

NASA Technical Memorandum 100016

Research Papers and Publications (1981-1987): Workload Research Program

Compiled by Sandra G. Hart

(NASA-TM-100016) RESEARCH PAPERS AND
PUBLICATIONS (1981-1987): WORKLOAD RESEARCH
PROGRAM (NASA) 124 P CSCL 05I

N88-12924

g3/53 Unclass
 0111142

August 1987

Research Papers and Publications (1981-1987): Workload Research Program

Compiled by Sandra G. Hart, Ames Research Center, Moffett Field, California

August 1987



National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

**RESEARCH PAPERS AND PUBLICATIONS (1981 - 1987):
Workload Research Program**

Sandra G. Hart
NASA-Ames Research Center
Moffett Field, CA

ABSTRACT

This document contains an annotated bibliography of the research reports written by participants in NASA's Workload Research Program since 1981. It represents the results of theoretical and applied research conducted at Ames Research Center and at universities and industrial laboratories funded by the program. The major program elements include: (1) developing a fundamental understanding of the concept of workload, (2) Providing valid, reliable, and practical measures of workload, and (3) creating a computer model to predict workload. The overall goal is to provide workload-related design principles, measures, guidelines, and computational models. The research results are transferred to user groups by establishing close ties with manufacturers, civil and military operators of aerospace systems, and regulatory agencies; publishing scientific articles; participating in and sponsoring workshops and symposia; providing information, guidelines, and computer models; and contributing to the formulation of standards. In addition, the methods and theories that have been developed have been applied to specific operational and design problems at the request of a number of industry and government agencies.

TABLE OF CONTENTS

Overview of the program	1
Background	1
Conceptual Framework	2
Workload Measurement Techniques	3
Primary Task Performance Measures	3
Secondary Task Measures	4
Subjective Rating Scales	4
Physiological Measures	5
Simulation and Inflight Validation	6
Summary: Phase I	7
Future Plans: Phase II	8
Workload Prediction	8
Relationship between Training and Workload	8
Appendix A: Grants and Contracts Funded	10
Appendix B: Comments by Program Participants	12
Appendix C: Research Publications (1981)	27
Appendix D: Research Publications (1982)	29
Appendix E: Research Publications (1983)	35
Appendix F: Research Publications (1984)	48
Appendix G: Research Publications (1985)	60
Appendix H: Research Publications (1986)	73
Appendix I: Research Publications (1987)	94
Appendix J: Research Publications (in press)	101

OVERVIEW

This document contains an annotated bibliography of the research reports written by participants in NASA's Workload Research Program since 1981. It represents the results of theoretical and applied research conducted at Ames Research Center and at universities and industrial laboratories funded by the program. The major program elements include: (1) developing a fundamental understanding of the concept of workload, (2) providing valid, reliable, and practical measures of workload, and (3) creating a computer model to predict workload. The overall goal is to provide workload-related design principles, measures, guidelines, and computational models. The research results are transferred to user groups by establishing close ties with manufacturers, civil and military operators of aerospace systems, and regulatory agencies; publishing scientific articles; participating in and sponsoring workshops and symposia; providing information, guidelines, and computer models; and contributing to the formulation of standards. In addition, the methods and theories that have been developed have been applied to specific operational and design problems at the request of a number of industry and government agencies.

BACKGROUND

The concept of workload has received an increasing amount of attention during the past decade, prompted by the realization that the human operators of advanced aircraft represent a limiting factor at the same time that their unique skills and capabilities remain an essential component. Automation has been offered as a solution to an increasing number of workload-related problems that have been found in existing systems or that have been predicted for systems under development. However, automation often simply replaces one source of workload for another, rather than accomplishing a significant reduction. In addition, there has been an ever-increasing tendency to reduce the number of crewmembers. For example, many civil transport aircraft now operate with two, rather than three, crewmembers and single-pilot operations have been proposed for the Army's most advanced helicopter (the LHX). Again, automatic subsystems are proposed to moderate the demands thus placed on the remaining crewmembers. Attempts to completely replace humans by automatic systems have failed, however, because human capabilities, adaptability, and flexibility continue to surpass those of the most advanced and sophisticated systems.

If pilots could perform all of the tasks that are required of them accurately and within the allowable time constraints using available equipment, workload would be of little practical importance. Because they often cannot, accurate predictions and assessments of workload are needed at all stages of design to develop optimal vehicle configurations, determine minimum crew complement, establish mission requirements and procedures, and specify the operational envelope for specific missions and vehicles. Thus, interest in workload, from an applied perspective, has stemmed from the assumption that workload has a direct impact on performance. Finally, the workload imposed on pilots is one of the final tests against which the adequacy and feasibility of operational requirements, system design, and training procedures must be tested.

Because the concept of workload includes numerous and diverse dimensions, many of which are not within the usual domain of experimental psychology, academic interest in workload lagged behind the research requirements of the operational community. Thus, most of the early work in this field was performed by engineers and designers tasked with implementing design requirements, and by military and civilian organizations tasked with evaluating the final products. Since these individuals generally did not have an extensive knowledge of the human performance, memory, and attention literature, they tended to rely on analytical approaches (that focused on observable activities and time lines) and informal subjective evaluations of engineering test pilots.

It was not until ten years ago that well-controlled, theoretically-motivated research in the field of workload began to be conducted in universities. During the same period, prompted by requirements to

specify the minimum crew complement for a new generation of transport aircraft and to evaluate the feasibility of single-pilot operations for advanced rotorcraft, interest in workload assessment and prediction peaked in the government and industry. However, much of the research performed during this period has not been directly applicable to the design and operation of advanced aircraft because individual reports were either microscopic in focus and phrased in psychological rather than engineering terms, or they were vehicle specific and proprietary. Nevertheless, it does form a data base upon which meaningful, valid, and reliable workload assessment tools and predictive models can be based.

In 1981, NASA formed a Workload Assessment Program to address many of the issues raised above. The goal was to merge the theoretical information about workload available from academia with the practical requirements of industrial and government organizations to develop a comprehensive definition, practical, useful measures and predictors, and workload standards. Throughout the program, basic research provided answers to theoretical questions in the well-controlled environment of the laboratory while simulation and inflight research provided verification that the results were valid and meaningful in the "real world."

Such issues as the relationship between workload and training, the relative demands imposed by vocal or manual inputs and visual or auditory displays, the association between imposed demand levels, achieved performance, and different measures of workload were addressed. In addition, the information provided by different types of measures, and when each can (and cannot) be used, were determined. Laboratory research provided answers to specific questions in a well-controlled environment, while simulation and inflight research verified that the results were meaningful in an operational environment. The results of this fundamental research effort are now being applied to a variety of vehicle-specific problems.

CONCEPTUAL FRAMEWORK

The first phase of the program was devoted to understanding the factors that influence pilot workload, evaluating existing assessment techniques, and developing new techniques. Because the workload experienced by pilots flying complex missions reflects many factors, developing a generally accepted conceptual framework within which to attack the problems of definition, measurement, and prediction proved difficult. Different researchers, focusing on whatever aspects of workload they included in their definition, manipulated and measured literally different phenomena. Yet, all used the same term (workload) in discussing their results.

The earliest conceptualizations of workload focused on the physical effort required to accomplish a task, defining workload in terms of physiological exertion. Analytical approaches focused on the number and duration of required activities, expressed in task and time-line analyses. Workload was defined as the relationship between the time needed to perform required tasks and the time available. Objective task demands were the foundation of this approach, rather than the behavior and responses of the individual performing a task. Both of these conceptualizations ignored the cognitive demands that were becoming an increasingly important component of the requirements placed on the pilots of advanced aircraft. In addition, early analytic approaches assumed that subtask elements would be performed serially. Since it is obvious from casual observation that people often perform several activities at the same time, concepts of divided attention, single or multiple "pools" of resources for information acquisition, processing, and response, and models of information-processing structures became important concepts in the field of workload assessment.

We defined pilot workload as the cost incurred by the human operators of complex airborne systems in accomplishing the operational requirements imposed on them. This cost reflects the combined effects of the demands imposed by mission requirements, the information and equipment provided, the flight environment, pilots' skills and experience, the strategies they adopt, the effort they exert, and their emotional responses to the situation. This is a pilot-oriented conceptualization, and reflects our belief that workload arises from the interaction between a task and the performer, and, thus, cannot be inferred from information about either in isolation.

The demands imposed on pilots are created by what they are asked to achieve (e.g., the objective goals of a flight and requirements for speed and precision) and when (e.g., schedules, procedures, and deadlines). Some flight tasks are intrinsically more demanding than others, and the difficulty of almost any task can be altered by a requirement for additional speed or accuracy. System resources (e.g., controls, displays, automatic subsystems, other crewmembers, and ground support) define how pilots accomplish task demands. Poor display design, inaccessible controls, poor handling qualities, and too much or too little information can increase workload substantially. Finally, where a task is performed (e.g., geographical location, altitude, time of day, weather) may also affect workload. For example, visual workload may be increased by low visibility, physical workload may be increased by turbulence, and threats from natural or man-made sources increase stress-related components. These elements may act independently or they may interact, enhancing or mitigating each others' effects.

Finally, the level of workload experienced by a particular pilot performing a specific task is determined by his basic skills, knowledge, and training; unskilled or novice pilots often experience greater workload than more skilled or experienced pilots. In addition, incorrect strategies, insufficient effort, or pilot errors can increase workload, due to the need for detecting, resolving and recovering from the problems created by the pilots themselves. Finally, pilots' expectations, previous experiences, and physical and emotional states affect their subjective experiences as well as their performance. Thus, although the "work" that is "loaded" on a pilot is an important contributing factor, workload reflects a number of other factors as well.

WORKLOAD MEASUREMENT TECHNIQUES

Despite its complexity, workload is assumed to be an important and practically relevant entity and a number of valid, sensitive, and reliable measures have been developed. However, it is clear that different measures are needed to evaluate different components of workload because the causes and manifestations of workload are so complex. Workload measures are usually organized into four categories: (1) objective measures of primary task performance, (2) objective measures of secondary task performance, (3) subjective ratings, and (4) physiological recordings. Each approach has advantages and disadvantages and there are limitations in the range of activities and questions to which it applies; the evidence they provide may or may not be useful, depending on the situation.

