine, 42,* 946-948.

Article type: Group Comparison Study
Keywords: Communication; Coordination; Situational Awareness; Situational Judgment; Training Program Development

The authors maintain that one of the most potentially dangerous conditions that exist in high performance multi-engine aircraft is when the Captain is experiencing partial functioning of the cerebral cortex. They also note that this is extremely critical during phases of the flight in which other crewmembers have less time to react to the decrease in functioning of the Captain (e.g., during takeoffs and landings). This study was conducted in two parts using a flight simulator. The first part examined crew communication of 36 B-727 flight crews during instrument/system failure, while in the second part, 18 of 36 Captains were privately told to cease functioning somewhere during the landing approach, while still appearing awake and keeping their hands on the stick. The other 18 Captains were also given the same instructions, except in this condition the crew was lectured on the importance of watching for signs of incapacitation of the Captain. For the first study, most of the crews tended to respond effectively to the instrument\system failures. However, in the second part of the study, 25 percent of the "surprise" group crashed and

even for those that did not crash, the range of time elapsed before detection of the Captain's condition ranged from 30 seconds to 4 minutes (mean time = 1.5 minutes). None of the planes crashed in which the crew had been briefed to "be alert to the possibility of a subtle type loss of function." The authors concluded that the simulator provides an ideal environment to develop flight crews' awareness of these types of problems and that continued education along these lines will stimulate more meaningful crew communication and enhance the "crew concept."

Helmreich, R. L. (1979). Social psychology on the flight deck. In G. E. Cooper, M. D. White, & J. L. Lauber (Eds.), *Resource management on the flight deck, Proceedings of a NASA/Industry Workshop* (pp. 17-30). San Francisco, CA: National Aeronautics and Space Administration.

Article type: Theoretical
Keywords: Communication; Leadership; Personality

The author recommends applying social psychological research findings relevant to such areas as leadership, group relations and communication patterns to the crew performance domain. He first discusses why this research has not been heavily utilized in the past. The use of personality assessment with air crews has been more concerned with screening out candidates than finding the most optimal characteristics. Another reason cited is the fact that researchers have not focused on the *interaction* between personality and situational factors. Next, those personality and situational variables which may be of interest to aviation were presented. Two dimensions have been found to be related to group and individual reactions. The first was called *instrumentality* or goal orientation and the second was called *expressivity* or group orientation. During individual performance, a highly goal-oriented person might perform best and expressivity would not be important. In group performance situations, however, it might be optimal to have a Captain high on both dimensions. The social environment we are interested in is also very important because it encompasses more people than just those on the flight deck. The Captain can become overburdened if trying to fly the plane and solve all the problems simultaneously and the quality of decisions may be marginalized. Crew members may also rely or be more dependent on the Captain during times of emergency. Another factor in emergency situations is the management of communications. The Captain must monitor communication and make sure there are no major breakdowns. Research on group cohesiveness may be better focused on training to facilitate group performance. According to the author, more data needs to be collected during overload and emergency situations in order to examine the effects of social factors on how people respond during these periods of high-stress. Next, training programs can then be developed to help crews deal with these situations effectively.

Helmreich, R. L. (1987). Theory underlying CRM training: Psychological issues in flight crew performance and crew coordination. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 15-22). San Francisco, CA.

Article type: Theoretical
Keywords: Coordination; CRM

This article describes the theory behind crew resource management (CRM) training. Helmreich notes that performance is a global concept, defined as the total effectiveness of an individual or crew in achieving the goals of safe and efficient flight operations. Superior performance, both at the individual and crew level has two distinct components; first, technical proficiency and competence and second, resource management or crew coordination. These two components of performance tend to be relatively independent of one another and Check Airmen display a high level of reliability in evaluating both parts of performance. The author suggests that personality may play a larger role in flight deck performance than has been found in the past due to the "Honeymoon effect" where personality becomes a predictor only after a period of adjustment on the job. Performance is determined by ability, personality and attitudes. The first two are very difficult to change, thus training has focused on changing attitudes. The author recommends that: (1) the training program should provide involvement for the participants, such as LOFT with videotape feedback; (2) time should be taken to train instructors, then monitor and evaluate their effectiveness; and (3) CRM training should be more than a one-shot deal. It should have support of all members of the organization and be incorporated in the total training and checking process.

Helmreich, R. L. (1991a). Strategies for the study of flightcrew behavior. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 338-343). Columbus: Ohio State University.

Article type: Theoretical
Keywords: Communication; CRM

Helmreich uses the term *organizational shell* to describe the context in which air crews operate. In order for research findings using air crews to generalize, they must be conducted in many different organizations under different conditions in order to account for the influences of the organizational shell. Some studies have examined the communication patterns of crews to determine how crews interact during times of high stress. A second source of information has been surveys of crewmember attitudes about crew resource management (CRM) using the cockpit management attitudes questionnaire (CMAQ). CMAQ scores have been validated as predictors of crew performance and results have also shown differences in CMAQ scores both between and within organizations. A third source of information regarding CRM has been check airmen and instructor ratings of crew behavior using the Line/LOS Checklist. This checklist consists of ten scales and behavioral markers that exemplify CRM concepts. Finally, crews' reactions to the training have been measured using the *Loft Survey*. Even though the field of CRM training has made great strides in recent years, there are still many additional issues that

need to be explored. For example, longitudinal data is just now starting to be collected to allow for the examination of the long-term effects of training. Increasing automation on the flight deck will also need to be incorporated into CRM training, Finally, cross-cultural investigations of performance and crew reactions to training need to be undertaken.

Helmreich, R. L. (1991b). The long and short term impact of crew resource management training. In *Proceedings of the AIAA/NASA/FAA/HFS conference, challenges in aviation human factors: The national plan.* Vienna, VA, January, 1991.

Article type: Anecdotal/Descriptive
Keywords: CRM; LOFT; Training; Training Evaluation

The author notes that most crew resource management (CRM) programs were developed in the 1980s. The next step, then, is to develop criteria to be used in measuring the effects of these training programs. Frequency of accidents is not an appropriate measure because the base rate of these incidents is so low. Helmreich describes a measure that assesses the attitudes of participants directly after going through training. Over 15,000 crewmembers have done so at this time. A second measure of attitudes is the Cockpit Management Attitudes Questionnaire (CMAQ). Crew members are given the CMAQ before and after CRM training to assess attitudes about crew performance. A third measure developed was the Line/LOS Checklist. Evaluators rate crews on ten aspects of crew behavior on a five point scale. A final approach has been analyzing crew communications from Cockpit Voice Recorder transcripts. All of this data supports the notion that CRM training is effective. Crews value the training and have shifted their attitudes and their behavior. Furthermore, crews under stress still perform in a manner consistent with the training. The training must become part of the organization, recurrent training and everyday line operations. The role of Check Airmen and Instructors is also critical and they should be provided additional training. How to deal with those crewmembers who reject the training is still a question to be addressed. LOFT is an important tool for the training, but scenarios need to be enhanced to include such variables as Air Traffic Control communications. Fleet differences have been found within organizations and the reasons for this still need to be explored.

Helmreich, R. L., Chidester, T. R., Foushee, H. C., Gregorich, S., & Wilhelm, J. A. (1989). *Critical issues in implementing and reinforcing cockpit resource management training* (NASA/UT Tech. Report No. 89-5). Moffett Field, CA: NASA Ames Research Center.

Article type: Review
Keywords: Attitudes; Training; Simulation

The authors' purpose in this paper is to "attempt to summarize the state of current knowledge about how to maximize the impact of crew resource management (CRM) training without statistical tables and psychological jargon." Twelve recommendations for optimizing CRM training are presented: (1) demonstrate total commitment to the program (why is CRM important); (2) communicate the nature of the program and what it does and does not do; (3)

assess the status of the organization before designing/initiating the program; (4) customize the training to reflect the nature and needs of the organization; (5) use course activities that actively engage participants; (6) recognize that the initial CRM course is just the first step (the FAA circular calls this the *awareness phase*, additional training and reinforcement must follow); (7) integrate CRM training with LOFT (authors believe LOFT is currently the most effective way to reinforce CRM concepts); (8) stress the critical role of check airmen and instructors; (10) institute quality control procedures to assure that CRM and LOFT programs are accomplished effectively; (11) consider extending training to cabin crews and other aviation jobs; and (12) make provisions for dealing with CRM "failures" (individuals who refuse to accept the CRM approach can undermine the credibility of the training and authors recommend peer pressure, counseling and/or more training to overcome this).

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

Article type: Review article
Keywords: CRM; training; group process; teamwork; communication; CMAQ

Crew resource management (CRM) has expanded from being cockpit resource management to crew resource management in the past few years. According to the authors, "CRM includes optimizing not only the person-machine interface and the acquisition of timely, appropriate information, but also interpersonal activities including leadership, effective team formation and maintenance, problem-solving, decision-making and maintaining situation awareness." In order to discuss group behavior, the authors break the group process down into three factors: input factors, outcome factors and group process factors. The interaction of these three determine the group's performance. Probably the least understood and studied component are the group process factors. Many have tended to treat them somewhat as a black box. The way that many of these processes are manifested is through communication. Simulators can be a great source to study these communications (e.g., Ruffell Smith, 1979). Decision-making is another factor that has been studied by researchers (e.g., Orasanu, 1991; Chidester et al., 1990) to determine its effects on crew interactions. Patterns of interactions vary between higher performing and lower performing crews. Group processes have been further broken down into two categories: interpersonal and cognitive functions and machine interface tasks. Effective flight operations will depend first on whether the crew can execute the proper procedures. These separate categories need to come together at some point to produce overall effective performance. The authors also discuss ways to optimize the input and group process factors. One short term strategy to enhance the individual factors would be to provide training to the existing workforce, while a long term strategy would be to select personnel who had the qualities which would contribute to effective team performance. Organizations can do several things to increase crew effectiveness. They can make sure that training, policies and procedures are congruent with one another. Checklists and cockpit documents should be consistent with the crew concept. Communication between the cockpit crew and other groups should be discussed as well. On the regulatory front, the Federal

Aviation Administration released an Advisory Circular (1989) which outlined the concept of CRM and its issues. A brief history of CRM is also presented by the authors.

Helmreich, R. L., Foushee, H. C., Benson, R., & Russini, W. (1986). Cockpit resource management: Exploring the attitude-performance linkage. *Aviation, Space and Environmental Medicine, 57*, 1198-1200.

Article type: Group Comparison Study
Keywords: Attitudes; Coordination; Decision-making

The objective of the research reported here is to demonstrate an *indirect* approach to assessing the potential effectiveness of crew resource management (CRM) training through assessment of attitudes. The respondents were 658 pilots (3 commercial airlines); the raters were five check airmen experienced in evaluating flight crew performance. The respondents completed the Cockpit Management Attitudes Questionnaire (CMAQ), containing 25 items reflecting factors related to effective crew coordination. The raters rated pilots on overall flightdeck management skills. A discriminant analysis was used to contrast the attitudes of pilots judged as superior with those *rated* below average. The results suggested there is a direct linkage between self-reported attitudes and independent evaluations of performance. Characteristics of effective flightdeck managers include: (1) recognizing personal limitations and diminished decision-making in emergencies and encourages other crewmembers to question decisions and actions; (2) being sensitive to personal problems of other crewmembers that might effect operations; (3) recognizing the need for the pilot flying to verbalize plans and importance of the Captain's role in training other crewmembers; (4) recognizing the need for a harmonious and relaxed flightdeck.

Helmreich, R. L., Merritt, A. C., Sherman, P. J., Gregorich, S. E., & Wiener, E. L. (1993). *Flight management attitudes questionnaire (FMAQ)* (NASA/UT/FAA Tech. Report No. 93-4).

Article type: Training Manual
Keywords: CRM

The Flight Management Attitudes Questionnaire (FMAQ) is a revised and extended version of its predecessor, the Cockpit Management Attitudes Questionnaire (CMAQ). Items were added from Hofstede's Values Survey Module, which include the four cultural dimensions of: (1) Power Distance; (2) Uncertainty Avoidance; (3) Masculinity-Femininity; and (4) Individualism-Collectivism. Further additions include the work of Bond at the Chinese Culture Connection and Hall's monochronic and polychronic concepts of time. Nineteen of the original 25 items from the CMAQ were retained for this new instrument and another 63 items were added. The Attitudes toward Automation scale was completely revised to reflect four aspects of automation: (1) preference for and comfort with automation; (2) recognition of possible increases in communication and workload due to automation; (3) when and when not to use automation; and (4) recognition of possible degradation in technical skills.

Helmreich, R. L., Wiener, E. L., & Kanki, B. G. (1993). The future of crew resource management in the cockpit and elsewhere. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 479-501). San Diego, CA: Academic Press, Inc.

