DOT/FAA/EM-81/16
DOT/FAA/CT-82/24

Pilot Performance Measurement: An Annotated Bibliography

Jacqueline T. Rehmann

Prepared By
FAA Technical Center
Atlantic City Airport, N.J. 08405

December 1982

Final Report

U.S. Department of Transportation
Federal Aviation Administration

Office of Systems Engineering Management
Washington, D.C. 20590

NOTICE

Technical Report Documentation Page

1. Report No. DOT/FAA/EM-81/16	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Pilot Performance Measurement: An Annotated Bibliography		5. Report Date December 1982
		6. Performing Organization Code
7. Author(s) Jacqueline T. Rehmann		8. Performing Organization Report No. DOT/FAA/CT-82/24
9. Performing Organization Name and Address Federal Aviation Administration Technical Center Atlantic City Airport, New Jersey 08405		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 161-301-150
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Aviation Administration Office of Systems Engineering Management Washington, D.C. 20590		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

One-hundred references relating to pilot performance were selected and summarized as part of the Cockpit Display of Traffic Information (CDTI) studies currently being conducted by the Federal Aviation Administration Technical Center in Atlantic City, N. J. A comprehensive search of the scientific literature was conducted using several sources, including books, scientific journals, proceedings of technical meetings, and computerized information retrieval.

Specific topics covered in this annotated bibliography, as they relate to pilot performance, include behaviors, measures, and instruments for performance evaluation in actual and simulated environments, task taxonomies and modeling, physiological measurement, and automated performance measures.

17. Key Words: Cockpit; Subjective/Objective Measurement; Automated Performance Measurement; Pilot Workload; Task Taxonomy; Inflight/Simulated Measurement; Performance Criteria; Factor Analysis; Performance Assessment; Data Collection; Stress Factors	18. Distribution Statement: Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	58	

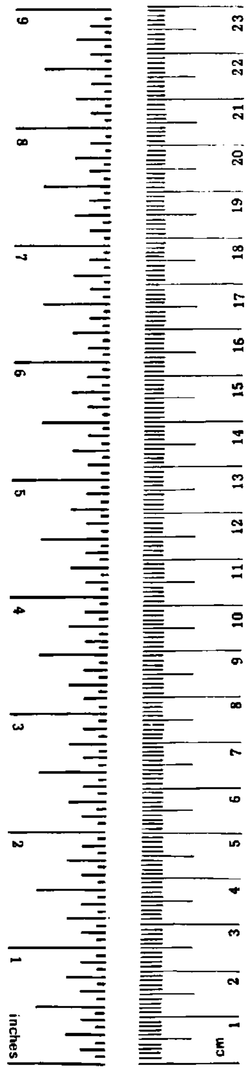
Form DOT F 1700.7 (8–72) Reproduction of completed page authorized

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
	LENGTH			
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
	AREA			
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
	MASS (weight)			
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
	VOLUME			
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
	TEMPERATURE (exact)			
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
	LENGTH			
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
	AREA			
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
	MASS (weight)			
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
	VOLUME			
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
	TEMPERATURE (exact)			
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F

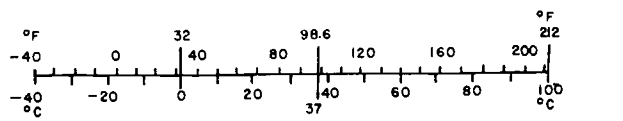


TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

The Cockpit Display of Traffic Information (CDTI) is a concept currently under investigation at the FAA Technical Center, Atlantic City, N. J. It assumes that traffic information can be made available in a pictorial display to the pilot and flight crew. Many benefits have been hypothesized, but limitations may also exist, and the effect of a CDTI on pilot performance is a major consideration.

One of the objectives of the General Aviation Cockpit studies involving CDTI is to assess pilot performance. As a result, an extensive review of the literature was conducted to evaluate pilot performance measurement research, and apply the findings to the development of a Pilot Performance Index (PPI).

METHODOLOGY

A comprehensive review of the literature related to pilot performance measurement was undertaken. Books, scientific journals, proceedings of technical meetings, and technical reports were consulted, and a computerized literature search was conducted to locate relevant sources.

Specific topics covered in this report, as they relate to pilot performance, include behaviors, measures, and instruments for performance evaluation in actual and simulated environments, task taxonomy development and modeling, physiological measurement used in conjunction with other performance measures, and automated performance assessment.

The criteria for selection of sources included in this survey are:

1. Performance - the emphasis is on the evaluation of pilot performance in the cockpit.

2. Cockpit (or pilot) - here again, the emphasis is on pilot performance versus operator performance, unless the information appeared relevant to an understanding of pilot performance.

3. Background - articles which provide information regarding the concept of pilot performance, overviews of the literature, etc., were included.

4. Subjects - work describing experiments with normal adults, preferably pilots, is included. Also, articles on pilot personality, including results of psychological profiles, are included.

5. Application - documents relating to pilot performance or aircrew in actual flight or simulated conditions are included.

6. Types of Measures - measures that are of interest include performance evaluation in actual or simulated environments, task taxonomy development, and automated performance assessment.

The sources contained in this bibliography are listed alphabetically by the author's last name. In addition, a subject index is included for the user's convenience.

RESULTS

The result of the literature search is included here, in the form of an annotated bibliography. One-hundred summaries are included all of which relate, in some way, to the areas outlined in the methodology section. In addition, articles dealing with multivariate statistical techniques, useful when reducing data from performance measures, are included.

Some consensus among the authors cited can be noted. For example, it is generally agreed that as greater automation is built into the cockpit, and as the complexity and difficulty of the pilots' task increase, conscientious efforts are needed to develop methods to evaluate and monitor an individual's performance. As the pilot becomes more involved in monitoring displays and operating controls, safety may be compromised when and if information processing and control actions go beyond the pilot's capability.

No fully acceptable or scientifically validated approach dedicated to the measurement of pilot performance exists (Grodsky, 1967). With the advent of new cockpit configurations, the problem becomes more complicated. The pilot's task is both complex and multidimensional, which makes the development of a single overall scale for evaluation of performance difficult. Additionally, intra-pilot variability in the performance of flight skills further complicates measurement (Farrell, 1973).

Several authors feel that the choice of a performance measurement method is dependent upon the question asked about the individual performance as well as the state of development of the system under consideration (Chiles, 1967; Grodsky, 1967; Koch, 1964; Farrell, 1973). Therefore, the applicability of any previous methodology and taxonomy may be limited.

Other considerations that arise in performance measurement are (1) the question of criterion, that is, to determine what information is necessary and how it will be used to provide valid assessment, and (2) the general procedure to be used for data collection - simulation versus inflight testing, to name two. A brief summary of the performance measurement techniques found in the literature is included, and the annotations follow.

Pilot performance measurement can be divided into two broad categories, subjective and objective. Subjective measurement refers to the evaluation of skill by an observer, for example, an instructor pilot, who is usually highly trained and experienced. Objective performance measures evaluate, more directly, how well the pilot performs on the basis of a predetermined objective scale. A major drawback in the use of subjective techniques is a general lack of correlation between raters on specific criteria that constitutes successful performance (Knoop, 1973). Objective measures, on the other hand, are not subject to rater variability, providing acceptable tolerance limits are established to account for pilot variability. Measures in the objective category include automated performance measurement, component/total task performance, residual attention, tracking tasks, dual-task performance, task taxonomy development, and modeling techniques.

The area of automated performance measurement represents a relatively new development in the area of performance measurement and can be attributed to the data collection potential available in simulators. It has as it's major advantage the fact that aircrew performance, by its continuously changing nature, requires automated data collection techniques. Several studies in this area are based on the assumption that flying performance is directly reflected in certain parameters (for example, airspeed and altitude) maintained close to some predefined criterion value (Henry, et al, 1979; Hill, Eddowes, and Knoop, 1973; Knoop and Welde, 1973). In many of the systems described, critical variables of aircraft control, navigation, and radio procedures are measured, and errors, i.e., departure of critical variables beyond the specified tolerance level, are reported. It has been shown that the "diagnostics" may not relate to pilot performance per se, but that interpretation of the error, if it takes into account the pattern of errors within segments and throughout the maneuver, reveals more meaningful data (Childs, 1979; McDowell, 1975). The major caution, as revealed through the literature, is that automated performance measurement systems be subjected to empirical utility testing, with particular attention devoted to specified tolerance levels for each measure.

Component-total-task performance measurement refers to the study of relationships among part-task and whole-task proficiency (Freedle, Zavala, and Fleishman, 1968). Results of research in this area indicate that the best predictors of total-task and multiple control subtasks were other multiple control subtasks. Prior practice on combinations of components yields a better prediction of total task performance than does prior practice on single components. In support of the "common ability" paradigm, it was noted that proficiency in unpracticed components was increased as a function of practice on other components (Fleishman, 1968). Implications for application in total task performance evaluation are summarized in the literature cited, and include the prediction of total task performance from prior knowledge of an individual's performance on task components.

Residual attention studies are often connected with pilot workload studies, and refers to the extra attention capacity which the pilot has, in addition to the workload of performing routine flight tasks (Damos, 1972; Damos and Lintern, 1979). Most studies on residual attention use a dual-task approach, with one task defined as primary, the other secondary. Residual attention research has application in the pilot performance area because it is often the reserve capacity that is used to deal with deteriorating situations inflight. An excellent review of the literature in this area is found in the study by Damos and Lintern, 1979, entitled "Comparison of Single and Dual-Task Measures to Predict Pilot Performance".

The use of tracking tasks in performance research has shown that compensatory tracking affords stable measures of human performance (McRuer, 1973; Damos and Lintern, 1979; Gerathewohl, Chiles, and Thackray, 1975; Gopher and North, 1974). The most important aspect is that the class of situations in closed loop control of aircraft are compensatory tasks in which a pilot acts on displayed error quantities between desired command inputs and comparable vehicle output motions, producing control actions (Krendel and Bloom, 1959). Tracking tasks have wide application in performance measurement and provide a number of models of pilot performance. Dual-task performance refers to the simultaneous performance of more than one task. Application of dual-task performance research includes the use of divided attention during the performance of several tasks. Results reveal a correlation between a dual-task test and the performance of a flight check.

Task taxonomy, in the general sense, represents a means of classifying objects and phenomena in such a way that useful relationships among them are established. It also involves the rules and principles concerned with the classification (Berry, 1980; Companion and Corso, 1977; Koch, 1967; Meyer, Laveson, and Weisman, 1975). Task taxonomy is especially useful in the aviation environment and can be used to describe the functions performed by the human component (behavioral taxonomy) and determine the relative position of each task on a certain dimension to the overall job ("true" task taxonomy). Rules and rationale for taxonomic classification vary according to the particular measurement problem, and one rule of thumb pervades the literature-simplicity in organization of the taxonomy (Berliner, Angell, and Shearer, 1964). Other considerations in task taxonomy development include the following criteria:

(1) Task element definitions should be objectively defined.
(2) Categories should be mutually exclusive.
(3) Taxonomy should be as exhaustive as possible.
(4) Taxonomy should permit analysis at different levels of details.

One example from the literature shows a taxonomic outline used to classify student pilot errors (Shannon, 1980). It is divided into discrete and continuous operations. Discrete operations, defined as tasks involving individually distinct movements or mediating responses, are further broken down into procedural, anticipation/planning, communication, and monitor roles. Each category is succinctly described and checkpoints noted. This minimizes interpretation error when the taxonomy is applied. Continuous operations, tasks involving multidimensional tracking responses, includes pitch axis control, roll axis control, yaw axis control, thrust axis control, and brake control. Again, each is described and specific parameters noted.

One of the major advantages of the taxonomic approach is that it remains creative and innovative, despite guidelines governing its development and use. Further, to the extent that the task taxonomy is defined and used, the task analysis can be objective, repeatable, and complete (Berry, 1980).

Models of many different kinds have been used for analyzing performance. They include those that model human behavior, models of environment, information, and machine behavior while observing the responses of the pilot and models of system task networks.

Control theory models are discussed and provide one example of a model of human behavior. First, the system, given certain characteristics, is described and human behavior within that system is modeled. The model, in this context, should include the various psychophysical limitations inherent in the human operator, i.e. time-delay, neuromotor, dynamic, and controller remnant. Advantages to a modeling approach include a high rate of agreement between theoretical and measured quantities (Kleinman, Baron, and Levison, 1970).

A comprehensive survey of man-machine models of human performance, based on information theory, control theory, and decision theory, is contained in the text Man-Machine Systems: Information, Control, and Decision Models of Human Performance by Thomas B. Sheridan and William R. Ferrell (1974). Information models covers probability estimation, Bayesian probability, updating, and human performance models of factors-like judgment and reaction time. Control models include manual control subsystems, linear control models, and system identification and modeling in both the time and frequency domain, to name a few. Decision models examine riskless and risky decisions, and the subjectively expected utility model, including signal detection. With the use of these models, it becomes possible to describe the behavior of the machine and the human components of a system in a single language appropriate to both.