Primary Task Performance Measures

Performance is the driving force behind workload evaluation in operational or manufacturing environments. It has been assumed, without empirical support, that high levels of workload will result in (1) an increase in errors, and (2) an abrupt and catastrophic decrement in performance. However, it is also possible that errors may occur when workload is too low (due to inattention) and that increased task demands will result in strategy shifts as often as performance breakdowns.

Performance measures often provide little indication of the effort that a pilot exerted in achieving them; as demands are increased, pilots generally put forth additional effort (to the limits of their capabilities) to maintain a consistent level of performance. In addition, many measures of performance reflect the characteristics of the system rather than the activities of operator directly. Finally, a common set of performance measures do not exist that can serve as workload indices across different tasks. Thus, although it is always necessary to obtain performance measures to determine the degree to which a pilot was able to accomplish the task requirements, these measures may not reflect the pilot's workload unless they reflect behavior directly and are sensitive to changes in the pilot's effort as well as to changes in imposed task demand levels.

Measures of flight-path deviation can provide an objective summary of how well a pilot managed his vehicle to achieve smooth and precise flight-path control. Deviations often indicate periods of time when a pilot is sufficiently overloaded by other actions that primary flight-path control suffers. In addition, the rate, content, and consequences of communications can provide an objective index of the workload imposed on pilots; a standardized taxonomy of communications has been developed in which *a priori*

estimates of the workload imposed by communications tasks have been quantified. In addition, errors and delays in response might indicate the presence of high workload levels. Because each measure of performance may provide different answers to questions about how well a pilot accomplished a complex task, a method of integrating the information provided by available performance measures is needed. The contribution of different measures to the weighted combination must reflect their importance to the overall success of a mission and accommodate the fact that performance on each task component may be quantified with different indices and are compared to different objective and subjective criteria.

Secondary Task Performance Measures

Because primary task performance measures do not always reflect the cost of task performance to a pilot, it has been suggested that additional tasks could be imposed that would provide an indirect indication of the resources required to perform the primary, flight-related tasks; as primary task demands are increased, secondary task performance should degrade in direct proportion. The intent was to discover a secondary task "yardstick" that could be used to compare the workload of different tasks. The fact that specific secondary tasks were found to be differentially sensitive to particular types of primary tasks prompted a remarkable increase in interest by the academic community in the field of workload assessment. Competing models of attention and performance were applied to discover the structure and allocation of human resources, and a more scientific approach to the field of workload assessment evolved. A driving force behind this research was the multiple-resources model which provided a very useful structure within which many experiments were designed and data interpreted.

A number of secondary task workload measures have been developed and tested in laboratory and simulation research. In general, they represent simple activities for which the input (visual and auditory stimuli) and the output (verbal and manual responses) can be quantified accurately and directly. The intervening cognitive processes are predicted from psychological models and inferred from variations in the speed and accuracy of performance. However, these tasks were designed for purposes other than workload assessment. Many of them, such as choice reaction time, memory search, and time estimation, were designed to develop and test theories of human performance, memory, and attention. Their focus is narrow, the range of factors manipulated limited, and their relevance to subjects questionable. Others were developed as simplified versions of "real-world" task components to answer specific questions in a controlled environment. They have better face validity, but lack the benefit of a theoretical foundation.

Although several of these tasks were found to be very sensitive to variations in task demands in simulation research, they are generally inappropriate for use in flight, because they are difficult to implement and might compromise the safety of flight. Some measures, such as time estimation, can be included in the primary flight task as a natural component -- an embedded secondary task -- with minimal instrumentation and intrusion on primary task performance, however. This and other embedded measures have been shown to be sensitive to the workload of different activities in simulated and inflight experiments.

Subjective Rating Scales

Subjective ratings have been used throughout the history of workload measurement because they have face validity and are easy to obtain. However, they were scorned by experimental psychologists for many years as examples of the discredited field of Introspectionism. Nevertheless, they may come closest to tapping the essence of workload because they provide a direct indication of the impact of flight-related activities on pilots and they integrate the effects of many workload contributors.

One of the earliest rating techniques used in the aerospace industry was developed by pilots and engineers: the Cooper-Harper Handling Qualities Rating Scale. This scale addressed workload only indirectly, however. Other scales developed explicitly for evaluating workload were not standardized or validated and never achieved general acceptance. Furthermore, the ratings were characterized by substantial variations of opinion among raters. One of the causes of this variability was that pilots respond to and consider different aspects of complex tasks when they provide ratings. In addition, the

factors that contribute to workload vary between tasks. Research on these issues, coupled with the emerging interest in creating tools for eliciting expert opinions by decision theorists and expert system developers, prompted the design of multi-dimensional rating scales that could deal with differences in the sources of workload among tasks and variations in workload definition among raters.

Several subjective assessment techniques were developed by participants in the program. One of the earliest was a modification of the Cooper-Harper Handling Qualities Rating Scale, worded so as to focus on workload more directly. This scale was tested in several simulation experiments, and was found to be one of the most sensitive of the many measures that were evaluated. In addition, the concept of using magnitude estimation methods to quantify subjective workload experiences was tested. Single-dimensional ratings of task difficulty were obtained for different single- and dual-task combinations of laboratory tasks in comparison to a reference task. Although the concept of providing a reference task to anchor workload ratings is extremely valuable, it was found that the magnitude of the ratings was influenced by the reference task used. This provided a note of caution about the importance of selecting an appropriate reference task.

The NASA Task Load Index (TLX) was developed to provide an estimate of overall workload based on a weighted average of six subscales: physical demands, mental demands, time pressure, own performance, effort, and frustration. These factors represent task-related, pilot-related, and environmental factors. Through extensive laboratory, simulation, and inflight research, they were found to be the minimum number of dimensions required to describe workload experiences across many activities. The weight given to each factor reflects its importance to each rater in creating the workload of a specific task. This technique is based on the assumptions that workload experiences are created by different factors in different activities, that the magnitudes of these demands vary within and between tasks, and that individuals faced with apparently identical task demands experience different levels of workload.

Physiological Measures

The earliest conceptualizations of workload focused on the physical exertion required to accomplish tasks. Measures of physical effort, such as oxygen up-take and heart rate, were used to quantify this component of workload, reflecting a medical, rather than a behavioral or psychological focus. Since these measures did not reflect variations in mental workload, other physiological responses that do reflect cognitive processes (such as event-related cortical brain potentials and heart rate variability) were investigated. This development brought psychophysicists and cognitive psychologists into the field of workload assessment.

Physiological measures generally have the advantage of being unobtrusive. That is, they can be obtained without requiring attention from the pilot or interfering with the flight. In addition, since they can be recorded relatively continuously, they can reflect momentary fluctuations in workload. Finally, they provide an objective indication of involuntary physiological changes that often accompany workload changes. Their primary disadvantage is that physiological measures reflect nonspecific responses to different sources of stress. These responses may reflect the demands imposed by the flight, the environment, or the pilot directly, or other factors that are only indirectly related to workload. Such measures may, however, provide an integrated indication of the total impact of a flight on the pilots that does not also reflect the characteristics of the system (as many performance measures do) or pilots' biases and misconceptions (as subjective ratings do).

Heart rate reflects the stress associated with specific flight-related activities; it increases as some aspects of workload are increased. For example, heart rates are typically elevated during take-off and landing and return to baseline levels at altitude. In addition, substantially greater increases are found for the pilot flying during take-off and landing than for the pilot not flying. It is possible that the feeling of responsibility and level of preparedness that must be maintained by the pilot flying could result in their elevated levels of arousal. Heart rate is relatively insensitive to variations in mental workload, however.

Heart-rate variability reflects even subtle variations in mental workload; it decreases as the difficulty of a task is increased. A method of obtaining online estimates of heart-rate variability has been developed that reflects workload variations. This technique measures the heart rate interbeat interval and computes

the power in the 0.1 Hz region of the frequency spectrum -- an adaptation of the "Mulder" technique. The signal-processing algorithms have been completed and a prototype device has been built. Validation studies performed with a laboratory simulation of a vehicle-control task have demonstrated excellent agreement between experimentally imposed variations in workload and the output of the device.

Event-related cortical potentials have been proposed as a measure of workload because variations in the amplitude of different components of the waveform that follows the presentation of relevant information can be used to evaluate the focus of the task performer's perceptual resources and as a measure of the information-processing load. This measure can be treated as any other type of dependent variable; it derives its meaning from the setting in which the measurement was made. If a task is designed so that a clear relationship can be drawn between variations in the amplitude and latency of specific components of the waveform, then this measure can provide an unobtrusive indication of the workload at that specific time. Its primary drawback is that it has not yet been tested in flight, and only limited simulation research has been performed. However, recent simulation research results suggest that it is a promising technique.

Simulation and Inflight Evaluation of Measures

Measures that demonstrated sensitivity to different types of imposed demands, methods of presentation, cognitive processing requirements, or response modalities, were then evaluated in the context of more complex activities. Part-task aircraft and supervisory control simulations provided an environment in which multiple, overlapping sources of task demands and response requirements could be imposed. Here, the sensitivity of each measure to specific or global sources of workload was evaluated. Some measures, such as subjective ratings, provided an integrated measure of the overall demands imposed during the interval evaluated. Others, such as secondary tasks and evoked cortical potentials, provided information about momentary workload levels at specific instants in time. Primary task performance measures generally reflected the effort exerted by the subjects, rather than the absolute levels of imposed demands. However, some aspects of performance were found to be more sensitive to variations in behavior (e.g., smoothness of control, timekeeping), than others, providing objective indicators of workload.