Article type: Review article
Keywords: CRM; Training;

The concept of crew resource management (CRM) has become widely accepted and is now being integrated into most flight training programs. The next challenge is refinement. People believe in the concept, but we must determine if there are better ways of teaching it. Five factors are identified which are seen as having implications for CRM programs: (1) the development of mega-carriers; (2) an increase in flightdeck automation; (3) the FAA's AQP; (4) the advent of manufacturers providing training; and (5) legal liability issues if a CRM program is *not* in place. The authors discuss several other issues also relevant for CRM. The first is that training programs are now including resource management in the early stages of training. By doing this, positive habits are learned from the start and only need to be reinforced. In present training programs, however, CRM is separate from technical instruction. This approach will need to be changed. CRM concepts will need to be integrated into all aspects of instruction. A challenge to the industry is to develop training that can integrate both the technical aspects of flying and CRM. Second, a continuing problem are those few individuals who reject CRM training. These people can be a significant threat to safety and the credibility of the organization if they are left to continue acting the way they have in the past with no repercussions. Selection techniques will need to identify those people who have the attributes associated with effective team performance. These new strategies will also need to be dynamic to take into account the changing technology of the industry. Third, the NTSB states in a recent report that there also needs to be some joint cockpit and cabin crew training. For some airlines, such training is not feasible because of the costs associated with bringing all crewmembers together for training. Others such as America West have already trained flightdeck and cabin crews together. This organization also has a flight attendant present in the simulator and plays a role dictated by a script. These line-oriented flight training (LOFT) sessions are even more realistic. Southwest Airlines has developed an extended crew training for new Captains. This is to prepare the new Captain for management responsibilities. It is hard to determine what the future will bring in training in this area. Research is just beginning in the air traffic control setting even though the parallels with the cockpit crew are plentiful. There is also interest in investigating controller-cockpit communications, as well as individual controller performance.

Helmreich, R. L., & Wilhelm, J. A. (1986). *Evaluating CRM training: I. Measures and methodology* (NASA/UT Tech. Report 86-8). Moffett Field, CA: NASA Ames Research Center.

Article type: Review
Keywords: Attitudes; Performance; Personality; Training

This article summarizes the program of research being conducted by Helmreich and associates in conjunction with NASA Ames Research Center. As part of this program of research, the authors

are collecting demographic information, attitudinal information (Helmreich's 25-item Cockpit Management Attitudes Questionnaire [CMAQ]), administering a personality battery (includes subscales measuring achievement motivation and Type A behavior patterns), LOFT/LINE check performance evaluation (measures individual and team performance, data on failed proficiency checks and archival data on incidents (where available). Preliminary analyses suggest that relationships exist between personality and performance, attitudes and performance, attitudes and CRM-induced attitude change and attitude differences (about CRM) among positions within an organization and between organizations.

Helmreich, R. L., & Wilhelm, J. A. (1989). When training boomerangs: Negative outcomes associated with Cockpit Resource Management programs. In R. S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 16-21). Columbus, OH: Ohio State University.

Article type: Training Development and/or Evaluation
Keywords: Attitudes; Personality; Training Evaluation; Training Program Design

This study compares pre- and post-crew resource management (CRM) training attitudes toward flight deck management in a large sample of airline pilots ($N = 5000$). The authors report that the majority of CRM training participants indicate highly significant attitude change (as measured by the Cockpit Management Attitudes Questionnaire (CMAQ) with respect to crew coordination and personal capabilities. However, there is a small subset of participants that react negatively to the training and show *boomerangs* (negative change) in attitude. In addition, there was also substantial *within* training seminar (i.e., for one instructor) variability. The CMAQ consists of 25 items summed into one of three composites: Communication and coordination, command responsibility, and recognition of stressor effects. Past research has identified three distinct subgroups of pilots from cluster analyses of personality data. The first cluster was termed *Right Stuff* and consisted of pilots who score high on both instrumental and expressive personality traits. The second cluster was named the *Wrong Stuff* and consisted of pilots who demonstrated high levels of negative instrumental traits (i.e., authoritarianism) and low positive expressivity. The final cluster was named the *No Stuff* and consisted of pilots who scored low in both the instrumental and expressive domains. Helmreich and colleagues demonstrated that the *No Stuff* pilots were the ones who accounted for the net boomerang effect. They also demonstrated that there was a much higher incidence of the boomerang effect in "bad" seminars (as evaluated by participants) with more than 30 percent showing negative change. The authors conclude that particular events in a CRM training group trigger particular patterns of reactions and that the dynamics of group processes drive attitudes and evaluation in one direction or another, in addition to personality and attitudinal characteristics of the group. They also advocate *selecting in* pilot candidates, in part, with personality characteristics important for success with respect to the CRM aspects of flying. In addition, they direct organizations to face the issue of what to do with CRM failures, defined as those pilots unwilling to accept CRM concepts.

Helmreich, R. L., & Wilhelm, J. A. (1991). Outcomes of crew resource management training. *International Journal of Aviation Psychology, 1,* 287-300.

Article type: Training Development and/or Evaluation
Keywords: Attitudes; Personality; Training Evaluation; Training Program Design

Helmreich and Wilhelm describe the multi-day crew resource management (CRM) course as providing experiential learning in interpersonal communication, decision-making, conflict resolution, stress/stress management and leadership. They note that belief in the effectiveness of this training is widespread, even though there is little empirical evidence to support this position. This report presents data collected from two major airlines ($N > 15{,}000$) using a CRM training evaluation questionnaire that assesses perceived usefulness of the training, both for self and others, in addition to data collected using the Cockpit Management Attitudes Questionnaire (CMAQ). The CMAQ contains 25 statements describing attitudes found to be indicators of crew effectiveness. Data show that even in these two major airlines, with independently developed CRM programs, most participants reported that the training was important and useful. They also show that attitudes demonstrated significant positive change from pre-training to post-training. However, despite these highly significant findings, there were some individuals who demonstrated either no attitude change or negative attitude change (boomerang effect). Helmreich et al. also clustered participants into one of three personality clusters and examined the attitude change from pre- to post-training separately for these types. Type 1 was termed *right stuff* and was characterized by those pilots who showed high levels of both instrumental and expressive traits, Type 2 was termed *wrong stuff* and was characterized by high levels of negative instrumental and low levels of positive expressive traits. Finally, they found that those subjects who were classified as *no stuff* (i.e., scored low in both expressive and instrumental personality domains) were the subjects who demonstrated this boomerang effect. The authors note that it is disturbing that those pilots who stand to benefit the most from this training are the ones who are least affected in the intended manner.

Helmreich, R. L., Wilhelm, J. A., Gregorich, S. E., & Chidester, T. R. (1990). Preliminary results from the evaluation of cockpit resource management training: Performance ratings of flightcrews. *Aviation, Space and Environmental Medicine, 61,* 576-579.

Article type: Group Comparison Study
Keywords: Coordination; Performance Ratings; Simulation

This study describes the impact of crew resource management (CRM) training on aircrews. The authors used check airmen and line-oriented flight training (LOFT) instructors as raters of simulator performance. Raters received videotaped vignette training, then made global ratings of overall crew performance (859 crews were rated). A five-point scale was used, but because of limited use of 1st and 5th scalar points, data were revised to reflect a three-point scale (below average, average, above average). In fact, one-third of the ratings were discarded because of raters' failure to "observe variability in performance" or because of an extreme positivity bias. Chi-square test showed significant deviation of the CRM-trained group toward "above average,"

whereas, the untrained group placed a greater than expected percentage of the sample in the "below average" group. The LINE/LOFT worksheet that leads to overall rating uses 14 dimensions: advocacy, abnormal management, briefings, communication, conflict management, critique, decision-making, distractors, group concern, inquiry, preparation/planning, proficiency, task concern and vigilance.

Helmreich, R. L., Wilhelm, J. A., Kello, J. E., Taggart, W. R., & Butler, R. E. (1991). *Reinforcing and evaluating crew resource management: Evaluator/LOS instructor reference manual.* (NASA/UT Tech. Report No. 90-2). Moffett Field, CA: NASA Ames Research Center.

Article type: Training Manual
Keywords: Coordination; CRM; Training Evaluation

It has been determined that many aviation accidents and/or incidents have been directly caused by a lack of crew coordination. In response to this problem, many organizations have implemented crew resource management (CRM) training. Many of these programs utilize the Line Oriented Simulation (LOS) developed by Helmreich and colleagues. There are eight specific crew effectiveness markers and two global ratings, one of overall technical performance and overall crew performance that are utilized as part of the LOS checklist. The markers are placed into one of four clusters: (1) communications processes and decision behavior; (2) team building and maintenance; (3) workload management and situational awareness; and (4) overall technical proficiency. For each concept, a series of behavioral indicators are provided. Ratings are made on a five-point scale rating from poor performance to excellent performance. If a rating of one or five is given, then comments describing why the rating was given should be provided. An example of the form and the behavioral indicators are provided in the manual. In order for these programs to be effective, however, all members of the organization must commit to them, especially the evaluators, instructors and check airmen. These people are in the key positions to teach and reinforce the CRM concepts. They also serve as role models and the primary source of information on the effectiveness of the programs. They need special training and this manual is only one means of helping them to evaluate CRM behaviors. There are currently two barriers to obtaining data on crew behavior. First, most evaluations of pilot performance is mostly completed using pass/fail grading criteria. Second, evaluation done on an individual basis have focused on technical performance. In addition, guidelines for evaluating and debriefing crew coordination have been lacking. Also, most of this past research has focused on accidents or incidents. Positive examples of crew behavior also need to be studied to understand how *effective* crews perform, that is, data needs to be collected at the group level.

Holdstock, L. F. J. (1979). British Airways' pre-command training program. In G. E. Cooper, M. D. White, & J. L. Lauber (Eds.), *Resource management on the flight deck, Proceedings of a NASA/Industry Workshop* (pp. 76-86). San Francisco, CA: National Aeronautics and Space Administration.

Article type: Anecdotal/Descriptive
Keywords: CRM; Training Program Design

British Airways utilizes a four stage process in training pilots for command: (1) an initial command potential assessment; (2) a pre-command management studies course; (3) a pre-command course; and (4) a command course. The initial command potential assessment takes place about three years after a pilot joins the fleet. A preliminary assessment is made as to whether this person will receive a command or not. After about eight years, a co-pilot is upgraded to senior first officer and a final assessment is made at that time. It takes on average about twelve years to be given a command. If a satisfactory assessment is made, the pilot progresses normally. On the other hand, if an unsatisfactory assessment is made, the pilot is informed. The pilot must request training to try and change this assessment, otherwise no training will be given. About two years before the pilot is to take command, he is given a management course which lasts for two weeks (this course is briefly described).

Hörmann, H-J., & Maschke, P. (1991). Exogenous and endogenous determinants of cockpit management attitudes. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 384-390).

Article type: Validation
Keywords: Attitudes; Personality

The purpose of this study was to examine the relationships between attitude factors, personality traits, flight experience and performance criteria. Two measures were administered to applicants for a German charter airline: Temperament Structure Scales (TSS) and Cockpit Management Attitudes Questionnaire (CMAQ). Some took the inventories in English, while others received German versions. A factor analysis conducted on the German version of the CMAQ found that the factor structure was similar to the English version. The authors also compared the mean responses on the three scales of the CMAQ between those who took it in English and those who took it in German and they found large intergroup differences. These differences may have been confounded by the position of the applicant, however, because both Captains and first officers completed the instrument. Some scales of the TSS were significantly correlated with the CMAQ. Preliminary data using performance criteria found that the CMAQ Command scale correlates significantly with the assessment of the supervision period. Flight experience is more highly related to the ratings made after the line check. Two scales from the TSS had significant negative correlations with the superior's rating: emotional instability and aggressiveness.

Hunt, G. J. F. (1991). Getting test items to measure knowledge at the level of complexity which licensing authorities desire: Another dimension to test validity. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 1169-1177). Columbus, OH: Ohio State University.

Article type: Test Development
Keywords: Civilian; Pilot Selection

The author notes that in the pursuit of test item validity, many aviation licensing test items have been developed that measure what they are supposed to measure, but not at the level of complexity at which it should be measured. Specifically, the author advocates a content-process approach to examining test item validity. That is, the extent to which an item matches the expected knowledge, or content type (e.g., declarative knowledge) with the required level of information processing is just as important for test validity as is that test item's correlation with some relevant criterion. One approach to content-process validity is to examine where a test item falls in a 4 (Content) X 3 (Thinking Process Level) matrix. The content of a test item consists of one of four different levels of information: facts, concepts, rules and cognitive procedures. The thinking process level that a student is required to use to solve a problem may be at one of three levels: remember, use, or hypothesize. The author sampled 150 test items selected from British, American and New Zealand commercial pilot license item banks and classified them with respect to where they best fit in the content X thinking Process Level matrix. He also used licensing personnel from both private and commercial licensing authorities to estimate the expected frequency distribution of items. The private licensing personnel predicted that most of the items would be from either "remember a concept" or "remember cognitive procedures," the commercial licensing personnel predicted that most of the items would be either "using rules" or "using cognitive procedures." Most of the examiners believed that aviation test items should mostly tap a student's ability to use rules and cognitive procedures, when in fact, most of the items sampled in this study (from all three countries) required students to *remember a fact, concept* or *rule.*

Hunt, L. M. (1991). Information processing in ab initio pilot training. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 1169-1177). Columbus, OH: Ohio State University.

Article type: Training Development and/or Evaluation
Keywords: Training Program Design

The author notes that in addition to specific technical flight training, pilot judgment may benefit from a broader context of training in information processing. If one examines common causes of poor pilot judgment, one can see that they often have an underlying information processing function, for example, making an inappropriate choice when the right choice was available. This article describes an information processing course that is taught as part of ab initio pilot training. Essentially, the course was designed to assist aviation students in "learning to learn." The seven course objectives were: (1) increased self-awareness of individual learning processes; (2) a more (cognitively) active role in classroom learning; (3) more control over and responsibility for

individual learning; (4) greater ability to evaluate and monitor individual learning performance; (5) a strategic use of cognitive strategies; (6) improved examination performance; (7) a deep holistic approach to learning rather than an atomistic, surface approach. As of this publication, the course was in the middle of its first term so a comprehensive assessment of its effects has yet to be completed. However, the author reports that students were able to identify a much higher number of strategies for remembering tasks and strategies for conducting deep and meaningful analysis of tasks than they were able to identify before beginning the information processing course.