Other factors which are considered in the literature, as they relate to pilot performance, are outlined in table 1. These factors are noise, age, stress, and personality. All influence performance in some measure. The investigators are named, and a summary of their work appears in the annotations.

CONCLUSIONS

The question arises as to which is the best situation in which to measure pilot performance. The two major categories are described as inflight and simulation. Inflight implies the measurement of performance on an actual system and on mission tasks in the actual environment (cockpit). Simulation duplicates the essence of the system under investigation and uses actual trained operators (pilots).

Inflight measurement is generally considered the most valid (Alluisi, 1967; Billings, Gerke, and Wicke, 1975). However, with the data collection systems available in simulators, the range of data collection potential has increased. Other benefits of simulation include the low cost of operation, once the initial expense of purchasing the simulator has been absorbed; safety, equipment reliability, ease of duplication of conditions, and relative ease of measurement (Obermayer and Vreuls, 1974; Mudd, 1968; Alluisi, 1967; Berliner, Angell, and Shearer, 1964). The question of validity remains, however, and extrapolation from simulated to actual flight must be made with considerable caution.

The question of arousal has been discussed, and some researchers feel that data from simulators and actual flight may never correspond, since different levels of arousal may be responsible for differences observed between the two (Grodsky, 1967; Alluisi, 1967). A pilot knows, for example, that his life is not endangered when attempting difficult maneuvers in a simulator, as it may be in flight.

The research summarized in the following pages covers a wide range of situations and details the type of results obtained. The final decision of performance measurement techniques, as well as the determination of an appropriate situation for data collection, remains with the researcher who must consider his own specific objectives, amount of time, money, and subjects available.

TABLE 1

OTHER FACTORS AFFECTING PERFORMANCE

FACTORS	COMMENTS	INVESTIGATORS
NOISE	Decreased performance in noise was not found, but an increased level of heart rate indicated the stress. Unevenly spaced and periodic intermittent noise on complex psychomotor task performance causes decrement.	Eschenbrenner, 1971 Strasser, 1977
	Reduction in spare mental capacity is a function of aversiveness of the noise stressor.	Finkelman, Glass, 1970
AGE	Reviewed the literature related to age differences as measured by standardizing tests of sensory, perceptual, mental, cognitive, and neurophysiological functions and objective assessment of personality in Part I. Isolated 114 variables important to successful pilot performance in Part II. Compared qualitative assessment based on subjective ratings and quantitative assessment based on recordings of pilot actions and aircraft response.	Gerathewohl, 1977; 1978
STRESS	Intellectual stress produces impaired flying skill which is predictable; emotional stress produces substantially more severe but fluctuating degrees of impairment of unpredictable duration.	Howard, 1975
PERSONALITY FACTORS, PSYCHOLOGICAL FACTORS	Male pilots tend to be "active-masculine, adventurous" (based on Edwards Personal Preference Schedule, and General Aviation psycho-social inventory). Female pilot profiles closely resemble "norm" for adult male personality traits. Several behaviors have been identified as relevant to the flying task and measured individually have not been used collectively in any attempt to predict aircrew performance.	North, Griffin, 1977 Novello, Youssef, 1974 Ware, 1964

BIBLIOGRAPHY

1. Alluisi, E. A., Methodology in the Use of Synthetic Tasks to Assess Complex Performance. Human Factors, 1967, 9, 375-384.

The task of performance measurement is a difficult one, and the results are often "ephemeral". The author discusses general difficulties in trying to design a measurement system for multiple-task performance assessment. The first problem mentioned is the criterion: that is, what information is necessary, and how would this information be used to provide a valid assessment of current performance? The second major problem discussed is the general procedure to be used - laboratory or simulation techniques versus full operational testing. Considerations in each method are discussed.

The multiple-task performance battery approach is discussed, and the author contends that this approach can provide levels of face validity adequate to maintain subject motivation while concurrently allowing the identification of changes in specific performance functions.

2. Altman, J. W., Improvements Needed in a Central Store of Human Performance Data. Human Factors, 1964, 6, 681-686.

The author suggests seven major improvements for central stores of human performance data:

(1) Update the current data stores.
(2) Improve schemes for classifying behavioral data.
(3) Develop more refined definitions and categories of human error.
(4) Develop more refined methods for scaling laboratory to operational performance.
(5) Add a wider variety of human activities to a behavioral taxonomy.
(6) Extend the range over which modifying factors to performance are studied.
(7) Provide a way that data stores can be applied to conceptual phases of system design.

Each of these areas is described in greater detail, and recommendations are made.

3. Baron, S., and Levison, W. H., An Optimal Control Methodology for Analyzing the Effects of Display Parameters on Performance and Workload in Manual Flight Control. IEEE Transactions of Systems, Man and Cybernetics, 1975, 5, 423-430.

The article describes a methodology which allows one to predict the effect of changes in display variables on performance and workload. Specifically, the methodology is applied to vertical situation displays and command information displays are examined in this study and compared for approach probability and workload questions.

4. Berliner, D. C., Angell, D., and Shearer, J. W., Behaviors, Measures, and Instruments for Performance Evaluation in Simulated Environments. Proceedings of the Symposium and Workshop on the Quantification of Human Performance. August 1964, 277-296.

This paper discusses the planning factors which are involved in the utilization of simulation equipment for objective proficiency measurement.

Four major areas are discussed as they relate to proficiency measurement. They are:

(1) The need for measurement capabilities in simulators including
- behavior quantification
- use of performance measures capability of simulators

(2) The development of criteria including
- selecting behaviors
- specifying performance standards
- records and scores

(3) Parameters of performance evaluation including
- development of a classification of behaviors (the authors present one which includes and subdivides perceptual, mediational communication, and motor processes.)
- choice of relevant measures
- types of instruments to be used

(4) A way to conceptualize the interrelationships among the parameters

The authors present a matrix of behaviors, measures, and instruments relevant to performance evaluation in order to illustrate not only the relationships but also possible combinations of classes of behavior, classes of appropriate measures, and types of equipment which can be used.

It is concluded that simulation devices and their data collection systems are not used to their fullest capacity.

5. Berry, G. L., Task Taxonomies and Modeling for System Performance Prediction. Proceedings of the Human Factors Society 24th Annual Meeting, 1980, 425-429.

Three models for operator performance prediction are briefly described. They are:

Category 1. Compares models which represent some facet of human behavior and explores characteristics of those models in a particular system.

Category 2. Involves simulation of the system, with the actual operator included, to respond to simulated system conditions.

Category 3. Analyses the behavior of a system, describing the tasks necessary to accomplish system goal(s) and examining the interaction of system components.

The advantages and disadvantages of each model group are discussed. The author then discusses the use of task analysis to abstract a task-oriented description from a real system. She found that a task-oriented network model of the same system gave significantly different results when based on different taxonomies. The recommendation for additional research is made to examine available taxonomies and determine which produces the maximum results, and under what conditions certain taxonomies can best be used.

6. Billings, C. E., Gerke, R. J., and Wick, R. L., Comparisons of Pilot Performance in Simulated and Actual Flight. <u>Aviation, Space and Environmental Medicine</u>, 1975, 46, 304-308.

In order to compare flight data from actual flight and a simulator, an experiment was conducted using five experienced professional pilots. The subjects performed instrument landing system approaches under simulated instrument flight conditions in a Cessna 172 and a Link-Singer GAT-I simulator, while under the influence of orally administered secobarbitol. The parameters evaluated were tracking performance in two axes and airspeed control.

The findings revealed that error and root mean square variability were about half as much in the simulator as in the airplane. Also, observed data were more strongly associated with the drug level in the simulator. The data are useful for studying the effects of mild stress on pilot performance.

7. Bremner, F. J., McKenzie, R. E., and Eddy, D. R., The Identification of Psychophysiologic Correlates of Cognitive Processing. Air Force Office of Scientific Research, Bolling AFB, DC, AFOSR-TR-81-0117, October 1980.

The efficiency of Electroencephalogram (EEG) measures in assessing the level of cognitive processing in aircraft pilots was evaluated. Several studies with different subjects under several conditions demonstrated the relationship between EEG and cognitive states. Results indicated that frontal electrodes alone were not adequate to gather the most significant data. Occipital electrodes proved to be the best statistical indicator of differences in cognitive states.

The statistics for analysing various treatment effects on EEG while subjects are used as their own control was called the INNOVATE (Individual Normal Variant Analysis Technique), and was developed for the purposes of this evaluation.

8. Brictson, Dr. C. A., Operational Measures of Pilot Performance During Final Approach to Carrier Landing. Proceedings of the AGARD Conference on Measurement of Aircrew Performance, AGARD-CP-56, May 1969, 7-1-7-5.

Measures of pilot performance during night carrier landings were found to differ statistically and practically from daytime performance in terms of altitude control. Night approaches were characterized by more altitude variability, a larger percentage of approaches below glide slope and higher bolter rates compared with day approaches flown by the same pilots. Practical application of the performance data is discussed in terms of pilot training, visual landing aids, and aviation safety. Empirical landing performance criteria are developed from the data and used to predict the probability of landing success as a function of deviations in final approach performance. (ABSTRACT)

9. Brictson, Dr. C. A., Ciaverelli, A. P., and Wulfeck, J. W., Operational Measures of Aircraft Carrier Landing System Performance. Human Factors, 1969, 11, 281-290.

Measures of final landing approach performance during day and night aircraft carrier recovery, including altitude error from glide slope, lateral error from centerline, sink speed at the ramp, approach speed at the ramp, arrestment wire, boarding rate, and bolter rate are analyzed and interpreted. The authors found that the major difference between day and night carrier approach performance was in altitude error control. With a diminished visual field at night, final approach performance showed more altitude variability, a greater percentage of aircraft below the glide scope, higher bolter rates, and less landing success as measured by overall boarding rate.

The point, then, is to determine if existing night approach performance is acceptable, and if not, emphasis must be placed on improving the presentation of visual cues in visual landing aids. Suggestions for future areas of study are made.

10. Brictson, Dr. C. A., McHugh, L. C., and Naitoh, Dr. P., Prediction of Pilot Performance: Biochemical and Sleep Mood Correlates under High Workload Conditions. Proceedings of the AGARD Conference On Simulation and Study of High Workload Conditions, AGARD-CP-146, October 1974, A13-1-A13-8.

The article discusses a preliminary longitudinal study of the factors affecting the carrier landing performance of naval aviators under high workload conditions. Biochemical, emotional, and experimental data were gathered from 26 F-4J pilots. The most significant finding was the effect of high cumulative workload on performance prediction. Under conditions of zero and moderate cumulative workload, more than 70 percent of the variability in individual landing performance could be accounted for by the predictor variable. Under conditions of high cumulative workload, only 40 percent of the individual landing performance could be accounted for by the predictor variables. Also, the pilot's assessment of workload became less predictable under high workload conditions and suggested that either the pilot used a different technique to assess workload under such conditions or marked central nervous activity during high workload interfered with the pilot's assessment.

11. Carter, V. E., Development of Automated Performance Measures for Introductory Air Combat Maneuvers. Proceedings of the Human Factors Society 21st Annual Meeting, 1977, 435-439.

The paper describes a study which was undertaken to develop an automated performance measurement technique. Seven F-4J student pilots and six F-4J instructor pilots flew 16 barrel roll attacks in an air combat simulator, mechanized to simulate the F-4J.

Five-hundred and twenty-two objective measures and thirty-five subjective performance measures were obtained. A master automated performance measurement tape was made which contained the subjective and objective measures obtained in each run. The format of data was such that it could be analyzed statistically. The outcome of the study was effective in that a preliminary set of automated measures, useful in augmenting instructor evaluation of performance, was identified.

12. Childs, J. M., Development of an Objective Grading System Along With Procedures and Aids for its Effective Implementation in Flight. Research Memorandum, Canyon Research Group, Inc., Fort Rucker, AL. May 1979.

The author describes the characteristics and test of two inflight scoring procedures designed to meet two specific requirements: (1) minimal data collection and (2) objective scores for Initial Entry Rotary Wing (IERW) student performance on basic instrument maneuvers. They were criterion-referenced procedures, employing different performance criteria, sampling techniques, and scoring algorithms. Tests in the UH-1 simulator assessed the potential value of various characteristics within each procedure for meeting requirements. Results provide a general indication of characteristics best discriminating student proficiency across training days.

13. Childs, J. M., Time and Error Measures of Human Performance: A Note on Bradley's Optimal-Pessimal Paradox. Human Factors, 1980, 22, 113-117.