The practical utility of these measures in complex environments was investigated as well. Here, it was found that some secondary task measures either interfered with primary task performance or were ignored when workload became too high, while others did not. Physiological recordings and primary task measures, which did not require overt, additional responses from the subjects, were obtained without degrading or altering primary task performance. In addition, some of the more sensitive performance measures (e.g., control variability and communications rate) were available in simulators without additional instrumentation. Physiological measures, on the other hand, did require additional recording devices. However, it was found that visual or auditory signals could be presented, even in the presence of competing information and responses required for primary task performance, that could evoke discriminable patterns of brain activity that reflected variations in primary task workload. This, at least partially, addressed one of the criticisms of this method. It was found that subjective ratings, which could interfere with primary task performance if given on-line, could be obtained without interference by using structured post-task debriefings. These retrospective ratings were surprisingly sensitive to segment-by-segment variations in workload and correlated highly with measures that were obtained during the flight.

A study conducted in the Ames Vertical Motion Simulator is one example of such a study. In this experiment, several stability and control augmentation systems, coupled with different levels of automation provided alone or in combination were evaluated to compare single- and dual-pilot performance and workload during low-level military operations in the NOE environment. In this experiment, two workload rating scales, the Cooper-Harper Handling Quality rating scale and heart rate measures, were used to evaluate the effects of the experimental manipulations on the pilots. All of the measures provided converging evidence that single-pilot workload levels were high, unless significant levels of automation were provided.

The final requirement in developing and testing workload measures is inflight verification. Although a simulation provides an analogue of the operational environment, elements are missing there that cannot

be replicated, and the practical constraints for applying some measures are less problematical than they are in flight. A number of the measures developed through laboratory and simulation research were evaluated in flight in the NASA Kuiper Airborne Observatory (KAO) and in an SH-3G helicopter. In the experiment conducted in the KAO, no experimental control was possible over the missions flown. The subjective and objective measures were obtained during roughly equivalent flight segments, and the results were compared across segments. Even with this complete lack of experimental control, it was clear that each of the measures provided useful and complementary evidence about pilot workload. This experiment provided information about the practicality of these measures in a flight environment, however, it did not provide a final validation of the measures because (1) the tasks each crew performed were somewhat different, (2) the demands of each task were not measured independently nor predicted in advance, and (3) no objective measures of aircrew performance were available against which to compare the workload results.

In the second experiment, conducted in an SH-3G helicopter, evaluating the utility of different workload measures was the primary focus of the experiment. Specific missions were defined in advance and flown by each crew. The flight scenarios included straight and level flight above 3000 ft and contour flight, visual landings at an auxiliary site, instrument landings at airfields, hover in and out of ground effect, visual search patterns, and visual and instrument navigation conducted between Moffett Field and Crows Landing. The workload measures included pilot ratings, secondary tasks, heart rate and heart rate variability, communications, and selected performance measures. Since portions of the flight were conducted on an instrumented flight-test range, objective measures of performance, often unavailable in flight, could be obtained.

In this experiment, it was found that similar estimates of workload were obtained when the same tasks were performed at different times in the flights. For example, all of the visual landings were given the same, low workload ratings. Subtle variations in tasks, however, prompted differences in workload measures that were in the predicted direction. For example, both primary and secondary performance measures and subjective ratings differed for hover tasks performed in and out of ground effect. As the environmental constraints imposed for different contour flight segments were increased, so did the measured levels of workload.

SUMMARY: PHASE 1

The first phase of the program has been essentially completed: the factors that contribute to pilot workload have been identified and a set of valid and practical measures have been developed. These measures are now being implemented to solve operational problems posed by the military, civil and public-use operators, and industry.

Since selecting an appropriate and practical measure of workload is difficult due to the multi-dimensional nature of workload and because different measures are selectively appropriate for different questions, tasks, and test environments, we developed a micro-processor-based expert system, WC FIELDE, which is available for public distribution to aid in this process. Although hundreds of articles have been written describing the results obtained with one or two techniques and a specific task, it is difficult for individuals who are not intimately familiar with the literature to know what measures are available, how well they have been tested, and when they can be used. Thus, the goal of this system is to integrate, organize, and evaluate information about workload assessment techniques and to make it readily available to human factors practitioners who are not experts in the field of workload per se.

The system suggests measures, in descending order of utility, based on a users' answers to questions about his goals, research environment, and available facilities. It draws from a data base of widely used measures and "rules-of-thumb" provided by experts in the field to propose alternatives. In addition, it provides sufficient information for the user to make an informed choice among the suggested alternatives and to implement the techniques included in the data base. Each measure is described and evaluated, studies in which it has been used are reviewed, and references are provided to allow the user to obtain additional information.

FUTURE PLANS: PHASE 2

The primary goals of the second phase of the workload program are to (1) complete and apply a computer model for workload prediction in advanced helicopters, (2) develop and publish criteria for workload (e.g., determine how much workload is "too much" or "too little"), (3) continue to support the workload research requirements of civil and military users and industrial designers and manufacturers, and (4) investigate the associations among workload, training and performance.

Workload Prediction

After several years of research on the structure of pilot workload, and developing and applying workload assessment techniques, a computer model to predict pilot workload in current and advanced helicopters is being developed. In a research environment, workload predictions are essential so that known levels of workload can be imposed to evaluate candidate measures. In an applied environment, such predictions are essential so that the potential impact of design decisions on pilots can be known early in the design process. Again, laboratory research provided the initial equations by which the workload levels of task elements were determined, measured, and combined to derive predictions for complex tasks. Here, it was found that the workload levels of subtasks performed individually, but concurrently, could be added together to predict the performance of the combined task. Subtasks that were functionally related or shared common information, processing, or response requirements, created lower levels of workload in the combined task than would be predicted from simply summing their individual workload levels.

Experienced workload is the integrated product of many factors in addition to the objective demands that are placed on a pilot. Although workload predictions, particularly those made during the design of a new system, must necessarily focus on the objective demands that are imposed on a pilot, there are a number of other types of information that might be included to enhance the predictive power of such a model. Our approach has been to start with nominal or typical flight segments or mission elements. Information about their duration, intensity, overall workload, and visual, auditory, information processing, and manual control requirements are obtained. A data base of additional tasks or events that might occur during any flight segment are identified and the same information that is obtained for the nominal segments is obtained for them. The functional relationships among specific segments and additional tasks are defined so the model can select the appropriate combination algorithms with which information about individual tasks and segments that must be performed concurrently can be combined to estimate the workload of the complex task.

A preliminary model was developed based on this structure. The predictions of the model were tested in simulation research, and were found to correlate well with objective and subjective measures of workload obtained in simulated flight. The full model is under development. The predictions of workload made by this model will be incorporated into the Army/NASA Aircrew-Aircraft Integration Program (A³I) model under development at Ames. These predictions of this Computer-Aided Design/Human Factors Engineering Workstation will allow the designer of system, subsystem, or mission element for an advanced helicopter to test the effects of the design element on the potential pilot-population in software using models of human performance, memory, perception, training requirements, and so on in conjunction with models of environmental factors and vehicle dynamics and control. With this workstation, potential problems can be identified during the conceptual stage, thereby avoiding expensive and time-consuming cut-and-try methods.

The Relationship Between Workload, Training, and Performance

Our interest in training evolved from its apparent influence on workload. Training is often proposed as a solution to workload problems, as it is assumed that both training and workload are equally improved by training. However, the two research areas rarely, if ever, overlap, and there is little empirical evidence to support such assumptions. Since training costs are escalating rapidly, it is imperative that training methods are developed that make optimal use of available time and facilities. To accomplish this it will

be necessary to monitor the workload of trainees to ensure that it is low enough to allow learning to take place (yet not so low so as to waste valuable training resources) and to make logical selections of training elements and promotion rules to optimize training time.

Ames sponsored two workshops jointly with the Army to initiate this program element. The topic of the first workshop was the relationship between workload and training. The topic of the second workshop was individual differences in pilot selection, training, workload, and operational performance. Participants were invited from academia, industry, and the government to discuss workload and training and their relationships in the context of advanced helicopter and space station operations. The first workshop has been described in an Executive Summary, and the information presented at both workshops will be published in book form. The meeting was a great success in acquainting members of different research communities, revealing their problems, and discussing how to improve the flow of information and support among industry, academic, and government research laboratories.

The training research portion of the program included theoretical studies about optimal training strategies, the development of evaluation criteria for training programs (that take trainee workload into account), and the application of these methods to operational problems.

SUMMARY

At each stage in the research process, information obtained in more realistic situations was used to refine theoretical models and provide the focus for well-controlled laboratory studies to address specific issues. By moving back and forth between these research environments, the requirements of theoretical development were balanced against the requirements of the "real world." Furthermore, operational relevance was ensured at the same time that the predictive advantages of a theoretical foundation was maintained. The program allowed theoretical researchers to become familiar with applied problems (through participation in simulation and inflight research) and exposed designers, engineers, and operational test and evaluation personnel to the advantages of experimental control, a theoretical foundation, and the use of validated measures. The verbal and written reports provided by participants in the program represent a theoretically sound, operationally tested body of information that can be used by industry and government organizations to estimate the impact of their design and requirement decisions on the users of current and advanced systems from early in system design to their operational use.

APPENDIX A: GRANTS AND CONTRACTS

Arizona State University (NCC 2-202) "Examining the Relation between Subjective Estimates of Workload and Individual Differences in Performance."

Principal Investigator: Dr. D. Damos

Behavioral Institute for Technology and Science (BITS) (NCC 2-228) "A Theoretical Approach to Measure Workload."

Principal investigator: Dr. B. Kantowitz

Behavioral Institute for Technology and Science (BITS) (NCC 2-228) "Toward a Dynamic Mathematical Theory of Mental Workload."

Principal investigator: Dr. J. Townsend

Douglas Aircraft Company (NAS2 - 11860) "Mental Workload Measurement."

Principal investigator: Dr. M. Biferno

General Physics Corporation (NAS2-11562) "Communications Workload for Transport Category Aircraft."

Principal investigators: Dr. F. Gomer, Dr. L. Silverstein, Dr. S. Eckel

Massachusetts Institute of Technology (NAG 2-229) "The Relationship between Aircraft Control Automation, Mental Workload, and Pilot Error in a Laboratory Simulator."

Principal investigator: Dr. T. Sheridan

Ohio State University (NAG 2-184) "Pilot Performance and Workload Assessment: An Analysis of Pilot Errors."

Principal Investigator: Dr. R. Jensen

Purdue University (NCC 2-235) "Timesharing Performance as an Indicator of Pilot Mental Workload."