Irwin, C. M. (1991). The impact of initial and recurrent cockpit resource management training on attitudes. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 344-355).

Article type: Single Group Study
Keywords: Attitudes; CRM; Training

In this study, the cockpit management attitudes questionnaire (CMAQ) was given to training participants from seven organizations, including U.S. and foreign carriers, as well as military personnel. The CMAQ was administered before and after initial crew resource management (CRM) training. It was then readministered before and after recurrent CRM training. Due to space limitations, the results of only one organization were reported. Positive attitude change was found on all three subscales of the CMAQ. Difference scores for each individual from before to after training were calculated for each subscale and placed the person into one of five groups: extreme negative change, negative change, no change, positive change and extreme positive change. Overall global attitude change groups, composed separately by *all negative*, *part negative*, *non-valenced*, *part positive* and *all positive*, were then formed. Overall, pilots had a positive reaction to initial training, but it was slightly lower at recurrent training. Some participants showed negative attitude change on one of the subscales, but few (3.2%) showed negative attitude change on all subscales. The same could be said for recurrent training, only the *all negative* attitudes increased somewhat to 5.2 percent and the *all positive* attitude decreased.

Jensen, R. S. (1982). Pilot judgment: Training and evaluation. *Human Factors, 24*, 61-73.

Article type: Theoretical
Keywords: Decision-Making; Training; Training Evaluation

This paper examines some of the decision-research literature and suggests training and evaluation strategies that could be employed for pilot judgment. The author notes that 80 - 85 percent of aircraft accidents can be assigned broadly to pilot error and almost 52 percent of these pilot errors are due to decision activities (data analyzed from the NTSB data base for the years 1970 - 1974). He defines judgment as the: (1) ability to search for and establish the relevance of all available information regarding a situation, to specify alternative courses of action and to determine expected outcomes from each alternatives; and (2) motivation to choose and

authoritatively execute a suitable course of action within the timeframe permitted by the situation. Potential decision-making training approaches are presented by the author, including: (1) classroom training in information integration and subjective probability estimation and training to anticipate decisions; (2) computer-assisted instruction; (3) flight simulation; and (4) situation training in the actual aircraft. Finally, the author discusses the importance of evaluating judgment capabilities before training, the effects of training on pilot judgment and the amount of training transferred to the operational flying environment.

Johnson, N. A., Shroyer, D. H., Grewe, J. B. (1991). A new generation of crew resource management training. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 668-673).

Article type: Training Development and/or Evaluation
Keywords: CRM; Training; Training Evaluation; Training Program Design

Hernandez Engineering, Inc. developed a crew resource management (CRM) training program for the U.S. Air Force Strategic Air Command based in part, on concepts suggested by the 1979 NASA/Industry workshop. This workshop suggested that current programs needed to be activity and skill oriented. This new program was designed to take advantage of both past experiences and current technology. Five initial decisions were made about the program: (1) the word cockpit would be replaced by crew; (2) a MOST session would be an integral part of the program; (3) the program would become part of the SAC command structure; (4) a multi-indicator evaluation process would be instituted; and (5) the goals would include operational effectiveness. The topics of the program were grouped into five categories: (1) crew communications; (2) situation awareness; (3) behavior styles; (4) stress management; and (5) mission management. The awareness training phase is delivered on Interactive Video Disc (IVD). The practice and feedback phase uses exercises which do not require a simulator. The facilitators are carefully selected and trained. Skills are practiced and critiqued. In the reinforcement phase, MOST sessions are designed to help further practice and reinforce the skills. Evaluations have been collected, but no details were given of the results at this time.

Johnston, A. N. (1987). Remedial training: Will CRM work for everyone? In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 108-118). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: CRM

The author discussed situations involving pilots who may be in need of remedial training and the link to Pilot Advisory Groups (PAGs). The PAG concept is one in which a peer group monitors and then reports any pilots who may be having difficulties. If reported, pilots are not subject to disciplinary action, instead they will receive treatment of some kind. Unfortunately, critiques of pilot performance tend to be very general and vague and training rarely attempts to specifically

address the problem behaviors. The author suggests that a new system of training needs to be developed in which the evaluative and checking elements are separate from the developmental or training elements of assessment. There needs to be a clear recognition of different skill requirements and their behavioral characteristics. New programs need to have peer involvement and be more sympathetic to the difficulties faced by the pilots.

Johnston, N. (1993). CRM: Cross-cultural perspectives. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 367-398). San Diego, CA: Academic Press, Inc.

Article type: Review article
Keywords: CRM; Foreign; Training

Countries all over the world share the common goal of safe and efficient flight, but there are frequently dramatic differences in how these goals are defined and achieved. Even the meaning of the words crew resource management (CRM) training may differ from one culture to the next depending upon the cultural and environmental context in which the person resides. It also may be difficult to translate some CRM-related terms because there are no equivalents in other languages (e.g., "followership"). There is a serious deficit with respect to cross-cultural research in general and this deficit becomes even larger when cross-cultural research in aviation is examined. CRM seeks to effectively utilize all available resources, these may vary from country to country, but crewmembers are always a resource. Communication and decision-making are also important and improving these things is a common objective. In order to make useful cross-cultural comparisons, the number of variables one needs to examine is staggering. There are different social processes at work, differences in the definitions of roles and how they are perceived, as well as variations in the applications of standard operating procedures (SOPs).

There are many different factors and issues addressed by CRM training. A question arises as to how these can all be considered across different cultural boundaries? The author presents work conducted by Hofstede, who examined work-related values across more than 50 countries. Hofstede found that cultures can be scaled and differentiated across four basic dimensions. The first dimension is power distance (PDI) which is how power is distributed between superiors and subordinates. In a high power-distance culture (e.g., India), superiors frequently make all the decisions and subordinates readily accept them. On the other hand, in a low power-distance culture (e.g., Austria), superiors and subordinates tend to treat each other as colleagues. Uncertainty avoidance (UAI) is the second dimension. This is how a culture views things such as novelty, ambiguity and uncertainty. High uncertainty avoidance cultures (e.g., Japan, Greece) are inflexible and have strict codes of conduct. In low uncertainty avoidance cultures (e.g., Denmark), people are much more tolerant and adaptable. The third dimension is individualism (IDV) which looks at the emphasis placed on individuals. In individualistic cultures (e.g., United States), individual initiative and achievement is highly valued. Alternatively, collectivist cultures (e.g., Iran) promote the collective interests of the group. The final dimension is masculinity (MAS). Ambition and performance are desired in masculine cultures (e.g., Italy, Australia) by being forceful. Feminine cultures tend to have a more ecological perspective. Sex roles are more

clearly defined in a masculine culture and men are expected to be dominant. The author applies these four dimensions and discusses what the implications could be for CRM training. Then, in the last part of the chapter, Johnston challenges the idea that CRM is appropriate and necessary around the world. Clearly, there are many underlying factors (besides CRM factors) which may cause differences in incidence rates from one culture to the next. Consequently, a "CRM solution" should not be implemented without exploring these and other issues very thoroughly.

Kahn, A. (1991). Behavioral analysis of management actions in aircraft accidents. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 674-378).

Article type: Theoretical
Keywords: Accidents; Decision-Making

This article looks at the distal causes of three aircraft accidents and one railway accident. The behaviors and decisions the pilots made were examined and many of them were not in line with safety. In addition, most of the poor decisions reviewed by the author had been reinforced by management in an attempt to keep costs down. Consequently, the pilots/engineer were reinforced for breaking the rules and/or not following appropriate procedures. Kahn concludes by stating that these cases indicated that the attitudes developed by managers are frequently powerful factors. These factors may convince individuals to ignore their better cognitive judgment and to proceed to commit dangerous acts which they would not otherwise commit.

Kanas, N. (1987). Psychological and interpersonal issues in space. *American Journal of Psychiatry, 144*(6), 703-709.

Article type: Review
Keywords: Interpersonal Skills; Personality

The author suggests that future manned space missions will require that crews become more heterogeneous, thus increasing the importance of psychological and interpersonal factors. A review of more than 60 U.S. and Soviet space simulation studies (and personal reports from space missions) identifies nine psychological issues (sleep problems; time sense disturbances; demographic effects; career motivation; reaction to isolation; transcendent experiences; postflight personality changes; psychosomatic symptoms; and anxious, depressive and psychotic reactions). Seven interpersonal issues were identified (interpersonal tensions; problems resulting from crew heterogeneity; anger displaced to outside personnel; need for dominance; decreased cohesiveness over time; task-neutral interactions; and types of leadership). The article suggests that future "space" research should focus on rapid screening methods that assess motivation and predict individuals that will be compatible crewmembers. Kanas also suggests the use of "sensitivity" training with crews before space flights.

Kanki, B. G., & Foushee, H. C. (1989). Communication as group process mediator of aircrew performance. *Aviation, Space and Environmental Medicine, 60*, 402-410.

Article type: Group Comparison Study
Keywords: Communication; Simulation

This study examined crew interaction as indexed by the communication process. The research objectives were to determine: (1) whether particular speech patterns could be found to differentiate crews who had just finished flying together from crews who had not flown together; and (2) relationships among the speech categories themselves. The subjects were 19 crews; 10 had flown together previously (FT); 9 had not flown together (NFT). The criteria were errors in performance, as measured by the Foushee and Manos method of voice data analysis (errors were analyzed in terms of severity) and categories of communication. The results suggest that crews in the FT condition tended to perform better than crews in the NFT condition and the relationship strengthened as severity of error increased. Five of the communication variables showed strong main effects. "statements of intent", "acknowledgments" and "total communication" were more frequent in FT crews; "tension release statements" and "non-task statements" were more common in NFT crews.

Kanki, B. G., Lozito, S., & Foushee, H. C. (1989). Communication indices of crew coordination. *Aviation, Space and Environmental Medicine, 60*, 56-60.

Article type: Group Comparison Study
Keywords: Civilian; Communication; Coordination; Simulation

The objective of the study was to examine the ways in which coordination problems are solved; that is, what differentiates flightcrews who are performing smoothly from those that are not. Using the voice reported content coding technique established by Foushee and Manos (1981) this study analyzed speech patterns in low and high error performances (5 of each; 10 2-person crews). Two main findings emerged: (1) low-error crews adopt consistent, standard form/pattern of communicating (high error groups showed diverse patterns); (2) groups who had more recently flown together showed more stable speech interaction patterns.

Kanki, B. G., & Palmer, M. T. (1993). Communication and crew resource management. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 99-136). San Diego, CA: Academic Press, Inc.

Article type: Review
Keywords: CRM; communication; situation awareness; leadership; personality

Illustrations of the importance of the link between communication and flight safety are clearly demonstrated in accident investigation reports completed by the National Transportation Safety Board (NTSB). They have been able to determine the causes of the crashes and have attributed

many of them to a breakdown in one of four crew communication links between: (1) the pilot and dispatchers; (2) the pilot and air traffic control (ATC); (3) within ATC flow control teams; and (4) within the cockpit crew. Recent research has investigated the effects of process variables and how they impact performance. Patterns of communication may be affected by the personality types of the crewmembers. Chidester, Kanki, Foushee, Dickinson, & Bowles (1990) and Kanki, Palmer, & Veinott (1991) have each conducted research in the area and found that higher performing crews have balanced two-way communication in the cockpit. The authors identify five ways that communication can affect crew performance: (1) communication provides information; (2) communication establishes interpersonal relationships; (3) communication establishes predictable behavior patterns; (4) communication maintains attention to task and monitoring; and (5) communication is a management tool. This, and other similar research programs have provided interesting findings, but the authors conclude the chapter by discussing some of the limitations of past research.

Kayten, P. J. (1993). The accident investigator's perspective. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 283-314). San Diego, CA: Academic Press.

Article type: Review
Keywords: Accidents; CRM

The primary method utilized by the aviation industry to identify and fix problems in air safety in the past was to look at the results of accident investigations. The National Transportation Safety Board (NTSB) has played a major role in recognizing the importance of crew resource management (CRM). This chapter discusses the events that have led the NTSB to focus on CRM training as a deterrent to other accidents and incidents. It was not until the report written about the United Airlines Flight 173 crash that the NTSB explicitly recognized the need for CRM-type training. The reason cited by the author for this was that up until this time there was not the "vocabulary, analytical models, or tools" to speak about these types of problems. The studies done by NASA and summarized by Cooper, White, & Lauber (1980) provided a framework for this type of research. The Ruffell Smith (1979) study began looking more directly at these issues. After these things became known, it was believed that a separate training issue was being discussed. In more recent years, the NTSB has recommended that CRM principles be incorporated into the standard operating procedures of the airlines and that the FAA should provide guidance to the airlines in doing so. Other accidents by smaller aircraft caused the NTSB to state that Part 135 carriers should also institute CRM training. Several positive examples were presented to show the effectiveness of CRM training. New directions for CRM training are also being explored. NTSB has identified several accidents in which there was a lack of coordination between the flight crew and others involved with the aircraft such as air traffic controllers or cabin crew. Thus, CRM training is expanding beyond the cockpit.

Klein, G. A., & Thordsen, M. L. (1991). Representing cockpit crew decision-making. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 1026-1031).

Article type: Group Comparison Study
Keywords: Decision-Making

This article examines how to observe a team making a decision. The authors attempted to develop a format for representing the decision-making of a team. Three crews were observed during a three leg simulator session. Their performance was videotaped, from which transcripts were made identifying all comments. It was hypothesized that team decision-making was similar to individual decision-making. The data was simplified into a few phases and coded into three different categories: (1) identifying the problem; (2) selecting a course of action; and (3) implementing the course of action. Crew comments were mapped according to a Problem Behavior Graph (PBG) approach. Examples are given by the authors. Results showed that different crews had different patterns of reactions to the same malfunctions.