The author attempts to demonstrate that Bradley's paradox, although theoretically sound, is grounded upon "assumptions that are empirically untenable." He gives alternatives to Bradley's recommendations when optimal-pessimal conditions exist.

Bradley's paradox, simply stated, is that if there are few errors, easy tasks, or ideal conditions, little variability results among time scores, and individual differences are masked. With numerous errors, difficult tasks, or poor conditions, diffused time score distributions result. Further, when performance conditions are optimal, parametric statistical tests tend to be pessimal and vice versa. The author suggests, as alternatives, the use of multivariate techniques, transformations, or repeated measures design.

14. Childs, J. M., The Identification and Measurement of Critical IERW Performance Variables. Research Memorandum, Canyon Research Group, Inc., Fort Rucker, AL, March 1979.

The purpose of the report is to describe the results of studies designed to identify and quantify critical Initial Entry Rotary Wing (IERW) performance variables in order to develop a method for assessing student flight performance objectively. Basic instrument maneuvers were used in the study. In addition, a review of the relevant documentation, personal interviews with instructor pilots, a questionnaire survey of students, and observation flights in UH-1 aircraft and flight simulator were used. Data from these measures were used to identify critical performance variables, with the results being descriptions of critical flight variables; airspeed, altitude, heading, trim, pitch and bank recovery, pitch and bank angle, control reversal, and rate of change coordination. The results indicate that the critical variables can be used to assess the quality of basic maneuvers objectively, and implications for further use of the findings are discussed.

15. Childs, J. M., An Analytic Technique for Identifying Inflight Performance Criteria. Working Paper, Canyon Research Group, Inc., Fort Rucker, AL., April 1979.

The author states that "inflight performance assessment should be focused upon instrument-referenced aircraft configurations rather than the pilot control inputs leading to that configuration." This statement forms the basis for the discussion of an analytic technique based on the work of M. L. Ritchie (1960), which demonstrates that inflight performance assessment should be focused upon the instrument-referenced aircraft configuration because it represents a valid index of pilot control skill. The technique illustrates the interrelations existing among pilot information processing variables, control inputs, and resulting aircraft configuration for several Initial Entry Rotary Wing (IERW) maneuvers included in the training program of Fort Rucker, Alabama.

16. Childs, J. M., Hennessy, R. T., Hockenberger, R. L., Barneby, S. F., Vreuls, Dr. D., Siering, G. D., and Van Loo, J. A., Summary Report: Human Factors Research In Aircrew Training Performance Enhancement, U. S. Army Research Institute for the Behavioral and Social Sciences, FtR-10-79, June 1979.

This report contains summaries of five reports having to do primarily with, (a) development of measurement techniques for assessing the performance of helicopter pilots in aircraft and simulators, (b) the development of the design requirements for an automated performance measurement and grading system for the UH-1 Flight Simulator, and (c) development of the requirements for a computer-based training management system.

The first study, which is directed toward the identification and measurement of critical performance variables, arose from the need to develop a set of objective and standardized assessment techniques. The methods used to establish a baseline were a review of relevant performance assessment documentation, consultation with instructors and operations officers, observation flights (in actual aircraft and simulators), development and administration of student questionnaires, and a detailed situation-specific maneuver analysis. Critical performance variables defined from the B1 maneuver analysis include airspeed, altitude, heading, trim, pitch and bank recovery, pitch and bank angle, control or airspeed reversals, and rate of change coordination. The author (Childs) concluded the following: the critical IERW (Initial Entry Rotary Wing) performance variables identified were observable instrument referenced flight variables and these variables make sequential, objective inflight evaluation possible by comparing observed configurations of the aircraft with desired configurations.

In the second report, Development of Procedures and Techniques for Inflight Performance Assessment, results describe an inflight scoring procedure for evaluating straight and level flight, climbs and descents, climbing and descending turns, standard rate turns, and steep turns. The instructor/observer records initial conditions to establish a baseline prior to the beginning of the maneuver; observed values are recorded at time-based sampling points during the maneuver. An objective evaluation of inflight performance is made possible by calculating the absolute deviation of observed values from desired values.

Finally, the purpose of the work described in Design Requirements for an Automated Performance Measurement and Grading System for the UH-1 Flight Simulator, was to design an automated performance measurement and grading system for instrument flight training in UH-1 flight simulators. The author describes the basic hardware and software that was used and the basic scoring system, including segment scores, composite maneuver grade, and diagnostic error messages to tell the pilot/ observer which elements of performance were within what performance bands during each segment of the instrument maneuver flown. The author recommends that the automated performance measurement system be subjected to empirical utility testing, paying particular attention to specified tolerance levels for each measure. In addition, it is suggested that a handbook, describing the purpose, function, use, and interpretation of the output of the performance measurement and grading system, be developed along with the implementation of the system.

17. Chiles, W. D., Methodology in the Assessment of Complex Performance: Discussion and Conclusions. Human Factors, 1967, 9, 385-392.

The purpose of this paper is to summarize the discussion which followed the presentation of papers for the conference entitled "Assessment of Complex Operator Performance". Those papers are included in this bibliography (ref. Alluisi; Christensen and Mills; Fleishman; Grodsky; Parker; 1967). Several statements were extracted from tape recordings of the conference and subjected to a semantic differential scale (agree vs. disagree) by 15 evaluators who participated in the conference. Four major areas were judged to be the most important:

(1) Criterion problem - There is a lack of both theoretical and empirical bases for specifying criteria for performance assessment.

(2) Task Taxonomy - There is a need for a taxonomy of performance functions.

(3) The reliability of Measures - The emphasis in this areas was on the expected magnitude of the statistical "error term" in relation to the expected magnitude of the "treatment effect". This is especially important when the subjects are highly selected and highly trained.

(4) Role of Face Validity - Face validity is most important when the subjects are operational personnel. Not stressed, but equally important, is "relevance to the real world."

The general results of the conference, in addition to the above, provide guidelines for performance measurement that have previously been undefined.

18. Chiles, W. D., Alluisi, E. A., and Adams, O. S., Work Schedules and Performance During Confinement. Human Factors, 1968, 10, 143-196.

Thirteen experiments spanning eight years were carried out to evaluate the effects on performance of several work/rest schedules during confinement. Four specific areas were addressed:

(1) How many men are required to maintain around-the-clock operations for a given number of work positions?
(2) What is the optimum duration of the duty period?
(3) How many days can such performance be maintained without decrement?
(4) What are the optimum work/rest schedules, with the least impact on man's performance reserves?

The approach used was a standard battery of performance tasks, and findings revealed reliable, face-valid results for assessing complex operator performance. Specific findings include the following: "a man can work 12 hours per day on a 4-hour work/ 4-hour rest schedule for periods of at least 30 days. For shorter periods, a person can work 16 hours per day on a 4/2 schedule but at a significant loss to his reserves. Circadian periodicities are found in psycho-physiological functions paralleled by similar periodicities in performance functions, the latter being subject to modification by special motivational instructions."

19. Chiles, W. D., and Alluisi, E. A., On The Specification of Operator or Occupational Workload with Performance Measurement Methods. Human Factors, 1979, 21, 515-528.

A review and analysis of the five performance measurement methods described in the current literature is conducted. The five measures described are laboratory, analytic, synthetic, simulation, and operation-system methods. Laboratory methods tend to be the most desirable for several reasons, including ease and precision of control over relevant performance demands, ease of duplication of conditions, relative ease of measurement, safety and equipment reliability and low cost, and the possibility of selecting tasks relative to the variables of known operational importance.

The author stresses the need for reliable and quantitative criteria which will reflect performance and concludes the article with a potentially useful paired-comparisons scaling procedure.

20. Christensen, J. M., and Mills, R. G., What does the Operator do in Complex Systems? Human Factors, 1967, 9, 329-340.

The authors attempt to describe, in phychological terms, the actual activities of systems operators and to develop estimates of the time spent on each activity. A review of the literature revealed little useable data on human activity under operational conditions, so a procedure to collect data was developed. The procedure included operational, test, simulation, and paper and pencil data. Paper and pencil studies referred to those conducted without benefit of any of the above facilities.

The taxonomy used classified behaviors into perceptual processes, mediational processes, communication processes, and motor processes. Each process was described in terms of activities and associated specific behaviors. The authors independently classified the activities of nine operators, and results indicate that, on the whole, satisfactory agreement can be obtained when classifying such data.

To strengthen rater reliability, several suggestions are made. It is suggested that operator activities be clearly defined and that classifiers be as familiar as possible with the job. Reliability could be increased if those who generate the activity data had the same behavioral taxonomy in mind. Again, reliability could be increased if those who classify the activities have sufficient practice.

Assuming that reliable and valid activity data could be gathered, the authors make several suggestions as to its use in an operational context.

21. Christy, R. L., Personality Factors in Selection and Flight Proficiency. Aviation, Space and Environmental Medicine, 1975, 46, 309-311.

The author first discusses personality in general, describing it as a "dynamic process involving life forces, instinctual drives, growth processes, life stresses, reactions to significant persons, both in and out of the family, and reactions to the environment in general, along with the mastery of methods and techniques for coping in solving problems and in providing for the satisfaction of one's needs." He stresses that the personality of pilots often involves similar motivation and conflicts, and deserves careful evaluation. Some adjectives used to describe expressed desires and feelings by those who love to fly include: experiencing thrills, freedom, excitement, power, speed, independence, competition, omnipotence, defiance; escaping (from earth), doing the forbidden; denying fear; or defying death.

Selection of flight personnel involves the assessment of the various factors involved in the individual's desire to fly. The author stresses the fact that these personality factors require close surveillance by flight surgeons, aviation medical and operational personnel in the interest of maintaining optimum personnel effectiveness and proficiency.

22. Companion, M. A., and Corso, G. M., Task Taxonomy: Two Ignored Issues. Proceedings of the Human Factors Society 21st Annual Meeting, 1977, 358-361.

The author begins by defining the concept taxonomy and states "a taxonomy is a means of classifying objects and phenomena in such a way that useful relationships among them are established." He cautions that a mere list of labels is not a taxonomic scheme - "There must be an inner syntatic structure around which useful relations can be established."

Two issues that have been ignored in the development of task taxonomies are then discussed.

(1) One must develop a set of criteria on which a judgment can be based for the evaluation of how well a task taxonomy accomplishes the goals responsible for its development.

(2) One should be concerned about the relations between taxonomic structure and empirical data.

An effective task taxonomy should do the following:

(1) The taxonomy must simplify the description of tasks in the system.

(2) The taxonomy should be generalizable.

(3) The taxonomy must be compatible with the terms used by others.

(4) The taxonomy must be complete and internally consistent, i.e. it should deal with all components of human performance in the system.

(5) The taxonomy must be compatible with the theory or system to which it will be applied.

(6) The taxonomy should help to predict operator performance.

(7) The taxonomy should have some utility.

(8) The taxonomy must be cost-effective.

(9) The taxonomy must provide a framework around which all relevant data can be integrated.

23. Connelly, E. A., Schuler, A. R., and Knoop, P. A., Study of Adaptive Mathematical Models for Deriving Automated Pilot Performance Measurement Techniques. Volume I. Model Development. Air Force Human Resources Laboratory, AFHRL-TR-69-7-VOL-1, 1969.

This report, first in a series, describes a new efficient approach to human performance assessment. The paper discusses model development and incorporates three models into one system, which is designed to evaluate pilot performance with the aid of a computer. The models are, briefly, state transfer measures, absolute measures, and relative measures. They are discussed in greater detail in Volume II, and an evaluation using the system is described.

24. Connelly, E. M., Schuler, A. R., and Knoop, P. A., Study of Adaptive Mathematical Models for Deriving Automated Pilot Performance Measurement Techniques. Volume II, Appendices. Air Force Human Resources Laboratory, AFHRL-TR-69-7-VOL-2, 1969.

A new, efficient approach to deriving human performance measures with the aid of a computer is described. Mathematical and computer models are used that examine performance data which correspond to known skill levels.

A system using three types of models was developed in order to derive a unique set of performance measures. The models are:

(1) State transfer measures, based on overall trends of performance.
(2) Absolute measures, based on a comparison of actual performance with some standard.
(3) Relative measures, based on the relationship between various performance variables.

An evaluation of the system was made, and the results are discussed.

25. Connelly, E. M., Schuler, A. R., Bourne, F. J., and Knoop, P. A., Application of Adaptive Mathematical Models to a T-37 Pilot Performance Measurement Problem. Air Force Human Resources Laboratory, AFHRL-TR-70-45, January 1971.