Principal investigator: Dr. B. Kantowitz

San Jose State University (NCC 2-34) "Flight Management Research."

Principal investigators: Dr. R. Ginsberg, Dr. K. Jordan

Search Technology (NAS2 - 12048) "Causes of Human Error."

Principal investigators: Dr. W. Rouse, Dr. N. Morris

SRI International "Comparison of Type A and Type B Individuals."

Principal investigator: Dr. M. Chesney

Structural Semantics (NAS2 - 11052) "Linguistic Methodology for the Analysis of Aviation Accidents."

Principal investigators: Dr. C. Linde, Dr. J. Goguen

Technion, Israel Institute of Technology (NAG 2-229) "Assessment of Workload in Engineering Systems."

Principal investigator: Dr. D. Gopher

Technion, Israel Institute of Technology (NAGW 1012) "Using Complex Computer Games as General Trainers to Improve Flight Skills."

Principal investigator: Dr. D. Gopher

United States Air Force Academy "The Subjective Measure of Workload: Individual Differences in the Perception of Factors that Influence Workload."

Principal Investigator: Maj. J. Swiney

University of California, Los Angeles (NAG 2-216) "Model-based Approaches for Partitioning Subjective Workload Assessment."

Principal investigator: Dr. J. Lyman

University of Illinois (NAG 2-169) "An Investigation of the Basis of Subjective Ratings of Mental Workload."

Principal investigator: Dr. C. Wickens

University of Illinois (NAG 2-308) "Human Performance and Workload in Automated Systems."

Principal investigator: Dr. C. Wickens

University of Illinois (NCC 2-380) "Workload and Training: An Examination of their Interactions."

Principal investigator: Dr. Emanuel Donchin

University of Illinois (NAG 2-369) "Event-related Brain Potential Indices of Cognitive Workload and Automaticity."

Principal investigator: Dr. A. Kramer

University of Southern California (NCC 2-379) "Temporal Factors in Mental Workload."

Principal investigator: Dr. P. Hancock

University of Toronto (NAGW - 429) "Development of Fuzzy Set Calculus for Estimating Pilot Workload as a Function of Modes of Operator Behavior."

Principal investigator: Dr. N. Moray

Wayne State University (NNC 2-230) "Analysis of Error Identification and Description in Simulation."

Principal investigator: Dr. R. Frankel

Western Aerospace Laboratories "Performance Assessment in Mental Workload."

Principal investigator: Mr. M. Bortolussi

APPENDIX B: COMMENTS FROM PROGRAM PARTICIPANTS

Dr. Emanuel Donchin
Department of Psychology
University of Illinois
Urbana-Champaign, IL

DEFINITION

The concept of mental workload arises most overtly, though not exclusively, within the context of the design of large, expensive, and complex systems, such as aircraft, in which operators are required to process large amounts of information, usually under conditions that leave little time for planning and reflection. One goal often set before the designers of such systems is the minimization of the mental workload the system imposes on the operator. As a general statement of the desirability of "good designs" this is indeed a desirable goal. Yet, it is evident that a systematic attempt to "minimize" workload requires that the term be defined with precision and that the designers have access to proper techniques for measuring workload.

This measurement problem appears deceptively simple. It is tempting to think that the workload associated with a task can be inferred directly from a description of the task. Is it not obvious that the more the operator has to do, the higher the workload? Unfortunately, matters are considerably more complex. It turns out that it is not possible to predict how a person will cope with a task solely on the basis of detailed information about the task. It is necessary to consider the capacities, mental and physical, cognitive and affective, that the operator brings to the task. The demands that a task imposes on a person will prove light, or excessive, depending on that person's skills, abilities, memory, attention, and basic knowledge.

It is convenient to adopt language and metaphors borrowed from economics in this context. The operator is assumed to have at his disposal an ensemble of "resources." The term resource refers here to whatever it is that a person needs in order to achieve successful performance of the task. These resources are assumed to be available in finite, limited, amounts and as they are required by many concurrently performed tasks, their allocation determines which tasks will be performed successfully and which would fail. Thus, the operator is viewed as purchasing performance as a "cost" in resources. Workload is the term applied to this cost. It is important to emphasize, however, that the cost that a task imposes on a person is best specified in its relative, rather than its absolute, value. What is critical are the demands that the task imposes on the resources relative to the resources available to the operator.

Workload, then, is a hypothetical construct that embodies the interaction between task demands and the available mental resources. As workload is a measure of an interaction, it cannot be obtained by measuring one of the interacting elements. Neither information about the task alone, nor on the operator alone, can serve as a measure of workload. To use an economic analogy, objective task "difficulty" can be viewed as equivalent to the price tag of a car, a price tag that is specified independently of the customer. Task "workload," by contrast, is the difficulty one experiences in buying the car, depending on one's current fortune. The term "costload" may be coined to refer to this relative cost.

IMPORTANCE

The measurement, and the prediction of mental workload is of considerable practical importance. Thus, for example, the development of reliable techniques for the assessment of workload is listed first in a ranking of the 157 research needs prepared by the Federal Aviation Administration. I note this ranking with great interest because for the past 15 years my colleagues and I at the University of Illinois have been examining the feasibility of using the endogenous components of the Event Related Brain Potential (ERP), with particular emphasis on the P300 component, as metrics for mental workload.

RELEVANT MEASURES

How would one go about measuring the costload of a car? Clearly, the fact that a person purchased the car does not tell us how difficult it may have been to do so. The customer may have borrowed heavily to acquire the required resources. One way of determining the impact the purchase of the car has on financial resources is to observe the pattern of expenditures on other items. If, for example, after the car has been purchased, the customer drastically reduces expenditures on other items, one would assume, though not necessarily know, that a large proportion of the financial resource pool has been devoted to acquiring the car.

Note, that in this financial analogy, this measure of costload is based on the assumption that (a) there is a fixed pool of resources that can be allocated to serve a number of goals; (b) resources made available to one goal are not available to others; (c) the person is in control of the allocation of resources; and (d) the person cannot choose to buy a less expensive car. Given these assumptions, we can use the level of expenditure on a secondary item as a measure of costload. The lower the consumption of secondary items, the higher the costload associated with the primary item.

Precisely this logic, and these assumptions, underlie one of the common methods for assessing workload, the "secondary task" technique. To measure workload associated with a given task, designated "primary," the subject is assigned an additional "secondary" task. The subject is instructed to perform the primary task to the best possible level of performance (assumption d, above) and the performance on the secondary task is monitored. The poorer the performance on the secondary task, the larger the relative demands the primary task imposes on the person's resources, and, therefore, the larger the workload.

We have proposed, and provided extensive empirical support for the proposition that the "odd ball" paradigm used in the study of the P300 can serve as a secondary task in the measurement of workload. This paradigm, which requires subjects to count or otherwise respond to one of two events presented in a Bernoulli sequence, is particularly useful as a secondary task because, unlike the more traditional secondary tasks, it interferes minimally with the primary task. The experiments we have conducted shared a similar structure. A subject was assigned some primary task and concurrently had to monitor a Bernoulli sequence of probe stimuli. One of the elements in the sequence occurred considerably less frequently than the other. The P300 elicited by these rare events was monitored. The independent variable was the "difficulty" of the primary task and the dependent variable was the amplitude of the secondary task P300. We assumed that, as the difficulty of the primary task increased, so would the subject's workload and we predicted that the amplitude of the P300 would decline as the workload increased. The experimental results demonstrated that the claim that the P300 can be used as a metric for workload can be asserted with some confidence. Further, we determined that the reduction in P300 amplitude would be graded as the subject moved from fully focusing on the event to fully ignoring it. These studies can be viewed largely as attempts to identify the effect that various experimental manipulations have on P300. The experimental manipulations can be objectively described in such terms as "increased bandwidth of target movements" or increased tension on the response button." One may or may not relate these manipulations to such terms as "task relevance." If one does, then we are committing ourselves to a psychological model within whose framework the term "task relevance" must take meaning. A different approach, and one which I espouse, accepts psychological concepts solely within the framework of the theories in which they are embedded. The value of relationships that one assumes between an ERP component and a psychological construct depends on the degree to which it is possible, within the theory, to derive critical studies that play a useful role in testing the theory. The theory must predict how ERP measures will emerge from an experiment. With respect to task relevance, the key theoretical step has been the adoption of Resource Theory and, in particular, its Multiple Resources version as the matrix within which "task relevance" need be interpreted. Indeed, it was this theoretical transition that made it possible to develop the P300 amplitude metric for workload. The amplitude of the P300 elicited by a secondary task stimulus is interpreted, within this theoretical framework, as a measure of workload because it is taken to be proportional to the resources that remain available after the primary task has taken its toll. Thus, the P300 becomes a measure whose meaning is established within the context of a theory and a data base. That it proves to have a useful application in Engineering Psychology is a bonus of some consequence. Yet, an even more important aspect of the approach is that the use of a theoretical model generates specific tests of the interpretations of the component. The validity of the interpretations one makes of the P300 are thus tested in the crucible of science.

Peter A. Hancock
Department of Safety Science and Human Factors Department
Institute for Safety and Systems Management
University of Southern California
Los Angeles, California

DEFINITION

As workload is a multifaceted concept, any attempted definition is required to reflect this multidimensional characteristic. This has produced considerable problems in the search for a definition. Unlike its physical counterpart, mental workload is a phenomena of recent origin and is generated by the incapacity of the individual to transduce meaningful input information into effective output action. Such incapacity may be structurally, functionally, or temporally mediated, dependent upon both input and required output. Consequently, a global definition of mental workload is the symptomatic representation of the failure of human cognitive adaptability to reconcile the content of input information with the execution of effector action of either perceived or actual necessity.

IMPORTANCE

In some of our recent work concerning the real-time adjustment of task structure and loading between human and machine, the signal which allows the assessment of current human capability is derived through mental workload measures. Consequently, the role of valid workload assessment is central to the design and operation of current and future human-machine systems. To enact compensatory action, which would commonly take the form of load shedding by the human operator, the prediction of future workload in response to time varying task demand is a key component. Without the overall ability to integrate human-into-system action, the unique capabilities of the human operator are lost. Valid workload assessment is the tool which will allow this necessary integration to occur.