Krey, N. C, & Rodgers, D. (1987). CRM for Part 91 and 135 operations. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 158-169). San Francisco, CA.

Article type: Review
Keywords: Communication; CRM; LOFT; Military; Training;

The authors discuss the concepts of Flight Deck Management (FDM). They suggest that two of the largest challenges that must be faced are: (1) how to get the average pilot to accept his or her limitations; and (2) how to train the pilot to accurately understand these limitations. The FDM Cycle is represented by a triangle. The base of the triangle represents the crew's development of an accurate and comprehensive concept of how the flight is going to be successfully accomplished (i.e., this process develops the *plan*). As the crew conducts the flight, they remain alert for events which conflict with the *plan*. Once these *challenges* (one side of the triangle) are recognized and validated, the crew must generate an appropriate *response* (the other side of the triangle). On the basis of the changed situation, the crew then returns to the base of the triangle, revising their *plan* for the successful completion of the remainder of the flight. During the course of the flight, the crew must use all of their resources. One important resource is the use of effective communication by each crewmember (especially the pilot). A training objective of the program is to make sure that the crew stays vigilant to the possibility of changes and challenges and communicates any such things to other crewmembers. In addition, a three day workshop on Flight Deck Management is conducted. Use of multi-media presentations of past accidents try to provide insight into problems. Materials used in the workshop include a reference handbook, which contains limitations and technical information about the aircraft and an operations handbook, which includes checklists and tabulated performance data. LINE and LOFT simulations are included, following NASA and FAA guidelines.

Lauber, J. K. (1979). Resource management on the flight deck: Background and statement of the problem. In G. E. Cooper, M. D. White, & J. L. Lauber (Eds.), *Resource Management on the Flight Deck, Proceedings of a NASA/Industry Workshop* (pp. 3-16). San Francisco, CA: National Aeronautics and Space Administration.

Article type: Review
Keywords: Accidents; CRM

The author cites the lack of effective management of available resources by flight-deck crews as the principle cause of many incidents and/or accidents. This conclusion arose out of the reviews of several research studies. First, a pilot interview study conducted by NASA (in 1973) showed that pilots frequently commented that training needed to be provided in areas such as crew coordination and effective leadership. A second source of information came from a full-mission simulation study conducted by Ruffell Smith (1979), which led to the identification of resource management as being an important variable in crew performance. Finally, a review of 60 jet transport accidents that occurred between the years 1968 to 1979 indicated that resource management problems played a significant role in many instances. Several of those accidents are discussed. A search of the Aviation Safety Reporting System was also conducted. Accidents which had resource management problems were identified. These accidents were analyzed and a summary of the skills, organization and process variables and resources is given. Several different approaches relevant to dealing with resource management problems are described (e.g., training, awareness programs, new operational guidelines or rules).

Lauber, J. K. (1987). Cockpit resource management: Background and overview. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 5-13). San Francisco, CA.

Article type: Review
Keywords: CRM; Communication; Coordination; Leadership

The author defines cockpit resource management (CRM) as the effective utilization of all available resources--hardware, software and liveware--to achieve safe, efficient flight operations. CRM is the integration of all these things; all crewmembers have responsibility and resources to add to the process, not just the Captain. The article describes seven major principles of CRM: (1) delegation of tasks and assignment of responsibilities; (2) establishment of priorities; (3) monitoring and cross-checking; (4) use of information; (5) problem assessment and the avoidance of preoccupation; (6) communications; and (7) leadership.

Lauber, J. K., & Foushee, H. C. (1981). *Guidelines for Line-Oriented flight training* (Volume 1, NASA CP-2184). Moffett Field, CA: NASA Ames Research Center.

Article type: Anecdotal/Descriptive
Keywords: CRM; LOFT; Training Program Design

These are the proceedings from a NASA/Industry workshop whose objective was to develop guidelines which could be used by any organization to develop a resource management training program to fit its own individual needs. There are seven chapters, each dealing with an issue relevant for designing, implementing and evaluating a line-oriented flight training (LOFT) resource management training program. A LOFT is defined as "the use of a training simulator and a highly structured script or scenario to simulate the total line operational environment for the purpose of training flight crews" (p. 5). When designing and developing scenarios, the particular needs of the organization should be considered. Scenarios should be designed to address specific objectives and state what the desired end products should be. The particular route structure of a carrier should be incorporated into the origin, routing and destination of a scenario. Other factors such as weather, climate and environmental factors should also be considered. The scenario should be as realistic as technically, economically and operationally feasible. The "rules" of the simulation should be fully explained to the participants, so that things will not interfere later with the realism of the scenario. After that initial briefing, the procedure should follow the normal pattern as closely as possible. Trip paperwork should be provided and should be identical to what a crew would receive before a line trip. Communications should be conducted through the channels normally used. External communications should be credible and realistic. Supplemental background communication may be channeled in as well. The LOFT environment should be created so that crewmembers are able to make mistakes and not fear repercussions or embarrassment. People should be open and enthusiastic about the process. On the other hand, to improve the training experience, crews may need to have their performance critiqued. An instructor must fill many roles in the simulation. It is felt that one instructor may be able to successfully *run* the scenario, but there will not be enough time to adequately evaluate the crews' performance. Thus, if possible, two instructors should be used. One additional use of LOFT is with initial training. The focus of the scenarios should be on "normal" line operations and should de-emphasize emergency and abnormal operations. It should help the new crewmember function as a team member. A full line crew may not be necessary. Use of LOFT in transition training is much like initial training. The focus is still on the individual functioning as part of a team. Scenarios should emphasize unique aspects of the equipment to which the trainee is transitioning. LOFT may be very helpful in upgrade training. A few other uses of line-operations simulation may be, evaluating new hardware in the cockpit, operating procedures, checklists, aircraft-operating manuals, charts and other system software. At least one carrier is using LOFT for proficiency checks. This practice is likely to increase in the future.

Lee, A. T. (1991). Aircrew decision-making behavior in hazardous weather avoidance. *Aviation, Space and Environmental Medicine, 62*, 158-161.

Article type: Group Comparison Study
Keywords: Accidents; Communication; Decision-Making; Situation Awareness

Since 1975, windshears associated with microburst events have caused the many air carrier accidents in the U.S. Investigations have suggested that information management and information transfer failures may be two primary contributing factors. The purpose of this study was to assess the situation awareness and decision-making behavior of aircrews. The author used a line-oriented simulation of a microburst/windshear encounter. Subjects were 18 commercial aircrews (Captain & first officer); one control group (conventional air traffic control transmission), two experimental groups (conventional plus visual display of simulated ground-based Doppler radar, at one of two different distances). Results revealed that the control group had difficulty discriminating conditions and making decisions, while improved situation awareness and decision-making were found for microbursts for the experimental groups.

Leedom, D. K. (1991). Aircrew coordination training for Army helicopters: Research overview. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 284-289).

Article type: Review
Keywords: Coordination; Military; Simulation; Training

This presentation details research conducted by the U.S. Army Research Institute Aviation Research & Development Activity on rotary wing flight operations. Helicopter flights are different than commercial flights both in terms of flight conditions and time available to deal with crises. For this group, aircrew coordination is defined as the interaction between crewmembers (communications) and the actions (sequence and timing) necessary for flight tasks to be performed efficiently, effectively and safely. A review of accident data was completed and the errors were classified into one of four categories: (1) quality and frequency of information exchange in the cockpit; (2) cockpit workload; (3) cross-monitoring of other crewmembers; and (4) team relationships and crew climate. The majority of the errors fell in the categories of poor information exchange or poor management workload. This is an interesting finding since these are areas not focused on in civilian crew resource management (CRM) training programs. It is believed that in the Army, the tasks are not designated by procedures as much as in commercial flights. Results from a simulated mission study showed that crews were making errors in mission planning and were not effectively using low stress periods. Six training issues were identified for Army aviators: (1) the importance of mental rehearsal for identifying and anticipating critical mission events and crewmember responsibilities associated with each unique mission; (2) the use of standardized terminology for cockpit communications; (3) the establishment of appropriate time horizons and priorities for crewmembers tasks and exchanging information in the cockpit; (4) the standardization of individual task responsibilities and information exchanges within the

cockpit; (5) the reinforcement of classroom training through mission-oriented flight simulator sessions; and (6) the systematic measurement and evaluation of crew level mission performance.

Lewis, S. T. (1975). Human factors in Air Force aircraft accidents. *Aviation, Space and Environmental Medicine, 46,* 316-318.

Article type: Single Group Study
Keywords: Accidents; Coordination

Lewis examined Air Force accident data (using flight surgeon completed Air Force Form 711g) between 1971 and 1973 (545 accidents) to determine the leading causes of these accidents. He found: (1) 2-3 "human factors" per accident; (2) *poor crew coordination* was the most frequently identified human factors problem; (3) when failure to use proper procedures was the cause it could be traced back to lack of training/experience and that (4) training was probably the key to preventing human factors errors.

Lubner, M., & Markowitz, J. (1991). Towards the validation of the Five Hazardous Thoughts measure. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 1049-1054).

Article type: Validation
Keywords: 16 PF; Accidents; Personality

This study examines the validity of the Five Hazardous Thoughts scale, a paper-and-pencil self-assessment measure. There are five subscales: anti-authority, invulnerability, macho, impulsivity, resignation. In this 25-item instrument, a scenario is presented in which a pilot reacts incorrectly. This is then followed by five reasons for the pilot's incorrect action. Each of the five responses represents one of the five hazardous thoughts. In this study, two groups of subjects participated; the first group consisted of pilots who had had a serious incident or accident within a recent time frame, the second group (control group) had not had an accident or incident. All were sent a questionnaire which contained the following measures: the Five Hazardous Thoughts scale; two scales from the 16 PF, Ego - C and Superego - G; three items from Gurin's Personal Efficacy subscale; Zuckerman's Thrill and Adventure Seeking subscale; and other measures of the pilots' safety behaviors, stressful life events, flight experience and demographic characteristics. Results showed that there were no significant differences on any of the subscales which differentiated the two groups, however, the some construct-validity evidence of the scale was demonstrated. In addition, there were significant differences between the two groups on flight experience and other personality variables; this may have contributed to the nonsignificant findings.

Margerison, C., McCann, D., Davies, R. (1987). Aircrew team management program. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 132-134). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: CRM; Training Evaluation; Training Program Design

In order to build the Aircrew Team Management (ATM) program of Trans Australia Airlines, the designers of the program met with check airmen, line crew representatives, union representatives, a steering committee and management representatives. It was decided that the training should focus on three key issues: (1) understanding oneself and others (using personal profiles and discussion via the team management index); (2) communication skills (using conversational control, communication styles and role plays); and (3) team skills (using team work and decision-making simulations). The first presentation of the training program took place during a 3-day live-in workshop. Prior to attending the workshop, each crewmember was required to go through a booklet of materials containing an outline of actual aviation incidents and accidents and other reading material relevant to the workshop. In the final session of the ATM workshop, participants were asked to complete a short semi-structured questionnaire evaluating the workshop; 100 percent of the participants rated the workshop as at least partly relevant. Ratings of the usefulness of the program, showed that 93 percent stated that the workshop was either useful, very useful, or highly useful. Of all aspects covered in the workshop, participants rated interpersonal skills and personal development most highly. Most participants rated the workshop as worth attending. Further follow-up studies will examine whether the training resulted in any observable differences in team performance via simulated flights.

McDonald, B. R., & Sullivan, D. J. (1991). Interactive video disk as an instructional tool in CRM programs. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 2* (pp. 923-928).

Article type: Training Development and/or Evaluation
Keywords: CRM; training; simulation

This article discusses the development of the Interactive Video Disk (IVD) which is part of the awareness phase of the Strategic Air Command (SAC) crew resource management (CRM) program. (Details of the entire program can be found in Johnson, Shroyer, & Grewe, 1991.) The authors list some of the reasons why this method of instruction was chosen. The student is able to interact with the program, responding to questions, training would be consistent across all locations, interruptions can be accommodated, the scenarios input more realism into the situation and scenarios can be individually developed, meeting the specific needs of the relevant aircraft and the needs of the trainees. Validation of these materials was underway at the time of this writing.

Moore, M. S. (1977, May). Complexities of human factors in aviation. *Aviation, Space and Environmental Medicine*, *48*, 471-473.

Article type: Theoretical
Keywords: Accidents; Coordination; Stress

Moore emphasized that the aviation community is beginning to realize that human factors are often the cause of aircraft accidents. Seven basic categories of human failure were identified from accident data: (1)visual illusions; (2) lack of altitude awareness; (3) poor inflight judgment or decisions; (4) non-adherence to standard operating procedures; (5) failure to be sure someone is flying the airplane during routine irregularities; (6) failure to monitor critical flight instruments; (7) poor crew coordination. The author provided recommendations for stress prevention, including: (1) selection (select pilots with stable personalities, good judgment, leadership, reliability, flexibility, adaptability); (2) training (training and retraining is essential for all aspects of the aviation industry to encourage social integration and professional proficiency); (3) task design; (4) system design; and (5) human needs (pay attention to medical, psychological, equipment, certification needs of pilots).

Morrow, D. G., Lee, A. T., & Rodvold, M. (1991). Collaboration in pilot-controller communication. *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 278-283). Columbus: Ohio State University.