This report, one in a series, discusses research on a new method of deriving performance measures and criteria for use in automated pilot performance assessment. It uses data recorded on board a T-37B aircraft which is submitted to a previously implemented system of adaptive mathematical models (AMM). The results are analysed to determine the capability of the AMM in deriving measures and criteria. Results indicate that the AMM system can be used to effectively address the problems of performance measurement using representative flight data. Also, face-validity of measures derived from the models is illustrated by comparisons and correlations with performance assessment made by the instructor pilot.

26. Damos, D. L., Cross Adaptive Measurement of Residual Attention to Predict Pilot Performance. Aviation Research Laboratory, University of Illinois, ARL-72-25/AFOSR-72-12, October 1972.

Two studies concerned with the measurement of residual attention (the amount of attention a pilot has left over while performing routine tasks) are discussed. The purpose of the first study was to determine if residual attention could be measured using adaptive tasks; the second study resulted in estimates of the predictive ability of the test battery as well as test-retest reliabilities.

Results indicate that further study in the area of residual attention is necessary to determine whether it is a good predictor of flight performance.

27. Damos, D. L., and Lintern, G., A Comparison of Single - and Dual-Task Measures to Predict Pilot Performance. AFOSR-TR-80-0325, 1979.

Following a literature review of the use of multiple-task performance measures to assess pilot performance, the author described an experiment which assesses the predictive validity of single- versus dual-task measures. Subjects received two trials on two identical one-dimensional compensatory tracking tasks. Following measurement of tracking task performance during several trials, the subjects were given a short basic flight course with ground instruction and practice in a GAT-2 simulator. When they finished, the subjects were asked to perform four repetitions of a descent, a descent followed by a stall, and a level turn. Performance during these maneuvers was scored by an instructor and observer. Performance of the tracking task was correlated with the performance in the simulator. It was shown that the predictive validity of the early single-task scores decreased with practice while the dual-task validity increased during the course of the test period. A discussion of the results and possible explanation concludes the study.

28. Ellingstad, V. S., and Heimstra, N. W., Performance Changes during the Sustained Operation of a Complex Psychomotor Task. Ergonomics, 1970, 13, 693-705.

This study used a primary tracking task and a variety of subsidiary tasks, and three physiological measures in order to investigate the relationship between the time subjects spent on the tasks and the effects of fatigue.

The results show a significant decrement for the tracking task. Rather than an abrupt occurrence, the decrement took place over the course of the experiment. In the subsidiary task measures, there was no clear decrement in performance, although a marked variability was noted. Performance on two of the subsidiary tasks (vigilance and reaction-time tasks) actually improved, contrary to expectations. The physiological measures (17-Hydroxycorticord, 17-Ketosteroid, and eosinophil count) produced no substantial results, except in the case of the 17-Hydroxycorticord where the increase was significant.

(Note: The 17-Hydroxycorticord analysis and the 17-Ketosteroid analysis measure adrenal cortical excretion from urinary output. An eosinophil is a white blood cell that has been shown to be sensitive to factors like stress or fatigue. There tends to be a decrease in their numbers during such situations.)

29. Engel, J. D., An Approach to Standardizing Human Performance Assessment. Proceedings of the Planning Conference of Standardization of Tasks and Measures for Human Factors Research, Texas Technological University, March 1970.

Two critical areas regarding standardization and evaluation of methods of performance assessment are discussed. They are (1) a task classification system and (2) a performance measure classification system. The author discusses some preliminary research which uses a performance measure classification system. Suggestions for further research are made, including:

(1) The need for the development and refinement of an interim task classification system and
(2) The development of an interim classification system for human performance measures.

30. Eschenbrenner, A. J., Effects of Intermittent Noise on the Performance of a Complex Psychomotor Task. Human Factors, 1971, 13, 59-63.

This paper discusses a study in which the effects of aperiodic (unevenly spaced) and periodic intermittent noise on complex psychomotor task performance are assessed. The task was manual-image-motion compensation. Twenty-four male subjects, with a mean age of 28.5 years, served as subjects. Performance was measured according to the total amount of time-image motion was held at or below 40 microradians per second.

The results indicate that white noise had a detrimental effect on image-motion compensation performance. The magnitude of the decrement varied as a function of the temporal pattern and level of intensity of the noise.

31. Farrell, Dr. J. P., Measurement Criteria in the Assessment of Helicopter Pilot Performance. Proceedings of the Conference to Assess Aircrew Performance in Army Aviation, U. S. Army Aviation Center, November 1973, 141-148.

Factors which should be considered in selecting a method for measuring pilot flight performance are described. The first consideration is, of course, to establish the purpose of the study. The author cites four major areas:

(1) Determine operational readiness of an aviation unit,
(2) Assess flight proficiency of individual pilots,
(3) Diagnose strengths and weaknesses, and
(4) Research equipment, procedures, selection, and training.

Factors complicating the measurement of pilot performance are summarized and include the multidimensional aspects of the pilot's task. It includes not only psychomotor events but also complex perceptual and judgmental elements. In addition, one pilot may be highly skilled on some tasks and less effective with others, creating a problem in developing a single overall scale for flight proficiency. Outside elements, such as weather, need to be taken into consideration as they affect the aircraft and pilot performance. The level of difficulty of particular maneuvers may not be the same for each pilot, due to effects from such variables. Intra-pilot variability needs to be taken into account. Finally, the establishment of a performance criterion is difficult since the amount of variation from a prespecified requirement may not keep a flight from being "successful."

Several performance measures are summarized and recommendations for further study are made.

32. Finkelman, J. M., and Glass, D. C., Reappraisal of the Relationship between Noise and Human Performance by Means of a Subsidiary Task Measure. Journal of Applied Psychology, 1970, 54, 211-213.

It has been shown in the literature that constant and expected noise resulted in minimal performance decrement, but that random and unpredictable noise bursts were responsible for substantial performance decrement.

The findings of the present study confirm the assumption that demands imposed by the task, with concurrent environmental stress, must exceed the operator's channel capacity for degradation to occur. That is, the use of unpredictable noise resulted in performance decrement on a subsidiary task. Peformance on the primary task was unaffected by either type of noise, predictable or unpredictable.

Based on these results, the authors conclude that a reduction in spare mental capacity is a function of the aversiveness of the noise stressor.

33. Fleishman, E. A., The Prediction of Total Task Performance From Prior Practice on Task Components. Human Factors, 1965, 7, 18-27.

The research documented in this paper is concerned with the intercorrelation among task components with each other, and with the total task. More specifically, the questions that were addressed are:

(1) To what extent is performance on task components, individually practiced, predictive of subsequent total task performance?
(2) To what extent is practice on combinations of components predictive of total task performance?
(3) What are the interrelationships among component performances?
(4) What is the relative contribution of various component performance to total and subtask performance?

Two-hundred and four Air Force basic trainees served as subjects performing seven different tasks on the Multidimensional Pursuit Apparatus. First, the subjects practiced on two components simultaneously. Three "double-component" combinations were used. The total task was administered as the final task, which required subjects to keep three dials centered through the simultaneous use of three controls.

Results of correlational and multiple correlational analyses showed that it was possible to obtain significant predictions of total task performance from prior performance on individual components. Other findings include (1) the best predictors of total task and multiple control subtasks were other multiple control subtasks and (2) particular components involved were not as important as the fact that simultaneous practice had occurred.

34. Fleishman, E. A., Performance Assessment Based on an Empirically Derived Task Taxonomy. Human Factors, 1967, 9, 349-366.

Three major issues are addressed as critical to the assessment of operator performance. They are:

(1) The need to identify behavior that will be assessed,
(2) The need to develop diagnostic measures of the behavior, and
(3) The need to develop relevant experimental and statistical methodology for the purpose of prediction.

Adherence to these three issues may alleviate the major drawback of performance evaluation to date; that is, seldom is a measurement system applicable to more than the specific setting for which it was designed.

The general outline of a research program that addresses experimental-correlational studies of complex performance is discussed. The author concludes by stating that "the centrality of the taxonomy problem is critical to military psychology, to problems of assessing complex performance, and to so many questions of generalizing from system to system."

35. Fleishman, E. A., and Ornstein, G. N., An Analysis of Pilot Flying Performance in Terms of Component Abilities. In W. Ronan and E. Prien (Eds.) Perspectives on the Measurement of Human Performance. New York, Appleton-Century Crofts, 1971, 346-362.

The study focuses on a factor analysis of performance in several different maneuvers. The authors attempt to specify variance in common between maneuvers to provide insight into the dimensions of individual differences in the flying tasks. The factors were identified as control precision, spatial orientation, multilimb coordination, response orientation, rate control, and kinesthetic discrimination. Results indicate that such ability categories are useful in describing complex skills.

36. Freedle, R. O., Zavala, A., and Fleishman, E. A., Studies of Component-Total Task Relations: Order of Component-Total Task Practice and Total Task Predictability. Human Factors, 1968, 10, 283-296.

Following a review of the literature, the authors describe the current evaluation which uses the Complex Coordination Test to examine component - total task relationships when the component tasks are practiced in different orders.

Correlational analyses were performed to determine the component-total task relationships and component to total task predictability. Results indicate that the order of part-task practice affects proficiency and that part-task scores may be combined to yield prediction scores on a more complex task. Practice had an effect on double-level tasks. That is, it led to better total proficiency than did practice on single-level tasks.

Implications for applications in total task proficiency evaluation are summarized, including the ability to specify the minimum amount of training necessary.

37. Fuller, J. H., Waag, W. L., and Martin, E. L., Advanced Simulator for Pilot Training: Design of Automated Performance Measurement System. Air Force Human Resources Laboratory, AFHRL-TR-79-57, August 1980.

This report traces the development of an automated performance measurement (APM) system. The underlying assumption that flying performance has several characteristics directly reflected in certain parameters is discussed. The parameters are (1) maintaining certain aircraft state parameters, such as airspeed or altitude, close to some defined criterion value, (2) avoiding excessive rates and acceleration forces so that the maneuver is executed smoothly, (3) accomplishing these objectives with the least amount of effort, by minimizing control inputs, and (4) not exceeding procedural or safety limits established for the maneuver. Several scenarios were implemented:

- Transition Tasks, straight and level, airspeed changes, turns, climbs, descents
- Takeoff/Approach/Landing Tasks, takeoffs, tech order climbs, slow flight, configuration changes, straight-in approaches, overhead patterns, touch-and-go's
- Instrument Tasks, rate climbs/descents, vertical S-A, vertical S-D, GCA, proceed direct to VOR
- Aerobatics, barrel roll, loop, splits, cloverleaf, cuban 8, lazy 8
- Formation, fingertip

These scenarios are discussed and recommendations for further study are included.

38. Gerathewohl, S. J., Psychophysiological Effects of Aging: Developing a Functional Age Index for Pilots: I. A Survey of Pertinent Literature. FAA Office of Aviation Medicine, 1977.

A survey was conducted regarding age and aviation-related psychophysiological functions, in order to develop a functional age index for pilots. The literature reviewed was related to age differences as measured by standardizing tests of sensory, perceptual, mental, cognitive, and neurophysiological functions and processes and objective assessment of personality. This report is first in a series devoted to the development of an age index for pilots.

39. Gerathewohl, S. J., Psychophysiological Effects of Aging - Developing A Functional Age Index for Pilots, II. Taxonomy of Psychological Factors. FAA-AM-78-14, 1978.

In an effort to identify psychological variables that are associated with pilot performance and proficiency, a study was conducted to identify, analyse and measure these variables. In the context of this paper, the use of psychological variables will be to determine whether functional age can be substituted for chronological age as a criterion for terminating a pilot's career.

Three specific approaches were used to determine psychological (and psychophysiological) variables. They are (1) the analysis of successful pilot behavior as displayed under simulated and operational conditions (2) the analysis of unsuccessful pilot behavior (pilot error) as related to aircraft accidents, and (3) the evaluation of pilot performance during the selection and training procedures as reported in the literature.

Fourteen factors were statistically isolated as important variables in pilot performance. They are perception, attention, reaction, orientation, sensorimotor, stamina, cognition/ mentation, interpersonal relations, decision-making, experience, learning, personality, mechanical ability, and motivation.

Although no attempt is made to rank the factors in accordance with their importance, their relationship to age and the aging pilot is discussed.

40. Gerathewohl, S. J., Psychophysiological Effects of Aging - Developing A Functional Age Index for Pilots, III. Measurement of Pilot Performance. FAA-AM-78-14, 1978.

The main objective of the study is to describe pilot proficiency measurement techniques to evaluate the effects of aging on pilot performance. The two major areas of pilot performance measures can be divided into (1) qualitative assessment based on subjective ratings and (2) quantitative assessment based on recordings of pilot actions and aircraft response. A review of the literature leads to the following conclusions about needed improvements in the area of objective measurement:

(1) Performance analysis should be conducted in order to establish quantifiable descriptors of performance (including the definition of errors, scales, and scoring techniques)

(2) Selection of a unit of measurement in regard to human subsystem or operator performance should be made.