RELEVANT ISSUES

Relationship between Workload and Performance.

It can be argued that the issues of relationship of workload to performance and the relationship of workload to error are reflections of the same problem. We have argued that a more detailed analysis of what composes a task and how errors may be generated might clarify this picture. (It is encouraging to see recent insights into forms of error as elaborated by a number of groups e.g., McRuer/Jex, Senders/Moray). However, the link to mental workload is far from clear although it seems to have often been assumed that overload equals poor performance and increase in error. Good data on error are sparse because of the observational frequency and this problem may be magnified as a more detailed taxonomy of error types is forthcoming.

Individual Differences.

As with the above issues, it appears that individual differences and the relationship of workload and training can be equally regarded as related issues. Individual differences focus on the inter-individual variability while issues such as training and practice reflect intra-individual variation. It is probable that the two issues might benefit from mutual interaction, although at present there appears to be a general malaise in studying individual differences few good ideas seem forthcoming at this time.

Relationship between Workload and Training.

Elsewhere, we have expressed our ideas concerning workload and training through the medium of attention. The dynamic change in the experience of workload with training may be related to the discrimination and assimilation of task relevant cues and to the effective reduction of viable task solution paths that occur with prolonged practice.

RECOMMENDED MEASURES

Very much a case of "horses for courses." With the somewhat meager tools available, whenever possible it is sensible to collect as much information as is feasible, as in, for example, a laboratory setting. However, in operational environments it is essential to follow a parsimonious approach, quite simply as few measures as will reliably accomplish the job. This depends to a large extent upon the arena of operation. In our work we have been concerned with both real-time and non-intrusive characteristics. However, we are aware of the opinion of others which advocates different measures based upon somewhat different criteria.

ISSUES TO RESOLVE

The list is potentially endless. However, from our current efforts a primary concern is the resolution of sources of workload into endogenous or internally originating factors and exogenous or environmentally (task) based factors. It is this, of course, that forms the focus of our present combined work. We are using the passage of effective time as a potential avenue through which to achieve a first-pass resolution of this issue.

Barry H. Kantowitz
Purdue University
West Lafayette, IN

DEFINITION

I currently define workload as an intervening variable, similar to attention, that modulates or indexes the tuning between the demands of the environment and the capabilities of the organism. When I first started this research I had no coherent definition of workload and instead used the assorted and inconsistent definitions that have been offered by practitioners. I am now convinced that future progress depends upon using a definition that can be related to theory rather than to the often conflicting statements of practitioners trying to demonstrate the unique benefits of their own approach. While each individual pragmatic definition is useful, it is impossible to put them all together without theory.

IMPORTANCE

Predicting workload allows human factors specialists to design systems that match human capabilities. This is important for any system where errors are expensive and people are a necessary system component. Predictions of workload cannot be evaluated without measuring workload.

RELEVANT ISSUES

Relationship between Workload and Performance.

Workload and performance are not identical. Performance is observable; workload is not. Workload must be inferred from performance just as any intervening variable must be inferred. For example, learning is not performance but an inference drawn from a change in performance. If performance is terrible, we might reasonably suspect that workload is excessive. But when performance is adequate, we cannot make any direct statements about workload without additional measures.

Individual Differences.

Since there are individual differences in learning, it seems reasonable to expect individual differences in workload. I am not sure how important these effects might be. They may be small compared to the effects of training.

Relationship between Workload and Training.

Training reduces workload. In our experiments, pilots with more flight time showed reduced workload based upon objective secondary tasks. However, we have not examined this systematically. I see two important issues concerning training and workload. First, how can training be optimized to teach operators how to decrease workload? Given a particular system, it is important for operators to learn as quickly as possible; this optimization of rate is a traditional concern of trainers. Second, how does training alter the asymptotic performance of operators? In operational systems we need to specify the tradeoff between additional training and workload. For example, assume we are training secretaries on word processors using commercial software such as Wordstar. Training time can be minimized by only teaching the first few chapters of the manual. This will allow the immediate production of text. However, a trainee who has the opportunity to work through the later chapters will learn more efficient strategies for manipulating text and so will asymptote at a higher level of production. Both secretaries may have equal workload, but the latter is generating more output. Similarly, two secretaries may have equal output but differ in workload due to different training histories.

Relationship between Workload and Training

Workload is both a cause and an effect of errors. This outcome is completely consistent with the definition of workload given above.

RECOMMENDED MEASURES

I recommend choice-reaction time, time estimation, and sinus arrhythmia as effective measures of workload. I abjure simple-reaction time probes and Sternberg memory scanning tasks since the data they produce as secondary tasks are often impossible to interpret without making unwarranted assumptions about attention and capacity. I accept subjective ratings for their ease of use, but have reservations about their measurement properties.

ISSUES TO RESOLVE

A. Development of a Workload Theory.

Since workload is inferred, rather than observed, it can be explained only by theory. As I have argued in both a general chapter on workload and a more specific chapter relating workload to aviation, the best practical tool is a good theory. Theory fills in the gaps and allows us to predict workload in new operational settings where we lack data. Therefore, I believe that development of theory to guide workload research should be a high priority.

While attention theory is an excellent starting place, it is crucial to realize that a theory of attention is not necessarily a theory of workload. I have argued (Kantowitz, 1986) that single-pool theories of attention are most suited for predicting workload because they make the most of the construct of spare capacity. Spare capacity becomes hard to define in multiple-resource models of attention making these models less appropriate for guiding workload research.

B. Converging Operations.

We need to find operations that converge on workload as an intervening variable. This does not mean that we should use 27 varieties of dependent variables in every experiment. Instead we should sample carefully from the three major categories used to measure workload: subjective ratings, secondary tasks, and biocybernetic indices. Thus, a typical experiment that looked for converging operations might use bi-polar ratings, choice reaction secondary task, and sinus arrhythmia simultaneously.

One especially valuable technique would be to take a behavioral task that is generally understood, for example, the psychological refractory period effect or a Fitts' law task, and use it to calibrate variables that are less understood but potentially easier to implement in operational settings. This approach may prove especially useful with biocybernetic dependent variables such as heart rate and event-related potential.

C. Attention Operating Characteristics.

Attempts to relate dual-task performance have often used Performance Operating Characteristics. However, when different tasks are plotted on the two axes, severe scaling problems are encountered that make results difficult to interpret. These problems are minimized with Attention Operating Characteristics (AOCs), a sub-set of Performance Operating Characteristics. I am unaware of any workload research that has used AOCs and we have just started this at Purdue.

D. Simulation.

While traditional laboratory tasks are essential for discovering basic principles, they (by design) lack the complexity of operational settings. The power of modern microcomputers now makes it feasible to bring moderately complex simulations into the laboratory. The best example of this is POPCORN which now runs on an IBM PC AT instead of the large expensive graphics system used to develop it. This new tool permits controlled investigation of a person-machine on an order of magnitude more complex than those typically studied by experimental psychologists. However, progress will be impeded until a formal model or theory is created for POPCORN.

E. Workload Prediction.

All of our models, theories, and paradigms will not aid the aviation community unless we can prove that they work. While it is difficult to acquire data in actual flight, modern simulators provide a close approximation. The fruits of our labors need to be demonstrated in a simulated flight. While we have made progress in this area using GAT simulator, it would be nice to "graduate" to the more sophisticated simulators used for jet transports. Such data will validate our workload procedures and enable us to study more realistic issues such as the effects of flight deck automation upon pilot workload. While I have suggested that automation can both increase and decrease crew workload, these suggestions were based upon theory and require empirical support.

James T. Townsend
Purdue University
West Lafayette, IN

DEFINITION

Most of my theoretical and experimental work has been with a closely related concept, that of mental capacity and attention. Many of the aspects of theorizing and methodology are similar in the two fields. My definitions of "workload" and its close relative, "capacity" (and capacity expenditure) have not changed over the past three or four years. My conception of workload is that it is a theoretical construct which must be embedded in a well-formulated model (preferably mathematical) in order for us to make due progress. It is perhaps too early to expect a universal model of workload, but it is high time that rigorous models were constructed for the more local experimental and applied situations with which the field is now working. Such models should be psychologically and physiologically realistic. In most circumstances, it would be possible to falsify the model. Otherwise, we learn little about the true underlying processes. Further, a parameterized model should be sufficiently flexible (e.g., contain sufficient parameters) to meet the particular demands of the researcher but not so "rubbery" as to lose important uniqueness properties. We should expect that as our experiments and modeling become more sophisticated, a "canonical system" should begin to emerge which is capable of subsuming a rather wide base of applied and basic phenomena.

At the risk of diverging a bit from the main question, it is interesting to review briefly some stages in the evolution of "workload" and "capacity." After a flurry of more or less rigorous definitions of workload and related topics in the fifties, the failure to find a panacea doctrine led, in the later sixties and early seventies, to a devolution of rigor and the incursion of often apparently all-encompassing but too often vague and confusing definitions and methodological constructs. Along the way, we have also seen some interesting fairly tight theoretical statements which, unfortunately, have made little or no contact with either laboratory or applied data. The situation is, however, looking up. This is due to a number of developments, only a few of which I have space to mention here. One is the review and evaluation of older more rigorous theories (e.g., linear and quasi-linear systems theory, optimal operator theory etc.) in an effort to salvage what is useful and to build more realistic models from these earlier models. Another is the growing sophistication in the use of physiological recording techniques and their amalgamation with behavioral perspectives. An approach which shows great promise is the use of analytic (i.e., with closed mathematical formulas) and computer simulation models of psychologically realistic processes (as opposed to normative or ideal models). This approach has as yet seen little implementation in the area of workload research. The modeling concept will play a crucial role in the discussion which follows.

IMPORTANCE

Within the mission of NASA, how pilots and astronauts perform as a function of environmental, psychological, and physiological variables must be of critical interest. To measure this in a way that is meaningful in the long term view, as well as permitting some generality of conclusions and description, it is necessary to carry out fundamental laboratory research in addition to the obvious efforts which are required in more immediately applied settings.