Article type: Single Group Study
Keywords: Communication; Coordination; Situational Awareness

The authors attempt to provide a mechanism to improve communication efficiency and accuracy of pilot-controller communication. Two methods are described: (1) establish base rate information concerning types of problems that disrupt routine communication; and (2) identify factors associated with the problem in order to provide possible explanations for why they occur and how they can be eliminated. One main aspect of their framework is that communication requires both an individual effort (e.g., working memory capacity) and a collaborative effort (the cognitive resources necessary to accept information as understood and appropriate). Three kinds of problems that can disrupt communications are: (1) procedural deviations; (2) inaccuracies; and (3) nonroutine transactions. These researchers examined 7,998 transactions between pilots and controllers (half from approach sectors, half from departure sectors) that occurred over a 42 hour period. Results suggested that controllers and/or pilots may produce longer messages to reduce transaction and turn-taking time, but that this, while decreasing individual effort, increases collaborative effort (especially during high-traffic periods). Thus, the likelihood of procedural deviations and nonroutine transactions is significantly increased. This study demonstrated that how well pilots and controllers communicate depends on more than how easily pilots/controllers produce and understand messages; it also depends on collaborative effort (i.e., how well they collaborate to accept messages as mutually understood).

Mosier, K. L. (1991). Expert decision-making strategies. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 266-271).

Article type: Validation
Keywords: Decision-Making; LOFT; Simulation; Situation awareness

The author bases much of the work in this study on research that has examined Recognition-Primed Decisions (RPDs). These are defined as decisions for which response alternatives are directly derived from recognition of critical information and prior knowledge. The emphasis in this model is on situation assessment, as applied in a team setting. Team members assess the situation and gather information to support the accuracy of information. In this study, the quality of information transfer was analyzed for crew performance. Situation assessment was predicted to have a large effect, that is, performance was expected to be related to the correctness and completeness of situation assessment. In addition, the Captain's personality was hypothesized to be a moderator variable. The subjects were 23 three-person crews flying a full-mission simulation Boeing-727 simulator. Each crew was rated on their coordination performance and on performance errors. Information was broken down into decision-related (D) information and non-decision (ND) information. Results showed that decision-related information helped to predict crew coordination and Type 2 errors (moderately severe errors). Non-decision information predicted Type 3 errors (major errors). There was an interaction between Captain's personality and which crewmember provided the decision-related information. Personality was predictive of performance on the first day of the simulation, but on the second day, information transfer was more predictive of performance.

Mudge, R. W. (1987). Cockpit management and SBO's. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 56-60). San Francisco, CA.

Article type: Theoretical
Keywords: CRM; Training Program Design

According to the author, the specific behavioral objective (SBO) is one of the primary tools of a training program. An SBO is simply a statement which specifically identifies a small segment of the final behavior sought, and a little more. The author believes it is critical to pinpoint *exactly* what it is you want the pilot to *do* after completing the training and what should be evaluated from the point of view of both the program and the pilot. There should be specific SBO's for each crewmember position. The author's program attempts to eliminate negative skills and add positive skills in order to provide a well-balanced repertoire of skills, which are in turn, protected by a strong "attitudinal umbrella and knowledge base." Eight factors are listed which have been found to be important for crew resource management training.

Murphy, M. (1980). Analysis of eighty-four commercial aviation incidents: Implication for resource management approach to crew training. *Proceedings of the 1980 Annual Reliability and Maintainability Symposium IEEE.*

Article type: Single Group Study
Keywords: Accidents; Attitudes; Personality; Communication; Coordination

Murphy utilized a "resource management approach" to analyze 84 aircrew incidents that were judged to be an exemplification of failure in maintaining a unified crew effort. Each of the incidents were classified into one of five resource management analytic categories: (1) resource management skills; (2) organization-process related variables; (3) personal factors; (4) material resources internal to the aircraft; and (5) resources external to the aircraft. In addition, each of the incidents were content analyzed to determine whether enabling factors (those factors allowing a situation to develop), associated factors (these are less directly contributable) and/or recovery factors (positive actions leading to recovery) were involved. The seven subcategories of skills deficits (or exemplary behavior) that resulted in the incident under the heading of resource management skills were: (1) management skills; (2) communication skills; (3) planning skills; (4) decision-making skills; (5) problem-solving skills; (6) leadership skills; and (7) social skills. The organization-process related variables that resulted in an incident were monitoring, workload and role. Personal factors involved experience, motivation, stress reaction, fatigue and proficiency. Material resources internal to the aircraft fell into one of two subcategories: textual (e.g., poor human engineering of tools) and equipment. Resources external to the aircraft generally involved poor coordination on the behalf of air traffic control (ATC). The author concludes that resource management training should be focused on one of three areas. The first area is concerned with interpersonal relations and mainly involves ATC clearance information. The second area is concerned with the management function of setting task priorities. The final area is concerned with planning and decision-making.

O'Hare, D. (1990, July). Pilots' perception of risks and hazards in general aviation. *Aviation, Space and Environmental Medicine, 61*(7), 599-603.

Article type: Single Group Study
Keywords: Civilian; Decision-Making; Simulation; Situational Awareness

This study cites a National Transportation Safety Board (NTSB) report that concluded over 50 percent of fatal general aviation (GA) accidents were attributable to faulty decision-making. Prompted by these results, the author examined pilots' perceptions of risks and hazards in GA and their impact on "in-flight" information acquisition and decision-making in a computer flight simulation. Subjects were 44 licensed pilots recruited from a safety seminar. They completed the Aeronautical Risk Judgment Questionnaire (ARJQ) which measured pilots' self-judgment, hazard awareness, personal invulnerability, risk awareness and various background variables. In addition, 18 of the 44 subjects performed a computerized flight simulation with marginal visual flight rule (VFR) conditions. With respect to self-judgment, subjects rated themselves as above average on skill and judgment and as much less likely than average to take risks in flying. Only

six of the 18 subjects who performed the computer task elected to takeoff, the remaining either canceled or delayed the flight. Deciding to takeoff was considered a "risky decision" because of the marginal weather conditions. There were significant positive relationships between self-rated risk taking propensity and the decision to takeoff. In addition, those who decided to takeoff also scored significantly higher on personal invulnerability and had poorer knowledge of the phases of flight during which accidents are likely to occur. The author concludes that in addition to providing information concerning hazards in general aviation to pilots, it is just as important to provide training which emphasizes the relationship between personal actions and risk.

Orasanu, J. M. (1993). Decision-making in the cockpit. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 137-172). San Diego, CA: Academic Press.

Article type: Review
Keywords: CRM; Decision-Making; Situation Awareness; Training

Decision-making should not be thought of as a unitary construct, but rather involves many different kinds and/or levels of decisions. According to the author, all types of decisions have three elements in common: (1) choice among options; (2) situation assessment; (3) and risk assessment. Problems that an air crew face may be either well-defined or ill-defined. If the problem is well-defined, the response to it will most likely be prompted by certain environmental stimuli (e.g., warning lights) and the crew will simply need to follow some prescribed procedure. *Rule-based decisions* are the easiest to make. A person must decide whether the present conditions fit with a prespecified pattern. For ill-defined situations, the cause of the problem is frequently unknown and the crew needs to decide how to best approach the problem. *Knowledge-based decisions* include those decisions which are clearly defined but there are multiple response options which are available or required. With ill-defined problems, two types of approaches are primary. The first type of decision is *procedural management*. The crew has noticed that cues are pointing to the fact that all is not normal with the aircraft, but they do not have a clear idea as to what is causing the problem. The second type of decision is the least frequent and consists of *creative problem-solving*. The problem must be diagnosed and then appropriate response alternatives must be generated. There are no past procedures to tell the crew how to deal with the situation. Situation assessment and response generation are both critical in these types of cases. Diagnosis is also very critical. The crew must create a solution which fits with the conditions that exist. The author also deals with how to determine the quality of decisions made in the cockpit. The criteria used should be different than those used to judge decisions made under laboratory conditions. *Process failures*, which are interpersonal processes, can contribute to a crew's performance under less than ideal conditions. These can come from the crew failing to share information or sharing misconceptions. Also crewmembers may refuse to work together or in the process of doing so make more risky decisions than they would have as individuals. A second problem is *performance failures*, these are difficulties in performing the task. Examples of this can be interruptions in performing a task, not giving information in a timely manner, not prioritizing tasks, or being provided with ambiguous goals or task assignments. Individual crewmembers may also exhibit *hazardous attitudes* which can reduce the effectiveness of a crew.

According to the author, four aspects of crew behavior have been identified which are associated with effective decision-making: (1) good situation awareness; (2) high levels of metacognition; (3) shared mental models based on explicit communication; and (4) efficient resource management. There are also certain skills that cut across all situations. The first step the crew must take is to assess the situation. Next, all decisions involve some type of risk assessment. Training should deal with the trade-offs that are involved. Metacognition may be one of the most trainable skills, although crewmembers must already possess relevant strategies. If the crew engages in appropriate communication, then a shared problem model will be developed. Finally crews need to be taught resource management skills. According to the author, there are still some specific things that the research has not told us. We need to have a better understanding of the decisions that are made in a cockpit and what knowledge, skills and strategies are needed to make those decisions. At this time, we lack good definitions of performance and criteria to judge it against. There also needs to be a better determination of which decision-relevant skills can be automated. Finally, more research needs to examine the impacts of stressors on decision-making performance.

Orasanu, J., & Fischer, U. (1991). Information transfer and shared mental models for decision-making. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 272-277).

Article type: Validation
Keywords: Communication; LOFT; Situation Awareness

According to the author, language and its effects on crew performance have not been fully studied. This article addresses two questions: (1) How is language used to manage problems in the cockpit?; and (2) Are there differences between 2-member and 3-member crews in their communication and problem solving strategies? Two data sets were used for this study, the first was from Foushee, Lauber, Baetge, & Acomb (1986). They had 2-member crews fly a single 45-minute flight simulation and two decisions needed to be made during the scenario: (1) when to abort the landing; and (2) the choice of an alternative landing site. The communication patterns of the five highest and the five lowest performing crews were analyzed. The second data set was obtained from a Chidester, et al. (1990) study that used 3-member crews flying a two day, 5-segment simulation. The communication patterns of the four highest and the four lowest performing crews from this study were included. In Study 1 and Study 2, higher performing crews had higher situation awareness and took more actions. In Study 1, Captains in high performing crews talked less when workload was high than when workload was low. Low performing Captains showed the opposite pattern. High performing crews did more planning, strategizing and obtaining information early in the flight. In Study 2, some opposite patterns were found. The high performing Captains talked more than the low performing Captains. They were using more Metacognitive/Problem Solving talk, but also formed more plans during the abnormal phase.

Orlady, H. W. (1993). Airline pilot training today and tomorrow. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 447-477). San Diego, CA: Academic Press, Inc.

Article type: Review
Keywords: Coordination; CRM; Training

Safety is one of the most important issues in aviation. If an airline is not safe, then it will not survive. Operations must also be efficient. This chapter discusses safety issues from a training perspective. According to the author, the two newest training concepts are line-oriented flight training (LOFT) and crew resource management (CRM). Many people around the world use these terms, but there is not total agreement on what they mean for the industry. Line-oriented flight training or LOFT was originated by Tom Nunn at Northwest Airlines. Also at this time, NASA researchers were exploring the relationship of intra-crew behavior on the accident rate. Both groups were converging on similar training ideas. One important distinction to make is between LOFT and full-mission simulation. They are not the same thing. LOFT is a training concept and should not include any sort of check or evaluation. Full mission simulation has uses that go beyond those of LOFT. The Ruffell Smith (1979) study helped to bring the concepts of LOFT and CRM to the forefront of aviation research. Changing pilot training has many implications for both managers and pilots. Management must communicate effectively with its pilots. An organization should also make sure that its philosophy, policies and procedures are adequately reflected in training materials. The jobs of check airmen and flight instructors are becoming more difficult and an important issue to be dealt with in the future is how to select and train people for each of these jobs.

Platenius, P. H., & Wilde, G. J. (1989, January). Personal characteristics related to accident histories of Canadian pilots. *Aviation, Space and Environmental Medicine*, *60*(1), 42-45.

Article type: Single Group Study
Keywords: Accidents

The authors developed a 302-item questionnaire designed to link personnel characteristics to accident histories. The survey was mailed to 70,000 Canadian pilots; and a data set was constructed which contained data from 8,819 English-speaking pilots (all male). One item asked the pilot if he or she had been an aircraft accident; all other item responses were then analyzed as to the item's relationship with the accident criterion. This resulted in 249 significant "accident markers." In order to develop scales to predict future accident-proneness, the authors grouped (both empirically and intuitively) these items into meaningful categories. These intuitively-derived categories were: (1) life events and preoccupations; (2) risk acceptance; (3) lack of humor appreciation; (4) asocial or sedentary hobbies; (5) medical symptoms; (6) perceiving oneself as unsuccessful; (7) lack of initiative/self control; (8) social disability or loneliness; (9) alcohol use; (10) automobile accidents/violations; and (11) under-attribution of accident causes. The authors hope to conduct a predictive study in the future using these accident-proneness categories.

Povenmire, H. K., Rockway, M. R., Bunecke, J. L., Patton, M. W. (1989). Cockpit resource management skills enhance combat mission performance in B-52 simulator. In R. S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 489-494). Columbus: The Ohio State University.