(3) Selection of appropriate parameters should be accomplished.

(4) Measurement system test and evaluation should be conducted.

(5) Calibration and standardization of the measurement system and validation of same should be performed.

(6) Calibration and standardization of the data and preparation of the information in a practical, manageable, and usable form should be accomplished.

41. Gerathewohl, S. J., Chiles, W. D., and Thackray, R. I., Assessment of Perceptual and Mental Performance in Civil Aviation Personnel, <u>Proceedings of the AGARD Conference on Higher Mental Functioning In Operating Environments</u>, AGARD-CP-181, October 1975, C7-1-C7-4.

The article describes a series of experiments designed to assess perceptual and mental functions important to aircrew, pilot and Air Traffic Control (ATC) performance. For the experiments concerning pilot and aircrew performance, the multiple task performance battery (MTPB) was used. Complex performance is represented by the complex demands placed on an operator in man-machine systems and requires that a pilot divide his attention among a number of different displays. The tasks used to assess complex performance were (1) monitoring of static and dynamic processes, (2) stimulus discrimination or target identification, (3) information processing in the form of mental arithmetic and visual pattern discrimination, (4) compensatory tracking task, (5) time-sharing task combining several of the above tasks, and (6) a task measuring procedural or group dependent behavior.

It was shown that performance decrements are conducive to error, failure, and accidents, and only by in-depth analysis of performance that errors can be identified and prevented. The probability of error is highly dependent upon the priority the operator assigns to a particular task, as well as upon the personality of the individual. Finally, significant correlations between mental and perceptual-motor attributes and some physiological functions were found. These correlations are of value in the assessment of the total human element in the man-machine system.

42. Gopher, D., and North, R. A., Manipulating the Conditions of Training in Time-Sharing Performance. Human Factors, 1977, 19, 583-593.

In order to assess three aspects of training under time-sharing conditions, the authors combined a one-dimensional compensatory tracking task and a digit-processing, reaction-time task.

Differential improvement on the two tasks resulted, and was attributed to the manipulation of relative task priorities under time-sharing, and repeated presentation of a single-task/dual-task sequence. The major source of improvement on the tracking task was specific to the skill of tracking, while digit-processing improvement may be a result of an improved ability to time-share.

Similarities and differences in the tasks are cited. Both tasks include visual processing and motor response under reaction-time pressure. However, the tracking exercise is an externally paired, continuous task, while the digit task is self-paced and discrete.

It is concluded that the observed differences have important implications when considering the design of training schedules.

43. Gopher, D., and North, R. A., The Measurement of Attention Capacity through Concurrent Task Performance with Individual Difficulty Levels and Shifting Priorities. Illinois University, Aviation Research Lab, ARL-74-13, 1974.

Unsolved problems in the application of secondary task techniques include the evaluation of relative changes in performance in dual-task situations, the prediction of possible interactions between different tasks and their components, and the extent of voluntary control of capability allocation. This article describes a three part experiment in which an effort was made to address these problems by a new methodological approach. The three successive phases included separate performance of the experimental tasks with adaptive adjustment of difficulty, simultaneous performance of the tasks with equal task priorities, and simultaneous performance with several manipulations of the two task priorities. The tasks used for the experiment were one-dimensional compensatory tracking task and a digit-processing, reaction-time task.

44. Grodsky, M. A., The Use of Full-Scale Mission Simulation for the Assessment of Complex Operator Performance. Human Factors, 1967, 9, 341-348.

The basic question that is considered in this paper is the choice of method which can be used to evaluate operator performance in a complex system. The author addresses the major reasons for assessing operator performance, and then discusses a number of methods available for measuring complex performance. These include:

(1) Inflight methods (most valid),
(2) Analytical studies (modeling performance of the operator by means of available data),
(3) Engineering or expert opinion (a possible substitute for quantitative means of determining operator performance because of the lack and substantiation of a generally accepted method.),
(4) Mockup techniques (yield information about operator performance only in the most general sense),
(5) Laboratory tasks and synthetic task batteries (though they demonstrate internal validity and sensitivity in stress situations, they lack applicability in the general sense to tasks in an operational situation), and
(6) Simulation studies (including part-task simulations and integrated missions simulations).

It is concluded that the integrated mission simulation, if applied according to certain criteria, can provide all the necessary data obtained by other techniques described, with the added advantages of the integrated mission itself. The rationale for this conclusion is presented, and associated problems are discussed.

45. Hammell, T. J., Pesch, A. J., and Lane, W. P., Decision-Making Performance Measurement for a Command and Control Training System. Proceedings of the Human Factors Society 19th Annual Meeting, October 1975, 315-320.

A performance measurement technique is described that is suitable for tactical decision-making training. Traditionally, performance measurement and feedback has been dependent on the culmination of the exercise. With the particular method described in this paper, measurement is made at interim points during and across segments of the exercise. The training exercise is directed by a combination of context-specific tactical objectives. The methodology required to provide this complex performance is described. Of major importance in the performance assessment system is the close relationship between elements of the technique, i.e. training objectives, trainee behavior, system effectiveness, diagnostic feedback, and the training methodology.

It is concluded that the technique is a useful tool for trainee performance measurement in complex man-machine systems.

46. Hartman, B.O., Benel, R. A., and Storm, W. F., A Review of USAFSAM Studies Employing Multiple Task Performance Devices. USAF School of Aerospace Medicine, Brooks Air Force Base, Texas, December 1979.

This paper contains summaries of studies that use apparatus classified as multiple-task performance batteries, used to measure performance and proficiency of systems operators. Task performance evaluation devices at the (USAF) United States Air Force School of Aerospace Medicine are geared toward medically-directed, applied problems from the operational environment. Problems studied were usually concerned with environmental stressors, and due to the low intensity of the stressors in many cases, significant performance decrements were seldom reported. However, the authors conclude that the test batteries developed show promising capability for performance evaluation.

47. Henry, P. H., An Automated System To Assess Pilot Performance in a Link GAT-1 Trainer. School of Aerospace Medicine, Brooks AFB, Texas, October 1974.

The authors describe a prototype control and scoring system which was developed around the Link GAT-1 trainer, permitting laboratory assessment of pilot performance. The system automatically presents pilots with an hour-long series of maneuver requests, providing an approximation of a cross-country flight. Pilot performance is scored electronically based on how closely pilots are able to stay within the tolerance prescribed. The report discusses basic design and circuitry details. Details of performance tests using this system are reported elsewhere.

48. Higgins, T. H., A Systems Engineering Evaluation Method for Piloted Aircraft and Other Man-Operated Vehicles and Machines with Hypothetical Examples of a Systems Evaluation and Quantified System Performance-Workload Rating Scales. U. S. Department of Transportation, FAA-RD-81-30, March 1981.

A systems evaluation method is presented which classifies and quantifies both PRP (pilot rating procedure) and ECP (engineering calculation procedure) measures of system performance on a logarithmic ratio of basic test aircraft configurations compared to a known selected standard aircraft vehicle configuration. The logarithmic units (10 log ECP_{test}/ECP_{std} and 10 log PRP $_{test}/PRP_{std}$) used in this system evaluation method are termed "decivals, d^V" as they are 10 times the log base 10 of the ratio of the ECP and PRP values obtained during tests for the test aircraft configuration, compared to the chosen standard aircraft configuration.

The system evaluation is for chosen time periods of selected flight operations which are critical to flight safety, such as may occur during takeoff, or approach to landing and may include emergency engine failure, flight control or instrument malfunction conditions. System equations are presented which answer the question as to how good is the test configuration in relation to the known standard configuration during these same flight conditions. Potential ECP measures are discussed and their correlation with PRP pilot ratings obtained during flight test or flight simulator test determines their retention as effective system performance and evaluation measures. The non-dimensional logarithmic nature of the retained ECP system performance descriptors allows their combination by logarithmic summation and their correlation with the PRP pilot ratings is determined. The combination of ECP measures having the highest correlation with pilot ratings is retained for final system evaluation. (ABSTRACT)

49. Hill, J. W., and Goebel, R. A., Development of Automated GAT-1 Performance Measures. Air Force Human Resources Laboratory, AFHRL-TR-71-18, May 1971.

In this study, three different experience level groups of general aviation pilots were given four flight tasks in a GAT-1 simulator in order to study the complex research problems of pilot proficiency.

For the experiments described, the GAT-1 was interfaced with a Line 8 computer system to continuously record eight flight variables (airspeed, altitude, climb, roll, pitch, heading, glidescope, localizer). Also, signals were supplied to three GAT-1 variables (roll, pitch, heading) for tracking tasks.

The four tasks used were:

- Task 1 - Altitude and Heading Holding
- Task 2 - Altitude and Heading Holding with Power Changes
- Task 3 - Flight Profile
- Task 4 - ILS Landing Approach

Processing of the data yielded 266 potential indices of pilot proficiency. An analysis of variance was conducted to identify a subset of the most important parameters and it was shown that 27 measures contributed significantly to perfect separation of the three experience levels of the subjects.

50. Hill, J. W., and Eddowes, E. E., Further Performance of Automated GAT-1 Performance Measures. Air Force Human Resources Laboratory, AFHRL-TR-73-72, May 1974.

This report describes a systematic, statistically-directed search for automated flight measurements that correlate with pilot proficiency. The approach is based on two separate experiments carried out in a GAT-1 trainer: (1) a basic experiment with 326 measurements on each of 30 subjects in three experience groups, and (2) an expanded experiment with 2436 measurements on each of 30 new subjects from the same three experience groups. Experiment (1) consisted of four different flight tasks, each about 10 minutes long, and experiment (2) consisted of these and six additional tasks. The results of these experiments show that there is little difficulty in obtaining measurements that correlate with experience. More than 5 percent of the measurements of each experiment were statistically significant (0.01) level. Tables of more than 400 important measurements are given with group means and standard deviations and further cross-tabulations to show which tasks and families of measurements are best at discriminating among pilots. Three different statistical methods were used to select a set of measurements from experiment (1), and combine them into two new canonical variables, each a linear-weighted combination of the measurements in the set, to discriminate optimally among the three groups of subjects. Applying the canonical variables to the repeated measurements of experiment (2), allowed several deductions about the best selection procedure to be made. (ABSTRACT)

51. Hockenberger, R. L., and Childs, J. M., An Integrated Approach to Pilot Performance Evaluation. Proceedings of the Human Factors Society 24th Annual Meeting, 1980, 462-465.

The article describes norm-versus criterion-referenced evaluation methods using undergraduate pilot training as an example. An approach that integrates both is described. The authors suggest that such an integrated approach could facilitate advanced training management concepts, such as proficiency advancement, while not posing significant user acceptance difficulty.

A 4-point criterion system was used in the research described in this study, but the authors recommend the use of more criterion levels that enable additional grading combinations (i. e. norm-referenced grade + criterion-referenced grade). The 4-point system is as follows:

4 = Criterion level performance (very small variations from standard)
3 = Small variations from standard
2 = Medium variations from standard
1 = Large variations from standard

52. Howard, Dr. L. R. C., Emotional Stress and Flying Efficiency. Proceedings of the AGARD Conference on Higher Mental Functioning In Operational Environments, AGARD-181, October 1975, C8-1-C8-5.

A comparative study of the effects of "emotional" and "intellectual" stress upon flying performance is described. Ten pilots with self-confessed emotional problems, but certified medically fit to fly, were matched approximately for age and flying experience with ten pilots confessing to no emotional problems. Emotional stress was induced by a clinical abreactive technique and measured in terms of psychophysical concomitants. Intellectual stress was induced by a technique which utilizes an automated numerical task designed to produce mental overload. Flying skill was measured in the presence and absence of both types of stress separately by means of a crossover design, using a specially modified D4 Link trainer. The results show significant differences in flying performance between the two groups and between the two types of stress. It is demonstrated that intellectual stress produces impaired flying skill which is predictable both in degree and duration, whereas emotional stress produces substantially more severe, but fluctuating degrees of impairment of unpredictable duration. The use of a psychometric technique, using personal keywords and physiological monitoring, clearly differentiates the two groups of pilots and indicates the origins of emotional stress. Its possible use for screening aircraft captains before important missions is examined. (ABSTRACT)

53. Jennings, E., and Chiles, W. D., An Investigation of Time-Sharing Ability as a Factor in Complex Performance. Human Factors, 1977, 19, 535-547.