RELEVANT ISSUES

Relationship between Workload and Performance.

Clearly important and can only be accomplished within the context of precisely specified and testable models. Otherwise circularity pollutes the research effort and stymies progress. That is, often operational definitions are given to theoretical concepts which involve an experimental result. If the result is found, the "theory" is proven. If it is not found, then the theoretical concept is not invoked so the "theory" is still saved.

Individual Differences.

Again of import. However, we need basic invariant laws which hold across individuals, and provide appropriate anchor points, in order to confer meaning on the concept of individual differences.

Relationship between Workload and Training.

I'm convinced much more could be done in modeling the learning process; from precise and completely testable laboratory models to more realistic and somewhat looser, but still eminently useful, approximative models for field training.

Relationship between Workload and Error.

There is emerging evidence on this (from some of our own work) that there is a feedback loop so that errors are a function of workload and that workload may be, in turn, affected by personal assessment of error rates. Again, this is an aspect which can and should be mathematically modeled.

RECOMMENDED MEASURES

1. Theoretical and experimental linkage of scaling variables (e.g., time pressure, task difficulty etc.,) with constructs in dynamic process models. This item is quite novel but could be of considerable significance in bringing measurement, theory, and methodology together.

2. Mathematical and computer modeling where possible. Where not, an intense effort to provide clean theoretical definitions of constructs with linkage to environmental (e.g., experimental, operational) variables.

3. Converging scaling operations which involve distinct techniques but that are all based on the same, hopefully important, variables.

ISSUES TO RESOLVE

At this stage, a great deal has been learned about a number of separate aspects of the workload problem. We have a pretty good idea of what won't work and what works approximately in certain situations. As mentioned above, several innovations and modifications of past efforts promise much progress on some fronts. We still need rigorous models that yield global, (but rigorous) qualitative, as well as quantitative, predictions even within relatively precise laboratory environments. In some settings, I expect it to be possible to formulate models that intercalate physiological parameters into psychologically based theoretical structures. Simulation models, in addition to intelligent use of psychometric and statistical techniques, should be extremely valuable in assessing concepts and theory in applied settings. As intimated earlier, one topic where these should be employed with benefit is the training scenario. My overall estimation is that our knowledge as well as our ability to apply that knowledge are emerging into a part of the curve that is positively accelerated so the next decade should be very exciting indeed.

DEFINITION

Workload is equal to the demands imposed by a task on the limited capacities of the human operator. Mental workload then is equal to the demands on the information and cognitive capacities of the operator. Because the human operator possesses multiple capacities or resources, workload is a vector rather than a scalar quantity. From the perspective of workload assessment, the two most important dimensions of this vector are perceptual/cognitive resources, and response-related resources. Although my views of what and how workload should be measured have been altered in the last three to four years; this fundamental definition has not been.

IMPORTANCE

I believe that these are extremely important issues. However, the two issues -- and prediction -- are quite separate and independent. The prediction of workload is important because it will allow system designers to identify periods of high workload ("choke points") as well as to predict with some degree of success which of two different system configurations may be preferable. These types of predictions, based upon the relative evaluation of two or more different systems, or different points in a mission, should be feasible to make. I am less optimistic about the prediction of absolute workload such as that involved in determining that the workload of system X at time Y will be "excessive" (i.e., above a cutoff "workload value"), or in certifying a particular system as adequate because its workload is less than some critical value. While I am not optimistic about the possibilities of attaining these absolute workload measures, I do realize the pressure on system designers to be able to provide them. In any case, models that will predict workload on either a relative or an absolute basis clearly remain a fundamental and extremely important part of the system design process. By acknowledging that system satisfaction is based upon criteria other than pure system performance, designers have clearly made a major step forward.

I see the measurement of workload as important for three separate reasons:

(1) Validating the predictive models.

To determine how well a model, as described in the preceding paragraph, works, it will be necessary ultimately to validate the model by assessing workload as the tasks whose workload was predicted are ultimately performed.

(2) Assessing and comparing systems.

Here again the measurement of workload is important to be able to evaluate the relative merits of one system over the other. As noted in the previous paragraph, however, I am doubtful about the success of assessing absolute measures of workload for the purposes of system certification.

(3) Assessing training.

In this domain, I believe that workload measurement is potentially important to determine the increase in "residual capacity" (decrease in workload) that occurs as training progresses. The goal of such measurement should be to decide when training regimes should be terminated and the learner transitioned to the operational environment, or when training of components should be combined, to initiate training of the whole.

In suggesting reasons why I believe the workload models and assessment techniques are important, I am constantly driven by the question of how these measures and models will be (and actually have been) used. Stated in other terms: How would a system be designed better when information from such a workload measure is used than when it is not? Or, rephrasing the question, how has a system been designed differently knowing about workload than not knowing about it? In this sense, I am constantly looking for, and would like to have on record, case studies that illustrate the utility of workload measures as an accurate design tool. While it is always nice to know that a workload measure confirms what a designer believed *a priori*, workload techniques will really have come of age when a workload measure is used to change a design or training system in a way that the system would not have been changed otherwise.

RELEVANT ISSUES

Relationship between Workload and Performance.

This relationship is an extremely important one and fundamental to the whole concept of workload. The range of task demands imposed by a system can be divided into two regions: a region where the demands are less than the available capacity, and a region in which the demands exceed the available capacity. In the second region workload is performance. During these overall conditions, poorer performance translates directly to greater levels of workload. However, in the first region, workload is merely the potential for performance and therefore the margin of demand increase before the breakdown occurs.

Individual Differences.

The importance of this factor in workload depends, to some extent, on what workload measures or models are being used for. Individual differences in the level of skill on a given task, which affect the relation between resources invested and performance, will change the performance resource-function. They are critically important in understanding the whole capacity-performance relationship. Individual differences may also provide a good way of getting a handle on this relationship. Individual differences in the style of using subjective measures or in the relative capacities of different operators (i.e., spatial vs. verbal) are perhaps less critically important whenever workload measures are used to design systems, and those systems are tailored for the average operator rather than being individually tailored for different kinds of operators. However, when issues such as custom-designing systems for different subgroups of the population are raised, then individual differences in workload become more relevant. In terms of designing for experts versus novices, my comments in the first sentence of this paragraph apply. That is, individual differences in the Performance-Resource Function related to skill level is clearly a relevant concept.

Relationship between Workload and Training.

This is an important relationship, but still remains to be firmly established in the extent to which workload measures, taken as a function of training, can reveal anything significantly more informative and useful than performance measures taken as a function of training. In both theory and in certain basic experiments it has been readily demonstrated that the concept of reserve capacity increasing even after performance has asymptoted. It will be important to demonstrate this phenomenon in other more complex tasks. But, even more important will be demonstrating the issues described above: How will a training paradigm be made differently knowing what workload is, rather than simply relying upon performance. If this question has a positive answer or can be shown to have a positive answer, then the values of studying the relationship between workload and training increases proportionately.

Relationship between Workload and Errors

To some extent this relationship appears to be an obvious one in the sense that a basic tenant of workload theory is that increases in workload will lead to losses in performance, and errors represent one measure of performance. In this regard it is a restatement of the Performance-Resource Function. When the causality is reversed and the question is asked: "Do errors cause increased workload?" there seems again to be a fairly intuitive answer. Errors should increase workload to the extent that the errors are either noticed or lead to degrading conditions of performance. But, they will not to the extent that errors are unnoticed or, for one reason or another, corrected by the system such that the system does not degrade. In short, I view the particular relationship between these two as a fairly atheoretical one. However, I would stress the importance of a theoretical model of errors such as that described by Reason and Norman. Important issues in this regard concern the relation between qualitatively different kinds of workload (e.g., perceptual/cognitive vs. response), and different kinds of errors (e.g., slips versus mistakes).

ORIGINAL PAGE IS
OF POOR QUALITY

RECOMMENDED MEASURES

Here again, I am going to revert to the traditional classification of measures that are subjective, physiological, or secondary task, as well as a tired old cliché that "more data are needed." It is clear to me that under conditions of relative overload primary task measures still remain the best measure of workload. However, these may well be supplemented with subjective or physiological measures (my hunch is that physiological measures such as heart rate or pupil diameter, if feasibly measured, should represent the best techniques in this situation). Under these circumstances, workload measures might influence system decisions if such factors as system cost are taken into account (i.e., the system with slightly poorer primary task performance has much lower subjective workload and is much cheaper).

During conditions where task demands are less than capacity, primary task measures are obviously insensitive and invalid and, therefore, I believe that either subjective, secondary task, or physiological measures provide realistic, plausible tools. Each, of course, has its costs and benefits. In spite of the research performed, I am still not sure the extent to which subjective measures are significantly diagnostic as to the locus or nature of task load, nor, as we have shown, do they index certain critical characteristics related to resource competition and single task demands. Secondary tasks have all the difficulties associated with "structural inference" or cost of concurrence, as well as the concerns about their obtrusiveness. However, I have argued elsewhere that obtrusiveness is not altogether a bad thing as long as priorities are appropriately stressed. I do have some concern about the use of secondary tasks for comparing workload across quite different structures or configurations. Here, differences in concurrence cost, related to the interaction of primary and secondary tasks, may introduce spurious effects into the level of secondary task performance. Finally, physiological measures, whether based upon ERPs, heart rate variability, or pupil diameter (these three still represent my best candidates) have either not yet received sufficient validation (the case of heart rate and ERPs) or are too limiting in many circumstances (such as pupil diameter). Obviously, physiological measures have a far greater cost of implementation than do many secondary tasks, and both of these are far more costly than are subjective measures. Therefore, the whole utility of using one of the three techniques depends considerably on a cost-benefit analysis.