Article type: Validation
Keywords: CRM; Military

There are several differences between operations at Strategic Air Command (SAC) and those of typical airlines. B-52 crews are permanently assigned (fixed) and tend to be younger and less experienced than their civilian counterparts. This study examined whether B-52 crews could benefit from crew resource management (CRM) training and if there would be an accompanying increase in mission performance. The authors observed a tactical Weapon System Trainer (WST) mission and evaluated both CRM and mission performance. Squadron Operations Officers were used to rank their own crews. The crews chosen to participate were ranked either in the top third or bottom third of the rankings (seven crews were included). CRM behaviors were rated by evaluators using the LINE/MOST worksheet and a relative mission ranking was given by the simulator instructors and the First Combat Evaluation Group (ICEVG). Positive correlations were found between both LINE/MOST Worksheet score and mission performance ranking ($r = .84$, $p < .05$) and with bombing rank ($r = .71$, $p < .05$). Other results showed that crews had higher scores when the Electronic Warfare Officer (EW) asked questions more frequently of the Navigator. Conflicts were correlated with LINE/MOST Worksheet scores; results showed crews with fewer conflicts did better, poorer crews left more conflicts unresolved and poorer crews' conflicts were more often about *who* is right rather than *what* is right. Squadron rankings and mission performance rankings were negatively correlated. Substitutions were made in many crews in the Radar Navigator (RN) position. This job has a large effect on the success of the entire crew. Thus, substitutions may have had a larger effect than previously thought. Situational awareness was also an important factor in performance. It was enhanced when there was confirmation or challenge of information. The authors concluded that CRM performance enhanced combat mission performance in the B-52 WST.

Praktiek, J. (1991). KLM feedback and appraisal system for cockpit crewmembers. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 410-415).

Article type: Anecdotal/Descriptive
Keywords: CRM; training program design and evaluation

KLM developed an appraisal system for non-technical skills, termed the Feedback and Appraisal System (FAS) for cockpit crewmembers. This system was developed using a seven phase process. In the first phase, critical incident interviews were conducted to get information about non-technical skills. Fourteen criteria were grouped into five general performing dimensions: (1) Work Attitudes - procedure orientation, assertiveness, exercise of self criticism, sense of responsibility, service orientation; (2) Information Management - information analysis, planning

and anticipation, decisiveness; (3) Stress Management - stress management; (4) Crew Cooperation - working with others, sensitivity, image; and (5) Captaincy - task oriented leadership and people oriented leadership. In the next phase, a relevancy rating was developed for each event, followed by the third phase in which all superfluous words were edited. The fourth phase called for a scaling of the relevant events. A five level ranking method was used. Next, two types of rating scales were developed: Behaviorally Anchored Rating Scales (BARS) and Behavior Observation Scale (BOS). An example was given for each type of scale. In the sixth phase, the rating scales were tried in both a simulator and on route training. The final phase assessed which of the scales was preferable for KLM. The BARS format was rated as more preferable.

Predmore, S. C. (1991). Microcoding of communications in accident investigation: Crew coordination in United 811 and United 232. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 350-355).

Article type: Single Group Study
Keywords: Accidents; Communication; Coordination

This study analyzed the communication patterns from two accidents, United 811 and United 232. In both of these accidents, crewmembers were able to work effectively together to minimize the number of casualties that could have occurred. The Cockpit Voice Recorders (CVR) were used to obtain the information. Communication levels increased considerably after the onset of the problem. Very little of the communication could be classified as being incomplete, which means that the crews were efficient in communicating information. Verbal behaviors were classified in terms of action decision sequences (ADSs). These are events which require coordinated actions on the part of the crew. The following ADSs were identified: (1) flight control; (2) damage assessment; (3) problem solution; (4) landing; (5) emergency; (6) preparations; and (7) social. Each of the two crews communications were broken down into these seven ADSs. The majority of both flights communications were in flight control. Crews could do little about the problems encountered so little time was spent on problem solution. United Flight 232 was focused on since it was in the unique position of having a check airmen from the cabin join the crew. By doing this, it freed the Captain or first officer to attend to other task demands. The check airmen's communications were focused almost exclusively on flight control. The flight engineer focused on damage assessment and problem solution. Patterns of communication changed at different points of the flight and the Captain was able to regularly shift his attention from one issue to another.

Prince, C., Chidester, T. R., Bowers, C., & Cannon-Bowers, J. (1992). Aircrew coordination: Achieving teamwork in the cockpit. In R. Swezey & E. Salas (Eds.), *Teams: Their training and performance*. Newark, NJ: Ablex.

Article type: Review
Keywords: Coordination, Training

This book chapter reviews current practice and research issues related to crew coordination. The authors suggest that accident data, research and pilot opinion led to a focus on crew coordination issues. Three approaches to crew coordination are discussed in terms of selection, task design, and training. The authors differentiate between attitudes (modifiable through training) and traits (stable, deep-seated predispositions to respond in a particular way). Traits are therefore an issue for selection. Research reviewed in this article revolves around two instruments targeted toward personality traits and attitudes. The personality measure was the Personal Characteristics Inventory (PCI; which is composed of the Extended Personal Attributes Questionnaire -- EPAQ, the Work and Family Orientation Questionnaire -- WOFO and the Achievement Striving and Impatience/Irritability scales of the Jenkins Activity Survey -- AS, I/I). The PCI revealed 2 trait dimensions, *instrumentality* (independence & goal level) and *expressivity* (interpersonal warmth or sensitivity). Crew attitudes were assessed by the cockpit management attitudes (CMAQ). The authors interpret three recent studies using the PCI as suggesting that personality may set limits on training effectiveness and attitude change. They note that if additional evidence begins to accumulate supportive of such notions, then organizations may need to afford the same selection consideration to personal attributes as they do to aptitude. The second approach to crew coordination is task design. There has been little research on the impact of task design on crew coordination. The third approach noted by Prince et al. is training. Aircrew coordination training (ACT) and crew resource management (CRM) training emphasize need to integrate and utilize all resources available to the cockpit. The authors discuss three main phases (recommended by an FAA circular): (1) awareness (seminar and group exercises focused on communication, decision-making, stress/workload management, leader/subordinate responsibilities, management styles); (2) practice and feedback (crew flies a realistic scenario in a simulator [like the LOFT program]); and (3) reinforcement (CRM concepts become part of the organization's overall training and operation practices). The chapter ends with a discussion of current training practices and problems. The authors suggest that communication training should include subtopics of cultural influence, barriers, assertiveness, participation, listening, feedback and legitimate avenue of dissent. They suggest that crew coordination training should include seven skill dimensions: (1) situational awareness, decision-making, mission analysis (these 3 cluster under task-work skills), communication, leadership, adaptation and assertiveness (these 4 cluster under teamwork skills).

Prince, C., & Salas, E. (1993). Training and research for teamwork in the military aircrew. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 337-366). San Diego, CA: Academic Press, Inc.

Article type: Review
Keywords: CRM; Military; Training; Training Evaluation; Training Program Design

The military developed an aviation program very soon after the invention of the airplane. Cockpit resource management (CRM) training, however, was developed primarily by civilian air carriers and mostly as a direct result of research conducted by NASA and an increasing public concern for aviation safety. Military aviation was slow to implement CRM-related programs for several reasons. First, military aircraft are frequently not equipped with the same recording devices as that of civil carriers. Second, concern is not as great with military accidents since fewer lives tend to be lost. Military leaders also felt that crew interactions were adequately covered by current training programs. Military aviation personnel soon realized, however, that many aviation incidents/accidents were caused by coordination errors. Errors that may have been avoided had these crewmembers been trained in the principles of CRM. Consequently, training programs were started by the Army, Air Force and Navy. These programs were generally called aircrew coordination training (ACT). This chapter describes the programs developed by each of these branches of the military. The military's reaction to the CRM programs has been fairly positive. It became evident though that the programs needed to be more tailored to the military setting. The authors also present seven skill behaviors that were identified and state that research support has been found by all services. The first is *mission analysis*, the skill in organizing and planning. This is one of the first tasks accomplished by an air crew and can be critical for a crew just beginning to fly together or that is flying a non-routine mission. *Assertiveness* is the second skill and has been seen in many CRM training programs. Research has determined that it is a trainable skill and it also requires practice and feedback. *Leadership* does not come as a surprise as an important skill. Much research has shown the importance of *communication*. It is a complex variable though and there needs to be a determination of which types of situations are best for certain types of communication. *Situation awareness* has been cited often as a cause for mishaps and is very important for the safety and effectiveness of a mission. Training for crews in *decision-making* has shown that it can help crews make quicker decisions with fewer errors. Finally, crews need to be *adaptable and flexible* in their performance in the aircraft. There needs to be further research that examines the effectiveness of various methods for training CRM skills. Since the military is focusing on behavior change instead of attitude change, the methods utilized are likely to be different from those utilized by civilian air carriers whose focus has mostly been on changing attitudes. Additionally, it has been determined that feedback is important, but there are still questions that remain unanswered such as what information should be provided, what is the appropriate channel for the information and what should be the format of the information? Other issues will also need to be addressed, such as how increased automation will change the type of training that is needed.

Rose, R. M., Helmreich, R. L., Fogg, L., & McFadden, T. J. (1993, September). Assessments of astronaut effectiveness. *Aviation, Space and Environmental Medicine, 64*(9, Section(1), 789-794.

Article type: Single Group Study
Keywords: Personality; Performance ratings

Rose et al. examine task and interpersonal measures of astronaut effectiveness. Based on previous research, the authors suggested that task and interpersonal factors were both critical to success as an astronaut. From previous experience and discussions with astronauts they identified 9 rating categories: (1) job competence -- knowledge; (2) job competence -- performance; (3) job competence -- performance under pressure; (4) leadership; (5) teamwork; (6) group living; (7) personality; (8) communication skills/external relations; (9) space station (would enjoy working with this person on a 3-month Space Station tour). The sample was 84 astronauts and 1 supervisor; the astronauts rated all fellow astronauts with whom they were familiar (peer ratings) and the supervisor rated all of the astronauts. The ratings consisted of rank ordering the top 5 and the bottom 5 astronauts on each of the nine dimensions. The supervisor assigned astronauts to 1 of 4 quartiles, then ranked all of those in 1st and 4th quartile. The authors found interobserver reliabilities for each dimension in the 50's and 60's. Then, to simplify data for further use, used a normalizing transformation to arrive at a single summarizing peer rating per dimension. Confirmatory factor analysis of normalized peer data identified 3 factors: (1) interpersonal skills (group living, personality); (2) performance/competence; and (3) leadership. In terms of supervisor ratings, since ranks did not lend themselves to normalized transformations, they looked at Cronbach alpha across 9 dimensions (alpha was .93, but personality ratings were less consistent, so personality was dropped, boosting alpha to .97). The authors decided to combine 8 dimensions (minus personality) to form a single rating factor. Finally, they looked at convergent validity and found that total flights, number of commanded flights and experience were highly intercorrelated (as a cluster) as were peer and supervisor ratings (as a cluster), but were not highly correlated across clusters. Within the factors that compose the ratings, the authors found that decision-makers place more value on technical abilities than interpersonal ability and thereby reasoned that there should be room to improve decision-making by incorporating measures of interpersonal ability.

Schwartz, D. (1987). CRM training for FAR parts 91 and 135 operators. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 170-177). San Francisco, CA.

Article type: Training Development and/or Evaluation
Keywords: CRM; Situation Awareness; Training Program Design

The author describes the training provided by Flight Safety International (FSI). FSI has analyzed accidents according to the "Safety Window." This is defined as encompassing all tasks dealing with the approach, landing, taxi, take-off and climb phases of the flight. The reason why these segments have been isolated is that this is a time of increased crew workload and the time where

over eighty percent of accidents occur. The heart of the training program is the concept of "situational awareness," defined by the author as "an accurate perception of the factors and conditions that affect an aircraft and flight crew during a specific period of time." There are five elements which contribute to situational awareness: (1) Experience and training; (2) physical flying skills; (3) spatial orientation; (4) cockpit management skills; and (5) health and attitude. All of these factors are part of individual situation awareness, yet the group situation awareness is not the sum total of each individual. The group situational awareness is only as good as the level of situation awareness of the pilot-in-command. The rest of the crew may have high situational awareness, but if the Captain does not, then problems can occur. Past accidents have shown this to be true. Usually accidents are caused by a series of errors called the "error chain." The author lists ten clues which may be used as a way to break the error chain and show the loss of situational awareness: ambiguity, fixation or preoccupation, empty feeling or confusion, violating minimums, undocumented procedures, nobody flying the aircraft, nobody looking out the window, failure to meet targets, unresolved discrepancies and departure from standard operating procedure.

Segal, L. D. (1989). Differences in cockpit communication. In R. S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 576-581). Columbus, OH: Ohio State University.

Article type: Theoretical
Keywords: Communication; Coordination

This paper examined how the spatial layout of the cockpit affected crew communication and how this, in turn, affected crew performance. According to the author, there are two sources of information available to pilots, the: (1) environment; and (2) actions of others in the environment. Communication can be verbal as well as nonverbal; in fact, any behavior can be construed as communication, interpretation of the behavior is what gives it meaning. Pilots assign meaning to certain actions because of their familiarity with the context. In today's technologically advances systems some controls have been consolidated into a single display, consequently, it is more difficult for another crewmember to interpret what the other crewmembers are doing because their actions are less visible to others. Cockpit designers need to recognize that the cockpit should be "supportive of crew interactions," and not ignore these issues in a rush to implement the latest technology on the flight deck.

Sellards, R. (1989). Testing for potential problem pilots and human error in the cockpit. In R. S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 582-587). Columbus, OH: Ohio State University.