The purpose of the paper was to investigate two different complex tasks using the factor analytic method in order to determine whether any of the performance measures exhibit properties that may be attributed to a time-sharing ability. Following a brief review of the relevant literature, the authors describe an experiment they conducted with 39 male subjects who were tested on six tasks. Their performance was measured on each task presented individually and on two complex tasks composed of three-task subsets. The multiple-task performance battery (MTPB) designed to test and measure skills important to aircrew performance was used. The test includes six tasks: warning lights, meter monitoring, mental arithmetic, pattern identification, group problem-solving, and a two-dimensional compensatory tracking task.

Results of the factor analysis revealed a factor that showed high loadings for two different monitoring tasks for complex performance but negligible loadings for these tasks for simple performance. In addition, separate orthogonal factors for the two monitoring tasks were found when they were done under simple task conditions; thus, the monitoring measures seem to possess properties that one would expect from measures of a time-sharing ability.

54. Kleinman, D. L., Baron, S., and Levison, W. H., An Optimal Control Model of Human Response, Part I: Theory and Validation. Automatica, 1970, 6, 357-369.

The paper discusses the development of a mathematical model of the human as a feedback controller using optimal control and estimation theory. The underlying assumption was that the human operator acts in a near-optimal manner subject to his task definition, inherent limitations, and constraints. The model contains time-delay, a representation of neuromotor dynamics and controller remnant to represent certain psycho-physical limitations.

The human's control characteristics are considered in terms of three linear operations: pure time-delay, equalization network, and neuromuscular/actuator dynamics. Observation noise and motor noise are representations of controller remnant.

Mathematical development of the model is discussed, along with model application, and validation (using three simple compensatory tracking tasks).

It is concluded from the results that the model is capable of reproducing all the essential data in the experiment, and the agreement between the theoretical and measured quantities make the use of modern control theory a valuable tool in human operator studies.

55. Baron, S., Kleinman, D. L., and Levison, W. H., An Optimal Control Model of Human Response, Part II: Prediction of Human Performance in a Complex Task. Automatica, 1970, 6, 371-383.

The model described in Part I of this study is used to analyse a manual control task involving the control of longitudinal position of a hovering VTOL aircraft. The problem is a more complex one, and it was found that the model was capable of reproducing most of the essential control characteristics of pilots. Also, system performance scores were, for the most part, faithfully reproduced.

The techniques used in applying the optimal control model of the human operator with reference to this task are discussed, and theoretical and measured data, are compared. The paper concludes with a simplified technique for predicting "average" behavior.

56. Knoop, P. A., Advanced Instructional Provisions and Automated Performance Measurement. Human Factors, 1973, 15, 583-597.

This paper describes research conducted by the Air Force Human Resources Laboratory on a basic simulation research program.

The report is divided into the following areas:

(1) Advanced instructional provisions, including automated demonstrations, variation of task difficulty, automatic malfunction insertion, instructor feedback, data recording, student feedback, automated task sequencing, and implementation;

(2) Performance measurement, including data acquisition and calibration, data collection phases, measurement approaches, and results. (In this section, the author uses her T-37 pilot performance feasibility study as a source of data.);

(3) Overview of current measurement technology for the described application, including the two complementary approaches described in this paper; and

(4) Simulator Implementation, including advanced instructional provisions.

57. Knoop, P. A., and Welde, W. L., Automated Pilot Performance Assessment in the T-37: A Feasibility Study. Air Force Human Resources Laboratory, Advanced Systems Division, AFHRL-TR-72-6, April 1973.

The purpose of the research documented in this very extensive report is to develop improved methods of pilot proficiency measurement. In the past, pilot performance was assessed by an instructor pilot applying a subjective rating scale. This method suffers from several drawbacks, so it was the intent of the authors to develop a more valid, reliable and sensitive measure of proficiency.

Instrumentation was developed for recording T-37 flight data, and software designed to automatically measure pilot performance using the recorded data. The approach used was to compute measures that were initially selected on the basis of Air Training Command maneuver analyses and possessed content validity.

During the study, data were collected on the lazy 8, barrel roll, and seven other maneuvers. The variables formed a sufficient data base for measurement, and techniques were developed for determining required sampling rates.

Lazy 8 performance assessment was accomplished using the parameters of roll angle, pitch angle, and airspeed. Barrel roll measurement was dependent upon roll and pitch angle, acceleration, and roll rate. A relationship between roll and pitch appeared to be critical to measurement.

It is concluded that automated proficiency assessment, if combined properly with subjective evaluation, can provide more valid and reliable measurement techniques than previously used.

58. Knowles, W. B., Aerospace Simulation and Human Performance Research. Human Factors, 1967, 9, 149-159.

Simulation facilities offer much potential for the generation of data on human performance. The author describes methodological limitations imposed by the basic characteristics of human performance, including versatility and variability. Efficient organizational techniques and development of more efficient and reliable techniques of experimental design and execution are necessary to fully assess the relationship between mission-oriented simulation studies and human performance research.

59. Koch, H., The Prediction of Human Performance: Theoretical and Practical Problems. Proceedings of the Symposium and Workshop on the Quantification of Human Performance, University of New Mexico, August 1967, 63-76.

The purpose of this paper is to discuss two significant aspects of performance prediction:

(1) Most performance analysis and prediction is concerned with a limited set of specific stimulus response relationships.

(2) The most meaningful predictions of human performance take into consideration the total environment and all performer responses.

The author discusses models of human performance, including the machine model, motivational model, rational model, and organization model. In addition, he outlines some basic aspects of the total stimulus situation, forming a basis for the development of a taxonomy of the total situation.

He concludes that no single approach is adequate, but if techniques and verified results of various approaches are unified, more significant predictions may be obtained.

60. Kohl, G., and Flamm, L. E., Effects of Observation on Behavior in Applied Research. Proceedings of the Human Factors Society 24th Annual Meeting, 1980, 466-467.

Many times, it is necessary for human factors researchers to observe the man-machine interaction for evaluation purposes. In order to determine what effect an observer has on facilitation or inhibition of operator performance, (outside of experimenter bias considerations) the authors conducted an experiment to determine if the presence of an audience distorts behavior.

It was shown that the presence of an observer did inhibit performance, but in a complex fashion, and was dependent on the experience and practice of the subject. The implications of these findings are discussed in terms of their relevance to human factors experiments in which an observer is used.

61. Krendel, E. S., and Bloom, J. U., The Natural Pilot: A Criterion System for Flight Proficiency Evaluation, Phase 1 Report, Franklin Institute Laboratory, September 1959.

The authors present a new approach to the evaluation of the flight proficiency of Naval Aviators which augments the subjective evaluation of piloting performance with a number of objective measures. They suggest that objective servomechanism descriptions of the "unstandardized" pilot are appropriate in pilot proficiency evaluations. The "Natural Pilot Model" is presented and specifies those uniquely human aspects of pilot behavior.

The "Natural Pilot Model":

(1) Consistency of System Performance - In order to make meaningful predictions and estimates of behavior, it is necessary that the pilot's behavior be in a state of statistical control.

(2) Human Adaptability - "Natural" pilot behavior is seen in the ability of an individual to generate adaptive behavior or as a function of changing dynamic requirements. It is also an effective discriminator between man and machine.

(3) Least Effort in Skilled Performance - Stable performance is best achieved if the pilot is operating according to some principle of effort conservation.

Finally, the authors present an experimental program which specifies and validates the performance criteria that are developed in the report.

62. Lees, M. A., Kimball, K. A., Stone, L. W., and Hoffman, M. A., Aviator Performance During Day and Night Terrain Flight. Proceedings of Human Factors Society 19th Annual Meeting, October 1975, 436-440.

The paper presents the results of an investigation which compared terrain flight for low level (LL) and nap-of-the-earth (NOE) profiles for three different conditions. They were (1) day flight with the unaided eye, (2) night flight with the unaided eye, and (3) night flight using night vision goggles. Data were acquired through use of the Helicopter In-flight Monitoring System (HIMS). The HIMS, which is integrated into the helicopter control systems, measures aircraft positions in six degrees of freedom while simultaneously recording cyclic, collective, and pedal inputs, and aircraft status values. These data were recorded in real-time using an incremented digital recorder. Selected measures from the HIMS are classified as (1) pilot control measures and (2) aircraft status measures, including collective control, instantaneous control reversals, cyclic left-right control, control position mean, cyclic fore-aft control, instantaneous control reversals, pedal control, control position, standard deviation and collective control position mean, mean roll rate, mean pitch angle, mean airspeed, standard deviation, heading and standard deviation, and altitude.

The total set of inflight measures for LL and NOE were analysed separately, and further analysis was performed on the subsets of pilot control variables, and aircraft status variables. Multiple discriminant analysis techniques were used to determine which measures best discriminate between visual conditions.

63. Lees, M. A., Stone, L. W., Jones, H. D., Kimball, K. A., and Anderson, D. B., The Measurement of Man-Helicopter Performance as a Function of Extended Flight Requirements and Aviator Fatigue. U. S. Army Aeromedical Research Laboratory, Fort Rucker, Alabama, July 1979.

A major review of the literature summarizing pilot fatigue and its effect on performance was completed. In this report, the authors define the concept of fatigue as the result of extended flight requirements, and in this sense, feel that it belongs to the more general area of research, that of "operator workload".

This study, as one of a series, was undertaken to address the impact of fatigue on aviator effectiveness, especially in situations where a pilot may be called upon to fly several successive missions, often with stress related factors. Six pilots were observed in an actual flight situation, and data collection included psychological, inflight, biochemical, and visual measurements. In addition to the literature review, the test methodology is described.

64. Leonard, J. J., and Wierwille, W. W., Human Performance Validation of Simulators: Theory and Experimental Verification. Proceedings of the Human Factors Society 19th Annual Meeting, October 1975, 446-456.

A simulator can be validated for some performance measures provided the following criteria are met:

(1) Simulator must have good fidelity corresponding to the variables measured.
(2) Parameters should be capable of adjustment.
(3) A sufficient number of independent variables with appropriate settings must be employed.
(4) Performance data should be available for the standard full scale vehicle whether it be an aircraft, etc., and for each adjustment of the simulator.
(5) Methods of experimental design must be acceptable to insure unbiased data and valid conclusions.

The above conclusions were reached by the authors in their research which used an instrumented full-scale test vehicle and driving simulator. Eight adjustments, with five driver performance measures for each, were used. Simulator conditions which most closely matched the full-scale system were identified. Additionally, it was determined that the concept of performance validation is both Alpha-level and sample-size dependent. Careful consideration should be given to the size of the experiment.

65. Loental, D. G., Feasibility of Implementing Specific Performance Measurement Techniques. Quest Research Corporation, McLean, Virginia, AMRL-TR-74-95, March 1976.

The report presents summaries of two techniques for performance measurement in a manned weapon system. The first technique uses a theoretical model of the human operator's flight control policies on the simulator. The second technique is empirical and derives performance measures from the simulator data. The system studied was the F - 106 coplanar attack simulator located at the Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

66. McDowell, E. D., The Development and Evaluation of Objective Frequency Domain-Based Pilot Performance Measures in ASUPT. Oregon State University Corvallis Engineering Experiment Station, AFOSR-TR-78-1239, April 1978.

This report describes the development of frequency domain pilot control, movement-based measures. The research was intended to address the need for the development of an automated objective pilot performance evaluation system. In particular, this would be used in the Advanced Simulator for Undergraduate Pilot Training (ASUPT).

Results indicate that the frequency domain-based measures are effective in discriminating between pilot experience levels. The findings may, in addition, have application in the development of an automated objective pilot performance measurement system.

67. McRuer, D., Development of Pilot-in-the-Loop Analysis. Journal of Aircraft, 1973, 10, 515-524.

The author describes human pilot behavior and the many factors that affect a pilot's behavior, such as physical, psychological, physiological, and experimental variables. In addition, fundamental concepts in pilot-vehicle analysis, and development chronology of pilot-in-the-loop analysis, including both a single-loop, full attention tracking model, and multiloop pilot model for tracking situations are discussed.

The author maintains that "the most important class of situations in closed-loop control of aircraft are compensatory tasks in which a pilot acts on displayed error quantities between desired command inputs and comparable vehicle output motions to produce control actions."

68. Mengelkoch, R. F., Muckler, F. A., and Monroe, R. D., Pilot Performance Measurement Equipment for an Electronic Flight Simulator. Martin Company, Baltimore, Maryland, Engineering-10977, December 1959.

This paper describes equipment designed to measure pilot performance in the YF-102 flight simulator. A manual and automatic technique were combined to allow for the measurement of five flight parameters collected during segments of the test missions. The author illustrates an example of equipment used to measure pilot performance during an experiment investigating pilot simulator performance with circular and vertical primary flight instrumentation.