ISSUES TO RESOLVE

Despite the tremendous amount of research in this area, I still believe there is a great deal that needs to be done. In our research program we have discovered and catalogued certain "dissociations" between workload measures with some degree of confidence. Examples are the relative insensitivity of subjective measures to the degree of resource competition, and the relative oversensitivity of these measures to the presence of concurrence cost. However, we still do not know enough about dissociations between other measures, or about other sources of dissociation with subjective measures. We also need to know far more about comparable scaling of relative measures. Most of this effort should be focused on using other measures to scale performance decrements. How do our performance decrements equate across tasks? What sort of invariance is there between changes in subjective measures and changes in performance on either primary or secondary tasks? What sort of invariant relationships hold between subjective measure changes and those on physiological measures? If workload is ever to be used scientifically to achieve more than a simple comparison of the levels of workload across a single task as one of its parameters of difficulty is varied, then we must directly confront this issue of equating scales across different tasks and different measures in terms of a single underlying construct. This particular issue brings us back to the nature of the Performance-Resource Function. Can it ever be defined as a real entity rather than an underlying hypothetical construct?

A final issue that I view as important relates to the stages-of-processing dimension in workload. There seems to be sufficient intuitive and experimental evidence that perceptual/cognitive load is different from response load. Is there any way of equating the relative loads across these two different stages? Or, is there an interaction between them in terms of total performance or in terms of any other workload measures? My view is that this is the most critical dimension for defining resources as a vector rather than a scalar quantity.

Walter W. Wierwille
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

DEFINITION

In our work we have treated workload as operator response (in a general sense) to operator loading. In other words, loading is an independent variable and workload measures then become dependent variables. Such an approach seems to avoid much of the controversy over definitions, but admittedly is too general. For example, measures of performance become measures of workload, because performance measures can be considered as response to loading. Nevertheless, I know of no way to restrict the definition without deleting known, useful measures of loading. I would like to leave the detailed definition of workload to others. However, I would caution that the definition must not be too restrictive.

IMPORTANCE

There is no doubt that workload is important. A system in which an operator works may underload, properly load, or overload that operator. Underload can cause inattention, boredom, and other vigilance-related problems. Overload can cause operator stress, error prone strategies, and outright blunders. Both underload and overload can therefore lead to dangerous situations. Ultimately, whether in the short run or the long run, performance suffers. When lives and property are at stake, such decrements in performance may lead to accidents. Workload is definitely important.

RELEVANT ISSUES

Relationship Between Workload and Performance

The previous brief statements have already addressed this relationship. Performance is one measurable aspect of workload, and if taken in a general sense, it is the ultimate measure. For example, an operator handling a very difficult task day after day may eventually have health-related problems, may burn out or develop other psychological problems, or may quit. Performance is thus ultimately effected. Thus, low or high loading may induce operator errors in the short term, in the long term, or both.

Individual Differences

It is important to recognize that specific aptitudes can vary enormously from individual to individual. Any given individual possesses various levels of a variety of aptitudes.

Systems on the other hand, must be designed so that all members of the user population can operate them safely and efficiently. For some operators, a system may be used efficiently and easily because aptitudes required by the system match those the operator possesses at high levels. Other operators, having a mismatch of aptitudes must learn to adapt or find a different kind of work.

Technology has not yet reached the point where operators' workspaces adapt to individual differences. Thus, individual differences will remain a problem in workload estimation for many years to come. Systems must be designed so that all members of the user population are accommodated to the maximum extent possible.

Researchers in the behavioral sciences have recognized that individual differences exist and have designed their research techniques to account for them. The use of statistics is a prime example. Fundamentally, statistical methods are used to determine whether or not for a given measure there are differences in population means, given only samples of the populations. Only means are compared because it is recognized that there are individual differences.

It would seem that it is more important, at present, to design systems so that workload measures for the user population are in the desired range. This statement implies that systems should be designed so that workload mean estimates are in the appropriate range. While individual differences are important, their study should be subordinated to the much more important task of getting mean workload levels set properly for the user population. The study of individual differences, like the study of dual tasks, is a never-ending process with diminishing marginal return on investment.

ORIGINAL PAGE IS
OF POOR QUALITY

Workload and Training

There is little doubt that training has a profound effect on workload. We have all been involved in activities which at first are found extremely difficult or impossible. Examples would include typing, performing long division, driving a manual transmission automobile, solo flying, playing a musical instrument, or programming a microcomputer. If asked about workload shortly after being introduced to these activities, we would probably indicate that the level was very high. However, after proper training and having performed these tasks routinely every day for several months, we would probably indicate that workload level had decreased.

Because measured workload may change as a function of learning, it would seem that experimenters should go beyond controlling for it and should actively examine it. Training and amount of practice should be treated as independent variables, just as loading is. It is to be expected that workload level will change with these new independent variables. We may find that a workload level shifts substantially with training and learning, and in ways that are surprising.

Workload and Errors

Errors are actually one class of measures of performance. Several researchers are already examining errors per se, as opposed to the relationship between workload and errors. They generally take the point of view that there are multiple causes of errors, with workload being only one of them. However, I take the point of view that errors caused by either high or low workload in systems can be very serious. If an operator misses warnings in a nuclear power plant because of underload, or if an air traffic controller blunders due to overload, the results can be catastrophic. Other categories of errors not dependent on workload may also lead to catastrophes. However, that does not relieve workload researchers from the responsibility of examining errors due to workload.

Earlier, the statement was made that performance was the ultimate measure of workload. A further statement might be that errors are the ultimate measure of performance. In any case, errors can be considered a very important aspect of workload, and should be examined along with other measures. However, as indicated earlier, errors and other changes in performance may be "late" indicators of a workload problem. Therefore, workload researchers should not direct sole attention to errors. Precursor measures of workload problems are also needed.

RECOMMENDED MEASURES

In 1983, we completed an experimental evaluation of approximately twenty-five workload estimation techniques in four aircraft simulator experiments. Each of the four flight task experiments emphasized one aspect of operator behavior (e.g., psychomotor, mediation, perceptual, or communications), but not to the total exclusion of other aspects. Using accepted methods of experimental design, we found that only a few measures were sensitive to load in each experiment. On the basis of the results, we made the following specific recommendations:

1. If the task at hand involves motor activities and manual control, use the following workload measurement techniques:

- a. Cooper-Harper scale
- b. WCI/TE scale
- c. Time estimation standard deviation

If stress (or danger) is normally associated with the task, also use:

- d. Heart rate mean

2. If the task at hand involves perceptual and mediational aspects, and if the information input is primarily visual, use the following workload measurement techniques:

- a. Modified Cooper-Harper scale
- b. WCI/TE scale
- c. Response time (to correct response)
- d. Error rate
- e. Time estimation standard deviation

3. If the task is communications oriented in nature and involves verbal input and output, use the following workload measurement techniques:

- a. Modified Cooper-Harper scale
- b. Communications errors
- c. Time estimation standard deviation

4. If the task involves a wide variety of behaviors and activities and is not easily categorized, use measures exhibiting global sensitivity, namely:

- a. Modified Cooper-Harper scale
- b. WCI/TE scale
- c. (A measure which reflects a shift in strategy or technique with load)
- d. (A measure which reflects response time, if quantifiable)
- e. Time estimation standard deviation

It should be mentioned that the WCI/TE scale is the forerunner of SWAT. It should also be mentioned that proper interpretations of the above results require careful use of experimental procedures and measure definitions. Thus, the reader interested in using any of the above recommended techniques should read the project final report in detail.

ISSUES TO BE RESOLVED

I recently wrote a chapter for a forthcoming book, Human Mental Workload. (P. Hancock and N. Meshkati, Eds.). The chapter is entitled, "Important Remaining Issues in Mental Workload Estimation." In this chapter, five important areas of further investigation are described. The chapter was written because workload research appears to be moving away from the applied and toward the esoteric. In the interest of brevity, it would probably be best to refer readers to the forthcoming book, rather than repeat the material here in abbreviated form. The titles of the five topics are as follows:

1. The importance of multiple experiments.
2. The concept of full mental load and its implications for system design.
3. Task analytic methods and momentary workload.
4. Workload estimation based on normal operating records.
5. Effects of learning and proficiency on workload.

APPENDIX C:
RESEARCH PAPERS AND PUBLICATIONS

- 1981 -

- Bird, K. A. (1981). Subjective rating scales as a workload assessment technique. *Proceedings of the 17th Annual Conference on Manual Control* (pp. 33-39). Los Angeles, CA: University of California.

The present study employs a multidimensional bipolar-adjective rating scale as a subjective measure of operator workload in the performance of a one-axis tracking task. The rating scale addressed several dimensions of workload, including cognitive, physical, and perceptual task loading as well as fatigue and stress effects. Eight subjects performed a one-axis tracking task (with six levels of difficulty) and rated these tasks on several workload dimensions. Performance measures were tracking error RMS (root-mean square) and the standard deviation of control stick output. Significant relationships were observed between these performance measures and skill required, task complexity, attention level, task difficulty, task demands, and stress level.

- Connor, S. A. (1981). *A comparison of pilot workload assessment techniques using a psychomotor task in a moving base aircraft simulator*. Unpublished M.S. Thesis. Blacksburg VA: Virginia Polytechnic Institute and State University.

A comparison of the sensitivity and intrusion of twenty pilot workload assessment techniques was conducted using a psychomotor loading task in a three degree of freedom moving base simulator. The twenty techniques included opinion measures, spare mental capacity measures, physiological measures, eye behavior measures, and primary task performance measures. The primary task was an instrument landing system (ILS) approach and landing. All measures were recorded between the outer marker and the middle marker on the approach. Three levels (low, medium, and high) of psychomotor load were obtained by the combined manipulation of windgust disturbance level and simulated aircraft pitch stability. Six instrument rated pilots participated in four sessions lasting approximately three hours each. Two opinion measures, one spare mental capacity measure, one physiological measure, and one primary task measure demonstrated sensitivity to psychomotor load in their experiment. These measures were Cooper-Harper ratings, WCI/TE ratings, time estimation standard deviation, pulse rate mean, and control movements per unit time. The Cooper-Harper ratings, WCI/TE ratings, and control movements demonstrated sensitivity to all levels of load, whereas the time estimation measure and pulse rate mean only showed sensitivity to some load levels. No intrusion was found for the physiological measures or for the spare mental capacity measures. The results of this experiment demonstrate that sensitivities of workload estimation techniques vary widely, and that only a few techniques appear to be sensitive to psychomotor load.