Article type: Single Group Study
Keywords: Accidents; CRM; Personality

Researchers at USAir have identified a personality profile that they termed *the Potential Problem Pilot Profile* (P4). Even though only about five percent of the total number of pilots exhibited this personality profile, it is still an issue that needs to be dealt with. Captains and first officers must deal with many different types of personalities, thus, training needs to focus on helping all crewmembers deal with many different personalities over a broad range of situations. Some of these pilots may be accident-prone and attempts should be made to identify and diagnose these individuals. Some types of psychological factors involved in accidents are delineated by the author. Some of these causes may be temporary or situation specific and these may be the ones that would most benefit from Crew Resource Management (CRM) training. Characteristics of these P4 individuals are listed and are assumed to contribute to pilot accident-proneness.

Shroyer, D. H. (1987). The development and implementation of cockpit resource management in UAL recurrent training. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop (NASA*-CP-2455) (pp. 47-49). San Francisco, CA.

Article type: Training Development and/or Evaluation
Keywords: CRM; LOFT; Training; Training Program Design

The author lists several guidelines which should be followed when developing a line-oriented flight training (LOFT) scenario. First, scenarios should be tightly scripted so that the objective and content are well-defined. Second, the content should be rigidly controlled to enhance the learning experience. Third, there should be a plan formulated for updating and continuing progress of the system.

Simon, R., Pawlik, E. A., & Bronkhorst, T. M. (1991). Aircrew coordination for Army helicopters: An exploration of the attitude-behavior-performance relationship. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 290-295).

Article type: Validation
Keywords: Attitudes; Coordination; Military; Simulation

This study examines the relationships between crew coordination attitudes, behavior and mission performance. A revised version of the cockpit management attitudes questionnaire (CMAQ) called the Army CMAQ was developed which contains 45 items rated on a seven-point Likert

scale. Behaviors were measured using the Aircrew Coordination Evaluation (ACE) Checklist and the Aircrew Training Manual (ATM) Tasks. Ratings were provided by trained instructor pilots. Crews flew a two-hour simulation. Twenty 2-person crews flew the simulation, while an additional 80 aviators responded to the Army CMAQ. The structure of the Army CMAQ was explored in a series of statistical analyses described by the author. It was decided that aircrew CMAQ scores should be used to assess the relationship between attitudes and behavior. Thus, individual CMAQ scores were combined. Results showed that aircrew coordination behaviors affect performance. Furthermore, attitudes toward aircrew coordination affect both crew coordination behavior and performance.

Singh, G. (1991). Comparative analysis of MMPI profiles in two groups of ab-initio flying trainees. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 694-699).

Article type: Group Comparison Study
Keywords: MMPI; Personality; Stress

In this study, the author examined the interaction between environmental variables and person variables, as well as the relationship between a personality profile and performance of flying trainees. The first group of subjects was 36 Indian Air Force cadets who underwent initial flight training under high stress conditions. The second group was 25 Army Artillery Corps officers who were also undergoing flight training, but under low stress conditions. The Minnesota Multiphasic Personality Inventory (MMPI) was administered to all subjects to obtain a personality profile. Performance ratings were also gathered during the course of training for all subjects. The first group of subjects was placed into two groups: those who scored in the normal range and those who scored in the aberrant range. All subjects from the second group fell in the normal range. Profiles for the two normal groups were almost identical. Scores for the aberrant group were elevated on some scales. It is felt that this group may use abnormal procedures for dealing with the high stress levels, although their performance ratings were similar to the normal group.

Sloan, S. J., & Cooper, C. L. (1985, October). The impact of life events on pilots: An extension of Alkov's approach. *Aviation, Space and Environmental Medicine*, *56*(10), 1000-1003.

Article type: Test Development
Keywords: Accidents; Biodata; Personality

The authors illustrate how the concept of "life events" may be extended beyond Alkov's previous work and examine underlying trends in pilots' perceptions of events/processes relationships. Alkov developed/used a 22-item "life events questionnaire". Current authors recommend use of factor analysis over Alkov's nonparametric statistics, hoping to uncover any complexities present in pilots' perception of features relevant to life events/pilot characteristics/incident potential. They used an adaptation of Alkov's methodology; used 6-pt. scale rather than yes/no; asked pilots

to rate Alkov's 22 items with an eye toward building a profile of the error-prone pilot. They received 442 returned surveys (52% return rate) and conducted a factor analysis (principle axes with iteration, oblique rotation and default Delta). Three factors emerged: "emotional losses" (marital separation/divorce, problems), "pilot characteristics" (professionalism, good team leader/team member, handles life difficulties well) and "emotional gains" (marriage/engagement, new addition to family).

Stone, R. B., Babcock, G. L., & Edmunds, W. W. (1985, February). Pilot judgment: An operational viewpoint. Special issue: Aviation psychology: I. *Aviation, Space and Environmental Medicine*, *56*(2), 149-152.

Article type: Anecdotal/Descriptive
Keywords: Accidents; Situational Judgment

The authors (affiliated with the Air Line Pilots Association) argue that pilots are a hard-working, dedicated group who are unfairly labeled as possessing a "macho attitude," and after-the-fact evaluation of accidents/incidents result in "pilot error" being unfairly identified as a cause of the accident. They applaud the introduction of the LOFT simulation programs that should help pilots train for "real time" decision-making activities. The authors examine a sample of reports (selecting 70 of 134,000 reports) obtained from the Aviation Safety Reporting System (ASRS) and conclude that no commonly recurring problems were found, however, most of the incidents involved time criticality or time pressures.

Sturges, C.A. (1991). Application of instructional systems development (ISD) principles to the Advanced Qualification Program (AQP). In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 416-421).

Article type: Training Development and/or Evaluation
Keywords: Training Evaluation; Training Program Design

The Instructional System Development (ISD) model was developed for the Department of Defense back in the 1960s. It provides a systematic approach to training development and consists of five distinct phases: (1) task analysis; (2) requirements definition; (3) objective development; (4) development of instruction; and (5) evaluation of instruction. The author uses a table format to delineate the relationships between the ISD phases and AQP tasks. First, a list of all tasks performed for each position in the aircraft should be developed. Next, the training requirements need to be determined. These can be compared with the task lists to define a list of tasks that require training. Instructional objectives must then be developed for those tasks which will require formal training. Following this, the researcher develops the syllabus for the program. Each objective is assigned to a learning domain, and medium for presenting it are chosen. Once this has been approved by the FAA, the developer will begin developing and implementing the tests and coursework. After this is completed and approved, the curriculum is implemented in a formal training environment. Upon final approval, it is implemented fully.

Swezey, R. W., Llaneras, R. E., Prince, C., & Salas, E. (1991), Instructional strategy for aircrew coordination training. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 302-307).

Article type: Training Development and/or Evaluation
Keywords: military; training; CRM

The authors state that there is little empirical evidence demonstrating that crew resource management (CRM) training programs have an effect on decreasing the number of coordination errors. The Naval Training Systems Center (NTSC) has developed an Aircrew Coordination Training (ACT) program which "specifically provides crew coordination training that is skill-based, behavioral, mission specific and capable of integration into existing training programs." According to the authors, skill-based training is preferable to knowledge or attitude based training. A review of existing training programs was done and seven skill dimensions were identified for effective cockpit coordination: (1) mission analysis; (2) decision-making; (3) assertiveness; (4) flexibility; (5) situational awareness; (6) communication; and (7) leadership. The modules of the training program provide training in each of these skill areas. Participants also will demonstrate the skills and receive feedback. Videotapes are used. Mishaps are presented to show how failures can occur. The training is designed to be mission-oriented. A standardized evaluation form has been developed which combines objective measures with subjective measures.

Taggart, W. R. (1991). Advanced CRM training for instructors and evaluators. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 356-361).

Article type: Training Development and/or Evaluation
Keywords: CRM; Training; Training Program Design

According to the author, early crew resource management (CRM) training programs provided very little specialized training for the instructors and evaluators. This trend has changed. Now, trainers and evaluators typically receive a two- or three-day training course. Ten training objectives are listed by the author. Videos are also being utilized to show examples of crew performance that ranges from good to bad. Participants are introduced to behavioral markers and then they practice identifying them in the videos. Differences in ratings can then be discussed. Debriefing and critiquing skills can also be practiced and evaluated. Organizations should develop a clear policy on the importance of CRM to help decrease any misconceptions or ambiguities. Examples of text that could be used in flight manuals are also given. They also recommend avoiding the overuse of well-known accidents and spending too much time responding to those with overly negative attitudes toward the training.

Tenney, D. P. (1988, January). Age and airline accidents. *Journal of Psychology, 122*(1), 15-20.

Article type: Group Comparison Study
Keywords: Accidents; Civilian

This article examined the relationship of pilot experience to accidents. The author hypothesized that the more experienced pilot would be less likely to be involved in an accident than an inexperienced pilot and that the experienced pilots' accidents would be less severe than those involving less experienced pilots. Main effects were significant for age, aircraft time and total time, but no interactions were significant. Correlations were also computed between age, aircraft time and total time for fatal and nonfatal accidents, separately for young and old pilots (median split). Age was significantly correlated with total flying time (overall and for young and old subjects individually) and for both fatal and nonfatal accidents.

Thordsen, M. L., & Klein, G. A. (1991). Training implications of a team decision model. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 296-301).

Article type: Theoretical
Keywords: Coordination; Simulation

The authors discuss a cognitive model of teamwork. A team is thought of as an intelligent entity with cognitive processes similar to individuals. According to the authors, the team should show signs of maturation, so they looked to developmental psychology for information. The team over time will develop an identity. After this, the team members will be more sensitive to the needs of the team and to issues of metacognition. The team will also have a shift in focus, from the present to anticipation of the future. Several behavioral markers were then identified: (1) the use of anticipatory and confirmatory cues; (2) metacognitive functions; (3) understanding of the team's intent; and (4) achievement of a more effective time horizon. Positive and negative examples of behaviors for all these behavioral markers were identified. Ten aircrews were observed performing a tactical mission in a simulator. The authors found that it was important for crews to use mental simulation to imagine how events would occur and very few crews actually engaged in this behavior.

Torrance, E. P. (1953, August). The behavior of small groups under the stress conditions of "survival." *Paper presented at the Annual Meeting of the American Sociological Society.*

Article type: Single Group Study
Keywords: Coordination; Communication

The focus of the article was to develop concepts concerning group functioning under survival situations. Two hundred interviews with Air Force personnel shot down in WWII or Korea generated 1,000 critical incidents. The author suggests that *structure of the field* and *structure of*

the group are critical for survival. Lack of *structure of field* (random, trial and error behaviors; feelings of hopelessness) frequently occurs because of failure of the aircraft commander to communicate with other crewmembers. Cooperative actions and a common goal help to hold the group together. The lack of *structure of the group* results from lack of organizing efforts. "Unorganization" can be fatal; pulling together as a team is essential. Also, the power structure (and how power is wielded by the AC) is crucial, with communication to all other crewmembers seen as critical.

Traub, W. (1979). Flight crew selection at United Airlines. In G. E. Cooper, M. D. White, & J. L. Lauber (Eds.), *Resource management on the flight deck, Proceedings of a NASA/Industry Workshop* (pp. 61-75). San Francisco, CA: National Aeronautics and Space Administration.

Article type: Anecdotal/Descriptive
Keywords: Pilot Selection

Errors in pilot selection can be costly for an airline. As of 1979, United Airlines had selected more than 6,000 pilots from over 90,000 applicants. Minimum qualifications for employment have changed over the years, in part due to recent court cases. A committee was formed at United to evaluate their selection process and make recommendations for changes. They developed a philosophy of pilot selection that is currently being used by United. The goal is to select pilots who have basic flying skills, plus the appropriate attitude and personality traits that make an outstanding employee and ultimately, an outstanding Captain. A multi-step selection and training process was developed that contained: (1) a computerized pilot applicant tracking system; (2) a series of psychological tests (attitude and personality traits are listed by the authors); (3) a simulator evaluation; (4) an in-depth interview; (5) the air crew selection test, or STANINE; and (6) an extensive medical assessment. From all this information, a profile of each applicant is developed which is presented to a review board who makes the final decision. New-hires go through a three-phase training program, followed by a one-year probationary period, during which time, the pilot has several performance evaluations.

Vail, G. J., & Ekman, L. G. (1986). Pilot-error accidents: Male vs female. Special issue: Ergonomics in aviation. *Applied Ergonomics, 17(4),* 297-303.

Article type: Group Comparison Study
Keywords: Accidents

The authors analyzed general aviation accident records from files kept by the National Transportation Safety Board (NTSB) and compared pilot error-related accidents for both males and females during the years 1972-1981. This data set contained 37,862 males and 967 females who had each been involved in accidents that were caused by pilot error. Eight variables were included in the analyses: (1) certificate of pilot; (2) age of pilot; (3) total flight time of pilot at the time of the accident; (4) flight time in the type of aircraft in which the accident occurred; (5) category of flying at the time of the accident; (6) phase of operation; (7) degree of injury; and (8)

causes of accident. For all different certificates combined, the yearly accident rates for females was significantly lower than for males. In addition, the fatality rate for males was twice that for females. Females walked away without injury in larger percentages of the accidents than did males (although the differences were small, 67 percent for females, compared with 60 percent for males). Males of all ages have higher accident rates than females, especially during the ages of 15 to 29. When first starting to fly, male and female accident rates are nearly equal. As flight time increases, however, the rate of accidents increases, but more so for males. When looking at other phases of operation, females had lower accidents rates in all phases except while taxiing.

Vandermark, M. J. (1989). Aircrew team dynamics. In R. S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on aviation psychology* (pp. 640-645). Columbus, OH: Ohio State University.