69. Meyer, R. P., Laveson, J. I., and Weissman, N. S., Behavioral Taxonomy Of Undergraduate Pilot Training Tasks and Skills. Proceedings of the Human Factors Society 19th Annual Meeting, October 1975, 266-270.

The purpose of this paper is to define and analyze undegraduate pilot training (UPT) flying tasks, and organize them into a classifiable structure of behavioral objectives. Flying tasks were organized through the application of a surface task analysis. The analysis was defined as "an investigative process that lists behavioral flying tasks in sufficient detail, in accordance with established ground rules, to be utilized as a tool for classifying those tasks into specific flying training categories."

A model of the flying process was developed to structure the surface analysis. The human operator was represented by three factors, including environmental and system stimuli, which affects sensory systems, mental actions, and motor actions.

The rules for accomplishing the surface analysis of the flying tasks are delineated and the resultant taxonomy described. Six basic cross-referenced data areas make up the taxonomy, including classification hierarchy, classification matrix, sorting slot content list, a task list, card file (used for cross-referencing all skill information in the taxonomy data system), and surface task analysis, which provided the information on which the taxonomy was developed.

Applications and recommendations for the taxonomy are discussed.

70. Mudd, S., Assessment of the Fidelity of Dynamic Flight Simulators. Human Factors, 1968, 10, 351-358.

The author discusses the evaluation of dynamic flight simulators in terms of efficiency and validity of pilot evaluations and assessment techniques. A general background of the fidelity assessment problem is described, and a set of requirements for an ideal fidelity measurement technique are presented. Two general approaches to fidelity measurement are also described, including:

(1) Empirical (subjective experience)
(2) Analytic (centering around a model-matching procedure where the dynamics of a given vehicle are approximated by the simulators)

The drawbacks of each measurement approach are discussed, and the author introduces a combined method which involves the use of pilot psychomotor responses rather than verbal responses. The subjective nature of rating scales remains, but additional information of an analytic nature results.

71. North, R. A., and Griffin, G. R., Aviator Selection 1919-1977. Naval Aerospace Medical Research Laboratory, Pensacola, Florida, October 1977.

This literature review provides an updated summary of psychological aviator selection research for the period 1919-1977. Briefly, the authors concluded that success in predicting aviator performance is high, and emphasize the need to assess additional factors which play a role in performance. To date, such factors have not been identified or quantified. They also recommend that highly relevant criterion-oriented performance measures be identified for use in evaluation of present selection/prediction variables.

The first section presents an overview of aviator selection, World Wars I and II. Postwar aviator selection research follows, summarizing psychomotor testing, perceptual/cognitive paper and pencil test selection research, personality test selection research, and neurological research. New directions in postwar aviator selection research follows, and current emphasis in aviator selection research concludes the study.

The paper presents an excellent overview of the history of aviator performance assessment and its current status.

72. Novello, J. R., and Youssef, Z. I., Psycho-social Studies in General Aviation: I. Personality Profile of Male Pilots, Aviation, Space and Environmental Medicine, 1974, 45, 185-188. The article describes a study in which 170 male General Aviation pilots were tested with the Edwards Personal Preference Schedule (EPPS), a modified early memory test, and the General Aviation psycho-social inventory. The results show that the EPPS fell between the norms for adult males and naval jet pilots; the two pilot profiles were more similar to each other than to adult males in general. Finally, it was shown from the measures used that male pilots tend to be "active-masculine, adventurous" types.

73. Novello, J. R., and Youssef, Z. I., Psycho-social Studies in General Aviation: II. Personality Profile of Female Pilots. Aviation, Space and Environmental Medicine, 1974, 45, 630-633.

Eighty-seven female General Aviation pilots were tested with similar measures used in Part I of this study. The results showed that the EPPS profile of these pilots was more similar to adult male norms than female adult norms. The EPPS profile was most similar to the profile for male General Aviation pilots.

74. Obermayer, R. W., and Vreuls, Dr. D., Combat-ready Crew Performance Measurement System: Phase I Measurement Requirements. Manned Systems Sciences, Inc., Northridge, CA., December 1974.

The study was concerned with a systematic definition of performance, development of methods for performance measurement, and provisions of usable measurement tools for attacking problems of combat-crew training. The report details information on academic, simulator, and flight training gathered from six different combat-crew training sites. Data were collected on training sequence, measurement possibilities, and specific new measurement developments. An extensive appendix includes a summary and sample measurements for 12 flying maneuvers.

75. Obermeyer, R. W., and Vreuls, Dr. D., Combat-ready Crew Performance Measurement System: Phase II Measurement System Requirements. Manned Systems Sciences, Inc., Northridge, CA., December 1974.

This study was undertaken to develop a systematic definition of performance criteria and measurement methods. This report, one in a series of seven, addresses the need for a standardized measurement system that utilizes areas of measurement delineated in the preceding reports. Several areas are discussed, including system criteria to guide design tradeoffs in performance measurement and preliminary systems analysis necessary to establish measurement systems requirements.

76. Obermayer, R. W., Vreuls, Dr. D., and Conway, E. J., Combat-ready Crew Performance Measurement System: Phase IIIA Crew Performance Measurement. Manned Systems Sciences, Inc., Northridge, CA., December 1974.

The paper describes a study in which a systematic definition of performance and development of methods of measurement are sought. Data on crews of fighter and heavy multiengine aircraft, specific measurement systems, and data processing problems are discussed. It is concluded that adequate training methods can be produced through research to resolve crew training and measurement problems, and to discover better methods of crew interaction.

77. Obermayer, R. W., Vreuls, Dr. D., Muckler, F. A., and Conway, E. J., Combat-ready Crew Performance Measurement System: Final Report. Manned Systems Sciences, Inc., Northridge, CA, December 1974.

This is the final report in a series directed to systematically define performance measures appropriate to US Air Force combat training needs and define a cost-effective measurement system. System criteria were based on an analysis of combat-crew training research procedures evaluated during data collection visits to nine Air Force training sites. A framework of maneuvers was compiled and formalized into measurement requirements. Measurement specifications were developed for both hardware and software requirements. In addition, an analysis of communications measurement was made from an examination of crew interaction. The performance measurement system which resulted from the research contains data acquisition, data processing, personnel, and facilities subsystems.

78. Ogden, G. D., Levine, J. M., and Eisner, E. J., Measurement of Workload by Secondary Tasks. Human Factors, 1979, 21, 529-548.

The purpose of the article was to review the post-1965 literature on the use of secondary tasks in the assessment of operator workload. Following a description of the underlying theory, the literature is summarized.

The secondary task technique is used to determine "how much additional work the operator can undertake while still performing the primary task to meet system criteria." The secondary task technique is characterized by an experimental situation in which two discrete and separate tasks are performed concurrently with a clear emphasis on the performance of one of the tasks. Use of the secondary task technique involves comparing the levels of performance obtained when the secondary or loading task is performed, along with the levels obtained on the same task when it is performed with the primary task. The difference between the performances obtained under the two conditions is then taken as a measure or index of the workload imposed by the primary task. The assumption is that the presence of the secondary task does not interfere with the performance of the primary task. The secondary task technique is also used as a loading task to assess the effects of stress on operator performance. Operators develop strategies of responding in order to contend with task demands. Typical strategies include response selection, response ommission, response queueing, criteria modification, and load balancing. Secondary task selection is very important since the type of secondary task employed has a major effect on the response strategies available.

A classification of tasks used in studies measuring workload by secondary task techniques was included in the discussion.

79. Parker, J. F., Jr., The Identification of Performance Dimensions Through Factor Analysis. <u>Human Factors</u>, 1967, 9, 367-373.

The approach outlined in the article is primarily concerned with perceptual-motor activities, or those involved with the dynamic control of a vehicle and in the manipulation of switches and controls related to vehicle operations.

The findings reveal that basic ability dimensions identified through factor analysis techniques for problems of performance are successful in accounting for the variance in laboratory criterion tasks.

In addition, tests based on dimensions identified through factor analysis are reliable and valid in content and construct.

Finally, an integrated measurement console was developed which allowed independent measures to be obtained on 18 primary perceptual-motor activities (representative of the perceptual-motor performance required to operate the Gemini spacecraft).

Implications of the findings for application to performance measurement theory are discussed.

80. Pickrel, E. W., and McDonald, T. A., Quantification of Human Performance in Large, Complex Systems. <u>Human Factors</u>, 1964, 6, 647-662.

The paper describes a methodology to systematically identify and eliminate sources of critical human-induced failures. The methodology is based on experiments in diverse environments, including manned aircraft, unmanned launch vehicles, and manned spacecraft.

Task criticality ratings are determined based on (1) the frequency with which the error is expected to occur, (2) the frequency that failure occurs as a result of the error, and (3) probable consequences of the failure. A sample worksheet used to evaluate task probability/criticality rating is included. On the basis of criticality ratings, design alternatives may be selected, for example, that are least likely to encourage human error. Other alternatives based on criticality ratings include reallocation of functions to men or machines, design changes to subsystems, procedural changes, use of warning or alert devices, or environmental changes.

In concluding, the authors maintain that the major strength of the method described is that it permits determination of the <u>probable</u> incidence of human errors, and their consequences early in system development.

81. Prophet, W. W., Performance Measurement in Helicopter Training and Operations. Human Resources Research Organization, Alexandria, Virginia, April 1972.

The study describes an ongoing research program (15 years) that involves research techniques for measuring the flight performance of helicopter trainees and pilots. This program addresses both the elemental aspects of flying and the goal-oriented aspects. Several approaches were used, with the major emphasis on nonautomated techniques feasible for operational use. Applications for the findings are discussed. In addition, automated human performance monitoring in flight simulators and implications for automated training is included.

82. Rabideau, G. F., Field Measurement of Human Performance in Man-Machine Systems. <u>Human Factors</u>, 1964, 6,663-671.

The author reviews problems associated with simulation data collection and interpretation. In addition, he discusses some practical questions related to the generalizability of field measurements.

Various types of evaluations are discussed, including explanatory, resolution, and verification evaluation, and each are defined in terms of their applicability to field measurement. The determination, definition, and control of independent and dependent variables, and modes of field data collection are covered. All of these are contingent on the objectives of the field study . A matrix is used as an analytical tool to assist in the determination of required measuring devices. It is concluded that field tests and studies are important, both scientifically and technologically.

83. Ray, J. T., Martin, O. E., and Alluisi, E. A., Human Performance as a Function of the Work-Rest Cycle: A Review of Selected Studies, National Academy of Sciences-National Research Council, Publication 882, 1961.

The authors conducted a comprehensive review of the literature regarding the effects of different work-rest cycles on man's performance (1894 to 1960). While further research in this area is indicated, the authors were able to reach some preliminary conclusions, namely:

(1) Performance does not appear to vary significantly as a function of the work-rest cycle, provided the work-rest cycle and sleep-wakefulness ratios are held constant (and the period of observation does not exceed one week).

(2) Man is apparently capable of maintaining high level performance on various tasks while maintaining a rigorous work-rest schedule for short intervals of time.

(3) Careful consideration must be given to the work-tasks and to the tests employed in measuring the changes in performance, i. e. passive tasks involving monitoring or vigilance have been found to be more sensitive to decrements than tasks which engage an individual more actively.

(4) Appropriate tests should be administered under conditions as similar to the operational context as possible.

84. Reising, J. M., and Krishnaiah, P. R., The Finite Intersection Test: A New Multivariate Statistical Technique Applicable to the Evaluation of Complex Systems. Proceedings of the Human Factors Society 24th Annual Meeting, 1980, 453-457.

For most contemporary systems, there is no one commonly accepted measure of performance which can adequately reflect their complexity. Performance is most accurately measured through a number of variables which the experimenter must analyze. Since behavior is multidimensional, it is imperative that multivariate statistics be used to determine significance of results during system evaluation. The author describes a new, multivariate, simultaneous comparison test, called the Finite Intersection Test (FIT), and illustrates its use with an example. It is concluded that with the use of the FIT, the researcher can determine which dependent variables are contributing most and where the differences lie among the levels of independent variables, resulting in a more precise analysis of the research findings.

85. Reising, J. M., Ward, S. L., and Ralet, E. P., Some Thoughts on Improving Experiments. Human Factors, 1977, 19, 221-226.

Human factors experiments that study such multidimensional concepts as operator workload and performance should take into consideration dependent variables, summary statistics, and statistical techniques. Deficiencies in experimental design currently in use include:

(1) A certain amount of vagueness regarding criteria, dependent variables, and summary statistics exists. Terms such as workload and performance are not operationally defined.

(2) Rationale for selecting summary statistics and dependent variables is vague or completely absent.

(3) Only amplitude measures are used as summary statistics, implying that other experimental effects do not occur or are not mentioned.

(4) Finally, dependent variables are rarely analyzed simultaneously (as is done with multivariate statistics).