Article type: Anecdotal/Descriptive
Keywords: Communication; Coordination; CRM; Training;

America West Airlines has expanded its crew resource management (CRM) training to include cabin crewmembers. This new training is called Aircrew Team Dynamics training (ATD). The author notes that this expansion should impact ground operations, as well as improve overall passenger relations. The ATD program is conducted over a three day period in a seminar/workshop format. Pilots attend all three days, while flight attendants attend only the first day. Brief descriptions of what occurs on each of the three days of training are provided. Reactions of the participants to the training have been positive and the author presents preliminary results from the first few seminars. The presence of female flight attendants as facilitators appears to result in a positive impact, however, at the time of this writing it was too soon to tell what the long-term effect is of having the flight attendants participate in CRM training.

Wagner, B., & Diehl, A. E. (1993, December). Your wingman is your co-pilot. *Flying Safety*, pp. 22-23.

Article type: Anecdotal/Descriptive
Keywords: Communication; Coordination; Combat/Fighter Pilots

The authors use anecdotal descriptions of single fighter jet aircraft accidents that likely could have been avoided had the flyers heeded crew resource management (CRM) concepts when flying in formation. They note that fighter pilots almost always fly in teams. This team also includes people other than those in the flight (e.g., ATC, AWACs, Command Post, etc.). The authors credit the National Safety Center as the first organization to institute CRM-like training for fighter pilots. Aircrew Coordination Training was introduced into their A-6 training wings in 1989. After introducing this program, their mishap rate dropped dramatically in the following years.

Warren, R. A., Hudy, J. J., & Gratzinger, P. (1991). The myth of adventuresome aviator. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 2* (pp. 700-705).

Article type: Single Group Study
Keywords: CRM; Personality; Stress

Pilots in the past have been thought of as being independent and adventuresome. This study explores whether this is true of pilots today. During the course of crew resource management (CRM) training, an inventory with twelve scales was administered to the pilots: humanistic-helpful, affiliation, approval, conventional, dependence, apprehension, oppositional, power, competition, perfectionism, achievement, self-actualization. The scales are grouped together into six clusters of scales. A person's type is derived from the clusters on which the person scores highest. Four main types were found in the sample: aggressive/autonomous, social/deferring, achievement oriented/social, mixed. There were no significant differences in the numbers of pilots who were in each group. Those who had the most experience flying reported the greatest amounts of stress. These results show that the pilot population in this sample is not characterized by a group of people who are adventuresome and independent.

Westerlund, E. J. (1991). The Pilot Judgment Styles Model: A new tool for training in decision-making. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 1061-1067).

Article type: Theoretical
Keywords: CRM; Personality

This is a model developed as part of crew resource management (CRM) training at Air Canada. Pilot perceptions were collected on their own flying operation. From these, two dimensions of judgment were to be used: (1) concern for system; and (2) concern for risk. System issues can include the flight mission, the overall organizational goals, attention to procedure and attention to the entire situation. Pilots can have a negative or positive attitude toward risk. These two dimensions were combined into a diagram with four quadrants. The grid was then rotated 45 degrees and encircled to come up with the Pilot Judgment Styles Model. The four judgment styles are listed with a description by the author, with the corresponding variations of each style: (1) "purposeful" (positive concern for both system and risk); (2) "traditional" (high on system, low on risk); (3) "adventurer" (high on risk, low on system); and (4) "crisis" (low on both system and risk).

White, G., & Baker, L. (1991). Inappropriate functioning of the cockpit dominance hierarchy as a factor is approach/landing accidents. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology, Volume 1* (pp. 706-711).

Article type: Anecdotal/Descriptive
Keywords: Accidents; CRM; Personality

According to the authors, many aviation accidents are a result of the prevalence of the dominance hierarchy propagated by the aviation industry. Examples are given of incidents in which the co-pilot had information which could have prevented it from occurring but did not tell the Captain. A review of ASRS data found similar results. The dominance hierarchy comes from the research Milgram has done on obedience. When entering the cockpit, the co-pilot enters the agentic state, which postulates that a person entering an authority system no longer views himself as acting out of his own purposes but rather comes to see himself as an agent for executing the wishes of another person. Conditions have rewarded the co-pilot for acting this way. There is little tension in the cockpit, reports are good and eventually the person will be promoted. There are consequences for speaking up in the cockpit which may adversely affect the co-pilot. Crew resource management (CRM) training programs must take into account this problem when structuring an intervention. The authors suggest that new standard operating procedures need to be proposed to deal with some of these situations.

White, L. C. (1987). Cockpit resource management training. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 123-131). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: CRM; LOFT; Training

The International Air Transport Association Flight Crew Training Subcommittee (FCTSC) conducted a survey of member airlines to see who had implemented resource management programs. Results were based on the responses of 24 airlines (or 17 percent of its membership). These airlines were located in Europe (8), the Americas (7), the Far East (6), the Middle East (2) and Africa (1). Over half of the airlines had some resource management training course and another 25 percent were planning to introduce CRM training. The content of the courses consisted of four major topics: decision-making, leadership, crew cooperation and communication. The methods of course presentation varied from airline to airline, but primarily included: line-oriented flight training (LOFT; 2), seminar/workshop (5), lecture (8), home study (2), audio visuals (1) and flight simulator (4). The lengths of the different courses ranged from three hours to sixteen days.

Wiener, E. L. (1985). Beyond the sterile cockpit. *Human Factors, 27,* 75-90.

Article type: Review
Keywords: Accidents

There is little consensus regarding whether automation will help or hinder pilot performance. Automation has even been cited as the *cause* of several accidents in recent years. Wiener lists eight reasons why there has been a push toward flight-deck automation and discusses each in detail. Reasons included: (1) the technology became available; (2) safety; (3) economy, reliability and maintenance; (4) workload reduction and certification of two-pilot transport aircraft; (5) more precise flight maneuvers and navigation; (6) display flexibility; (7) economy of cockpit space; and (8) special requirements of military missions. In addition, the nature of the flying task has changed over the years. According to the author, this is the result of two developments: (1) the increasing complexity of the environment in which pilots fly; and (2) a vast number of computer-based devices at the pilots' fingertips, replacing the demand for manual control and mental arithmetic. Pilots are concerned that in the future, decisions will be taken out of their hands and given to the computer. Many have remarked about feeling "along for the ride," or "out of the loop." This overuse of automation may result in a deterioration of piloting skills. Others have stated that automation increases the likelihood of larger, more potentially disastrous mistakes. The shooting down of KAL 007, a B-747 en route from Anchorage to Seoul, by Soviet interceptors is discussed in detail. The most widely accepted theory is that the crew made errors in inputting information into the automation system; and this caused them to drift off course. Some of these types of problems have been solved by automation, but new ones have also developed. Past human factors research demonstrates that humans are not good monitors, yet, we continue to expect people to do such a job and automation seems to be forcing this issue. Future cockpit designs need to take these and other similar issues into consideration. Tasks need to be allocated in such a manner as to not take the pilot "out of the loop." The monitoring role should be constructed in such a manner as to make piloting an aircraft an interactive, not passive task.

Wiener, E. L. (1993). Crew coordination and training in the advanced-technology cockpit. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 199-229). San Diego, CA: Academic Press.

Article type: Review
Keywords: CRM; Training Evaluation; Training Program Design

Pilot training for highly automated cockpits has only recently begun to be addressed by the airline community. Many line-oriented flight training (LOFT) scenarios now being used are not based on the characteristics and capabilities of the advanced technology cockpit (ATC). The author argues that future LOFT scenarios need to be more ATC-intensive. Crew resource management (CRM) programs that are now being instituted will help crews to deal effectively with the coordination problems that are present in flying high-technology aircraft. A difficulty that may be occurring is that many of the new cockpits are technology-driven instead of human-centered. Cockpits are being designed depending on what technology can provide, not based on what pilots

need. Research is being currently undertaken to study the problems of automation and to try and make it more human-centered. For this to occur, the crew must be supported by such things as policies, procedures, checklists and training (both CRM and LOFT). Over the years, technology has greatly enhanced the cockpit. Glass cockpits have integrated features that could have only been dreamed about years ago. With these advances, questions have been raised about problems that may occur. Thus there has been research in to what the so-called "automation problem" would be. As three person crews have been giving way to two person crews, there has also been discussion about the safety of phasing out the flight engineer position.

Research is just beginning to address the effects of advanced technology on crew coordination and it is assumed that it will have some impact. The evidence for this has come from five methodological sources: (1) non-systematic observation; (2) experimentation and systematic flight observation; (3) controlled simulator experimentation; (4) questionnaire/survey data; and (5) incidents from reporting system databases. Evidence from this type of research has demonstrated many interesting findings. For example, it is physically difficult for crewmembers to see what the other crewmembers are doing in these newer cockpits. This has an impact on communication because it is difficult for the pilots to monitor what the other is doing. Also, research has shown that there is less communication during low workload periods in ATCs, as compared to more conventional cockpits, but that there is more communication during high workload periods. These shifts in communication patterns will need to be addressed by CRM training.

Wilhelm, J. A. (1991). Crewmember and instructor evaluations of line-oriented flight training. In R. S. Jensen (Ed.) *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 362-367). Columbus: The Ohio State University.

Article type: Validation
Keywords: LOFT; Training Program Evaluation

The author reviews the results of LOFT Surveys which have been administered for four airlines. The usefulness of LOFT training to its participants was rated six on a seven point scale. One European carrier participated in this study and they distributed detailed information regarding possible scenarios that were to be used in training. Ratings from this carrier were more positive about the training value of LOFT. High ratings were also given for a scenario's value with respect to crew coordination skills and technical skills, but varied across organizations. The lowest ratings were given for how much was learned which will be used on the line (i.e., transfer-of-training), but were still fairly high. No relationship was found between workload and training outcomes. About 69 percent of the respondents stated that the workload was about right in the scenario. The highest ratings for the instructors were for "knowledge and helpfulness." Ratings were much lower for the importance of feedback obtained from the videotape. In all organizations, individuals rated the crews' performance as being higher than individual performance. Students rated themselves higher than the instructor did. Self-ratings of crews on CRM behavior performance varied across organizations, depending upon which areas were stressed by the training and recurrent programs. There were large differences between the instructors'

evaluations of these behaviors from the individuals' evaluations. Good instructor ratings correlated highly with good scenario ratings, with good CRM behaviors and overall ratings of LOFT. There were also differences in scenario ratings between fleets in an organization.

Wilhelm, J. A., Helmreich, R. L., & Gregorich, S. (1989). *LOFT from the airman's perspective* (NASA/UT Tech. Report No. 89-3). Moffett Field, CA: NASA Ames Research Center.

Article type: Group Comparison Study
Keywords: CRM; LOFT

The line-oriented flight training (LOFT) Survey was administered at two carriers during 1988 and 1989, as airmen exited a LOFT training session. One organization had just initiated its LOFT program, while the second organization had used LOFT for some time. Ratings were given on a seven point scale. Both organizations rated LOFT training as useful, although Organization 2 (the one that had LOFT in place for a longer period of time) provided slightly higher ratings than Organization 1. Engineers rated the training most useful and first officers rated it least useful. The next section asked about the realism of the scenarios. The scenarios were rated as being "moderately" realistic, of average difficulty and that the training was beneficial for both crew resource management (CRM) skills (mean = 5.87) and technical skills (mean = 5.43). Ratings of crew performance were higher than ratings for individual performance. This varied by position, with Captains giving the crew the highest ratings and Engineers giving the lowest. Ratings of how well the crew performed in the simulation were correlated with crew self-reports of good CRM behaviors. Instructors were rated as being knowledgeable and helpful. The debriefing was rated as being helpful, but viewing the videotape was not rated as highly. An Appendix contains responses of participants to the open ended questions broken down by organization and crew position.

Yamamori, H. (1987). Optimum culture in the cockpit. In H. W. Orlady & H. C. Foushee (Eds.), *Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop* (NASA-CP-2455) (pp. 75-80). San Francisco, CA.

Article type: Anecdotal/Descriptive
Keywords: CRM; foreign;

The author discusses similarities between the Japanese and the U.S. airline industry in terms of accident rates, pilot error and crew coordination. Some cultural differences were noted. For example, American pilots tend to be very task-oriented, whereas Japanese pilots tend to be more group-oriented. American pilots were also reluctant to listen to another's opinion, whereas, the Japanese pilots frequently did not voice an opinion, even when a dissenting vote was necessary. Thus the author concludes that the cockpit should be considered "culture-free," because neither of these styles is appropriate in the cockpit. The most important quality of any program is to find the approach to the goal of ultimate safety.

Yamamori, H., & Mito, T. (1993). Keeping CRM is keeping the flight safe. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 399-420). San Diego, CA: Academic Press.

Article type: Review
Keywords: CRM; Training Evaluation; Training Program Design

Japan Airlines (JAL) developed ten safety fundamentals which are a part of their safety campaign. The top item, "keeping CRM is keeping the flight safe," was deemed to be the most essential factor in safety. JAL has also developed it's own crew resource management (CRM) seminar and over 2,300 crewmembers have completed the course. The course has three main objectives: (1) getting the crew to work together as a team under all circumstances; (2) getting crewmembers to understand how they communicate with one another; and (3) understanding the process of decision-making and CRM. The JAL CRM seminar consists of three distinct parts. In the first part, participants rate eight items concerning cockpit management, behavior and performance. Participants then discuss what is the best way to solve problems with other crewmembers. Forty situations are presented, each having two possible alternatives. Participants weight the alternatives against each other and select the alternative representing the more effective way to manage. In the last portion of the seminar, participants compare their pre- and post-seminar data on performance values and Grid styles. Crews also compare their ratings of their own behavior.