The findings were based on the authors' review of over 200 articles in Human Factors and technical reports.

It is suggested that the rationale for a choice of criteria and dependent variables should be documented as fully as possible. Summary statistics, other than amplitude measures, should be applied where appropriate, and multivariate techniques should be used with more than one dependent measure, especially for the examination of complex human performance.

86. Rogers, D. B., Holden, Dr. F. M., Replogle, Dr. C. R., Potor, Dr. G., Day, C. N., Van Patten, R. E., Smiles K. A., and Mohr, G. C., Performance Measurement Using Pilot Controlled G_z Maneuvering With A Simulated Operational Task. Proceedings of the AGARD Conference on Performance and Biodynamic Stress-Influence on Interacting Stresses on Performance, AGARD-CP-101, 1972, C11-1-C11-5.

A technique for human performance measurement using a closed loop centrifuge has been validated. The study was performed on the Dynamic Environment Simulator (DES) operating in a closed loop mode. The simulation utilized the pitch and roll dynamics of a high performance aircraft. The measurement criteria was hits on target using a display-generated, heads-up gunsight on a maneuvering target aircraft. An important consideration was the relationship between man as a passive rider versus man as an active participant in the generation of the G_z stress. Two important demonstrations resulting from this study are: (1)

there is a significant difference in the ability of subject pilots to perform in closed versus open loop configurations and (2) it is feasible to provide a mission-related human performance metric in a selective simulation in which the $+G_z$ forces are dynamically realistic. A predictive heads-up gunsight display is utilized with target trajectories representative of aerial combat maneuvers; and on-line performance measures and immediate performance feedback are provided. (ABSTRACT)

87. Salvendy, G., and Stewart, G. K., The Prediction of Operator Performance on the Basis of Performance Tests and Biological Measures. Proceedings of the Human Factors Society 19th Annual Meeting, October 14-16, 1975, 457-466.

The authors tested the hypothesis that production performance can be better predicted from a combination of measures, namely biological and manipulative, rather than just manipulative test performance.

Several physiological measurement techniques were used, including electromyogram (EMG), heart rate, and sinusarrythmia. The one-hole test was used for test performance and production performance as a criteria measure.

Results indicate that sinusarrythmia at rest was the most powerful biological predictor of performance.

88. Seibel, R., Levels of Practice, Learning Curves, and System Values for Human Performance on Complex Tasks. Human Factors, 1964, 6, 293-298.

Many times, the values of expected operator performance in a system are a function of the level of practice of the operator. Unfortunately, there is usually not the time, money or subjects available for extensive study on an experimental task to obtain values of expected performance. The author recommends that asymptotes be estimated for each individual with the "cautious application of curve-fitting techniques, and these estimates averaged for the appropriate sample of subjects."

89. Shannon, R. H., Task Analytic Approach to Human Performance Battery Development. Proceedings of the Human Factors Society 24th Annual Meeting, 1980, 354-358.

The author uses task analytic methodology to determine what would be critical elements of Naval student flight performance. A total sample size of 78 students was used and student maneuver behaviors during primary training were compared to overall phase grades of primary, basic, and advanced using factor and regression analysis. Results indicate that flying ability is best measured by basic transitions, coordination flying and entry to dirty configurations. The author concludes that the task analytic procedure is useful in the development of performance measurement systems. The taxonomy used depicts pilot performance as a function of both continuous and discrete communications scanning, and controlling operations occuring in four-dimensional inflight environments of pitch, roll, yaw, and thrust. Further, these tasks are performed as part of an information processing loop between pilot, aircraft, and the environment. The components of display, sensory, cognition, motor, and control are contained in this loop.

90. Shannon, R. H., The Validity of Task Analytic Information to Human Performance Research in Unusual Environments. Proceedings of the Human Factors Society 24th Annual Meeting, 1980, 325-329.

The author described three task analytic methods used to specify human capabilities necessary for criterion performance. The three methodologies are (1) Maneuver Task Analysis (2) Critical Incident Technique of Grading Sheets, and (3) Flight Rating Forms. It was found that valid maneuver task descriptions could be outlined, and a clearer and more meaningful picture of student pilot behavior emerged. Inflight maneuver ratings were developed which did elicit valid and reliable data. It is concluded that the information gathered from this study could have application in Naval student selection, training and assessment.

91. Shipley, B. D., An Automated Measurement Technique for Evaluating Pilot Skill. Arizona State University Tempe College of Eduation, AFOSR-TR-76-1253, February 1976.

The purpose of this study is to establish specific indicators of pilot performance skill. Three different investigations were carried out, using an algorithmic, performance state evaluation model for instrument flight maneuvers with performance times and deviations from a standard flight path as performance indicators.

The results were as follows:

Study 1. Performance times can be used to enable an observer to discriminate between performances by two experienced pilots.

Study 2. Means of total performance time were found to discriminate between differences in treatments used in a training experiment. A priori predictions of differences in effects of these treatments on variability of group performance at specified locations were found to be significant.

Study 3. A small set of specific indicators could be used to replace a summary indicator of variability in performance.

92. Smillie, R. J., and Blake, T., The Case for Repeated Measures in Ergonomics Experiments: Another Look. Proceedings of the Human Factors Society 21st Annual Meeting, 1977, 339-343.

In order to evaluate order effects, a study was carried out which combined in one experiment, a between-subjects and within-subjects design. The experimental design included order as an independent variable and found a significant main effect for treatments. Also, treatments were found to interact with order.

The author suggests that research be conducted to determine if order effects are always present, with what particular variables and combinations they occur, consistency of effects in magnitude and direction, and the statistical and practical implications of order effects. Rather than eliminating the use of repeated measures design in ergonomic research, it is suggested that careful deliberation be given to the particular situation where repeated measures designs can be used effectively.

93. Stave, A. M., The Effects of Cockpit Environment on Long-Term Pilot Performance. Human Factors, 1977, 19, 503-514.

A fixed based helicopter simulator was used to examine pilot performance under conditions of vibration, noise and fatigue. Flight periods ranged between 3 and 8 hours, using test subjects with a minimum of 1,000 hours flying time. Results indicated that performance seemed to depend on motivation rather than fatigue; despite the onset of fatigue, performance not only did not degrade, it actually improved with time. The author explains this phenomena in the following way: as pilots feel the onset of fatigue, they put forth increased effort to compensate, which results in an initial improvement in performance.

Subjects in the study suffered from "lapses" resulting in abnormally poor performance. The lapses are short and occur at unpredictable times, perhaps the result of the body's reaction to fatigue products in the nervous system. They are also attributed to a reduction in the amount of oxygen in the air breathed by subjects, which increased the number of blocks and prolonged their duration. A momentary lack of oxygen in the brain will cause a block.

If it can be shown that such "lapses" occur in actual flight, they may provide a probable explanation for many "pilot error" accidents.

94. Strassor, H., Physiological Measures of Workload Correlations Between Physiological Parameters and Operational Performance. Proceedings of the AGARD Conference on Methods to Assess Workload, AGARD-CPP-216, April 1977, A8-1-A8-7.

Stress is defined as the input load of man at work and may be assessed by means of operational measures and time studies. Strain, defined as the individual output, cannot be quantified without considering physiological data . Variability of some physiological parameters, used mostly in the field of workload studies, are described and their practicality discussed for application to field studies. Methodological problems and improvements to counteract these problems are also discussed.

In experimental research on stress and strain, correlations between operational, physiological, and subjective rating parameters of workload are expected. By means of data from laboratory studies with simultaneously registered physiological and operational performance measures, it is demonstrated when correlations can be found and when they cannot.

The influence of different hypoxic gas mixtures on pursuit tracking and on physiological parameters is studied. From the results, several conclusions can be made, namely: already in a state of relatively mild hypoxia, physiological changes were present, but normally were concealed by reactions due to prolonged rest time. In spite of statistically significant physiological effects, no noticeable deteriorations of performance in tracking could be measured in hypoxia down to a hypoxic gas mixture of only 13 percent O_2 in inspired air. Not until before 11 percent O_2 could significant and mentionable impairments of tracking performance be found.

The same is true of noise. Decreased performance in noise could not be found, but an increased level of heart rate indicated the stress. Results indicate that physiological indices definitely react even in low workload situations in order to bring into action reserves of energy which will guarantee a normal performance. Only in high workload situations can correlations between performance and physiological measures be expected.

95. Swain, A. D., Some Problems in the Measurement of Human Performance in Man-Machine Systems. Human Factors, 1964, 6, 687-700.

This paper deals mainly with the quantification of human performance as it relates to man-machine system reliability. It is especially focused on the systems in pre-production stages.

The author reviews several studies on man-machine reliability. Three specific obstacles to human performance quantification are cited. They are:

(1) Complexity and subjectivity of available quantification methods,
(2) Grossness of assumptions behind these methods, and
(3) Disapproving attitudes of some psychologists toward these points.

These obstacles are discussed and recommendations for improvements are made. The author suggests that a human performance data bank, containing quantitative data in a wide variety of contexts be established. He cites the need for better models of human behavior which are appropriate for predicting error rates and their effects in man-machine systems.

It is concluded that available quantitative methods referenced in the paper are sufficient to make estimates of error rates for many tasks, but that current methods are still in the "infant stage". A large amount of research in this area needs to be done.

96. Thackray, R. I., Jones, K. N., and Touchstone, R. M., Self-Estimates of Distractibility as Related to Performance Decrement on a Task Requiring Sustained Attention. Ergonomics, 1973, 16, 141-152.

The paper describes a study in which 50 subjects performed a monotonous, but perceptually demanding task over a period of 30 minutes. In order to determine the level of distractibility of subjects, a questionnaire was administered to subjects prior to the experiment. High-distractibility subjects showed increasing lapses of attention during performance. Low-distractibility subjects did not show a decline in attention.

Physiological measures used with the task included respiration - period variability, heart-rate variability, and skin conductance. While significant changes were recorded, they did not differ significantly between the two groups.

The authors conclude that certain temperamental and cognitive variables are significantly related to performance.

97. Vreuls, Dr. D., and Obermayer, R. W., Selection and Development of Automated Performance Measurement. Proceedings of the Conference on Aircrew Performance in Army Aviation, November 1973, 168-174.

The problem of defining a useful set of measures to aid in human performance measurement has posed a major challenge in human factors research. Aircrew performance measurement, by its continuously varying nature, requires automated data collection techniques as well as a measurement definition that is both analytic and empirical.

This paper discusses an automated performance measurement based on simulation research. The research approach described by the authors implements the use of candidate measurement in a trial situation, collection of empirical data with human subjects in environments approximating the operational setting, and the use of mathematical selection techniques. It is concluded that the major measurement constraint is related to the amount of time and effort required to define and test performance. The recommendations of the report include a suggestion to researchers in this area to collect empirical results that are broadly applicable, so that future data collection efforts may be reduced. Also, further research is needed to identify (1) better computer algorithms for definition of measures, (2) implementation methods so that measurement can be computed, and (3) selection techniques which include diagnostic as well as discriminating and predicting measurement properties.

98. Waag, W. L., Eddowes, E. E., Fuller, J. H., and Fuller, R. R., ASUPT (Advanced Simulation in Undergraduate Pilot Training) Automated Objective Performance Measurement System. Air Force Human Resources Lab, Brooks AFB, Texas, AFHRL-TR-75-3, March 1975.

This report describes the approach taken for the development of performance measures, and presents the results of data collected from two preliminary evaluation studies. The findings indicate that objectively derived measures correlate highly with instructor ratings, and discriminate between pilots of different experience levels. These findings suggest the potential of the present approach for generating the needed automated objective pilot performance measurement system.

99. Ware, C. T., Individual and Situational Variables Affecting Human Performance. Human Factors, 1964, 6, 673-674.

The author makes a distinction between two types of variables that affect human performance. They are (1) individual and (2) situational. In the class of individual variables, such parameters as age, sex, skill level, and personality are listed. Situational variables refer mainly to task characteristics, system organization, test characteristics, and physical environment.

The two sets of variables are described as intervening variables that do not directly control human performance. Therefore, no linear relationship between one variable and the performance of a particular task can be found.

The intervening variables form a complex of factors, and to account for the effect on performance, it is suggested that a multi-dimensional independent variable be correlated with a unidimensional dependent variable.

100. Zollars, G. F., Stress Factors on Pilot Performance, May 1974 - 1980. New Mexico University, Technology Application Center, and National Technical Information Service, Springfield, VA, July 1980.

This review of articles from the international literature covers all aspects of stress factors on pilot performance, including biological aspects of flight stress, psychophysiology, acceleration stresses, flight fitness, and physiological effects and responses. A total of 185 citations are included in this bibliography.

SUBJECT INDEX

AUTHOR INDEX