Hart, S. G., Childress, M. E., and Bortolussi, M. (1981). Defining the subjective experience of workload. *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 527-531). Santa Monica: Human Factors Society.

Flight scenarios that represent different types and levels of pilot workload are needed in order to conduct research about, and develop measures of, pilot workload. In order to be useful, however, the workload associated with such scenarios and the component tasks must be determined independently. An initial study designed to provide such information was conducted by asking a panel of general aviation pilots to evaluate flight-related tasks for the overall, perceptual, physical, and cognitive workload they impose. These ratings will provide the nucleus for a data base of flight-related primary tasks that have been independently rated for workload to use in workload assessment research.

Wierwille, W. W. (1981). Instantaneous mental workload: Concept and potential methods for measurement. *Proceedings of the 1981 International Conference on Cybernetics and Society* (pp. 604-608). New York: The Institute of Electrical and Electronic Engineers.

This paper provides an initial conceptual framework for instantaneous workload and describes potential methods for short-term measurement. Many existing estimation techniques can be modified for use as short-term assessment techniques. Techniques in the 1) opinion, 2) spare mental capacity, 3) primary task, and 4) physiological categories are discussed. The limitations involved in instantaneous workload, which are real and fundamental, are also described.

APPENDIX D:
RESEARCH PAPERS AND PUBLICATIONS

- 1982 -

Casali, J. G. (1982). *A sensitivity intrusion comparison of mental workload estimation techniques using a simulated flight task emphasizing perceptual piloting behaviors*. Unpublished Ph.D. Dissertation. Blacksburg, VA: Virginia Polytechnic Institute and State University.

This research represented a first attempt at examining the sensitivity and intrusion of workload estimation techniques in a flight task emphasizing perceptual load. In a similar manner, the application of these techniques to other tasks, such as mediational or communicative, warrants future investigation. Subsequently, the stability of the measures over time (reliability) needs research attention.

Casali, J. G. and Wierwille, W. W. (1982). A sensitivity intrusion comparison of mental workload estimation techniques using a flight task emphasizing perceptual piloting activities. *Proceedings of the 1982 IEEE/SMC* (pp. 598-602). Santa Monica, CA: Human Factors Society.

There are many flight task situations in which perceptual activity on the part of the pilot or aircrew member is emphasized. Unfortunately, the sensitivity, that is, the relative ability of conventional workload estimation techniques to discriminate between perceptual load levels, is largely unknown. Because of this lack of basic knowledge, an experiment comparing several workload techniques was conducted in an instrumented GAT-1B flight simulator. The initial sensitivity and intrusion results of the experiment are reported in this paper, and a relative categorization of techniques is presented, based on demonstrated sensitivity.

Childress, M. E., Hart, S. G., and Bortolussi, M. R. (1982). The reliability and validity of flight task workload ratings. *Proceedings of the Human Factors Society 26th Annual Meeting* (pp. 319-323). Santa Monica, CA: Human Factors Society.

Twelve instrument-rated general aviation pilots each flew two scenarios in a motion-base simulator. During each flight, the pilots verbally estimated their workload every three minutes. Following each flight, they again estimated workload for each flight segment and also rated their overall workload, perceived performance, and 13 specific factors on a bipolar scale. The results indicate that time (a priori, inflight, or postflight) of eliciting ratings, period to be covered by the ratings (a specific moment in time or a longer period), type of rating scale, and rating method (verbal, written, or other) may be important variables. Overall workload ratings appear to be predicted by different specific scales depending upon the situation, with activity level the best predictor. Perceived performance seems to bear little relationship to observer-rated performance when pilots rate their overall performance and an observer rates specific behaviors. Perceived workload and performance also seem unrelated.

Hart, S. G. (1982). Theoretical basis for workload assessment research at NASA Ames Research Center. *Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics* (AFFTC-TR-82-5). (pp. 455-470). Edwards AFB, CA: Air Force Flight Test Center.

Workload may be thought of as a collection of experiences, requirements, feelings, demands, and circumstances that are referred to in summary form by the term "workload." When one person says that he really worked hard, he may mean that he is physically tired, while another person may provide a rating of equivalent magnitude because he was required to do more than expected, even though his actual output and effort did not increase. There are many factors associated with the term workload as it is usually applied that each exist independently and can be analyzed as such most profitably. Task demands are just that -- task demands. No additional meaning or value can be associated with renaming this factor "workload." Physical effort and emotional stress are also independent, unique entities that can each be measured by specific and unique assessment techniques, but again neither is synonymous with "workload" per se. Performance is also an independent, important entity, but again it is not "workload." Measures of performance are most relevant to determining how successful an individual was in meeting task demands but do not reflect how hard he worked, what his expectations were, his stress level, the time pressure felt, and so on.

The one factor that does reflect the effect of all of these factors on each individual is the subjective experience of workload. If an individual feels loaded, he or she is. This may be the only factor in the constellation of elements variously call "workload" that is purely "workload" and nothing else. This subjective experience is obviously derived from the other factors -- task demands, success in meeting demands, effort, and so on -- but it is the product of a weighting process that may be unique to each individual. The weights or importance that each individual places on the various elements that may affect his experience of workload may differ from person to person, although they should be fairly consistent with an individual. By determining what factors enter into this weighting process and how they are combined, it may be possible to develop methods to assess this subjective factor -- the one element that may be uniquely "workload" -- to use in the interpretation of subjective ratings, variation in performance, and physiological recordings. The assumption is that if a person feels loaded -- he is -- and that this will not only affect his or her subjective evaluations of workload but also physiological measure of stress, arousal, fatigue, etc. and the individual's ability to perform the primary task as well as additional tasks effectively.

Hart, S. G. (1982). *Workload Assessment Research Program*. Invited address at the Air Force Office of Scientific Research in Biocybernetics and Workload Annual Review, Alexandria, VA.

The goal of this program is to develop relevant and reliable measures of pilot workload to assess and predict the impact of aircraft and ATC system changes on aircrews. Although pilots typically adjust to advances in technology, there may be unacceptable costs associated with the adjustment: pilot overload, stress or fatigue, additional training, or reduced safety. The effectiveness with which aircrews use new and existing equipment is usually defined by their performance whereas the cost to the aircrew of producing such performance is pilot workload. Measures of performance and workload may not be correlated, however, as pilots may or may not be willing or able to meet increased task demands. Further, existing measures of physical workload and overt performance may not reflect the cognitive and perceptual activities which are a major element in piloting current and future aircraft.

The term "workload" serves as a convenient label for a number of events, ideas, states, and dimensions. These factors may either relate to the operator or to the task, they may covary or not, and they may derive from the task at hand or simply coexist with it. There may be only one of these factors, however, that is uniquely "workload" and not something else: the operator's perception of his experience. If an operator feels loaded, then he is loaded and this will be reflected in physiological, subjective, and objective measures, although not necessarily in performance. This experience is derived from the other factors, but the importance placed on different components varies from person to person. Because workload measures typically reflect a fraction of the total situation and may not focus on dimensions that are relevant to that operator, available measures are often unreliable and uninformative.

Due to the complexities involved, many fundamental issues must be resolved before appropriate and reliable measures can be developed and applied: (1) Standardize the selection and combination of flight-related tasks so that predictable types and levels of primary task demand can be imposed; (2) determine the effects of many factors, such as task demands, fatigue, time pressure, effort, success, and the circumstances under which single or multiple tasks are performed on the perception of workload; (3) identify the effective level of task demand and effort as a function of the level of automatic processing and control; (4) determine the sensitivity and intrusiveness of commonly used workload measures; (5) analyze pilot errors and communications as primary task measures of workload; and (6) produce a practical guide for the analysis of workload.

Hart, S. G., Childress, M. E., and Hauser, J. R. (1982). Individual definitions of the term "workload." *Proceedings of the 1982 Psychology in the Department of Defense Symposium* (pp. 478-485). Colorado Springs, CO: United States Air Force Academy.

A study was conducted in which four groups of raters (51 researchers, 28 college students, 12 general aviation pilots, and 26 high school students) assigned 19 possible components of workload to one of three categories: (1) not related to workload; (2) related to, but not a primary component of workload; and (3) a primary element of workload. These ratings were factored to determine the relationships among the items. The analysis yielded seven factors: fatigue/stress, task difficulty, effort, performance/motivation, task type, interest in task, and purpose of task. The 117 participants were clustered on the within-subject standardized factor scores. This analysis yielded seven patterns of responses about the relative primacy of the different factors to different individuals' definitions of workload. The results indicate that patterns of estimating the primacy of components in subjective workload evaluation exist which cross working group lines.

Rahimi, M. R. (1982). *Evaluation of Workload Estimation Techniques in Simulated Piloting Tasks Emphasizing Mediation Activity*. Unpublished Ph.D. Dissertation. Blacksburg, VA: Virginia Polytechnic Institute and State University.

An experiment comparing the sensitivity and intrusion of eight workload estimation techniques was conducted using a mediational loading task in a three-degrees-of-freedom moving-base aircraft simulator. The primary task mediational loading required the pilots to solve a variety of navigational problems while maintaining straight-and-level flight. The presented problems were sorted prior to the experiment into low, medium, and high difficulty problems. The eight techniques included opinion measures (modified Cooper-

Harper rating scale and multi-descriptor rating scale), spare mental capacity measures (time estimation and tapping regularity), primary task measures (mediational reaction time and control movements per unit time), and physiological measures (pulse rate variability and pupil dilation). One opinion measure (modified Cooper-Harper rating scale), one spare mental capacity measure (time estimation), and one primary task measure (mediational reaction time) demonstrated sensitivity. These results suggest that sensitivity and intrusion of workload estimation techniques vary widely when applied to mediational task, and that care must be taken to select sensitive measures. It must not be assumed that all measures are equally sensitive.

Rahimi, M. and Wierwille, W. W. (1982). Evaluation of the sensitivity and intrusion of workload estimation techniques in piloting tasks emphasizing mediational activity. *Proceedings of the 1982 IEEE/SMC* (pp.593-597). Santa Monica, CA: Human Factors Society.

In this experiment, pilots flew an instrumented moving-base simulator. Mediational loading was elicited by having them solve a variety of navigational problems. The problems were sorted into low, medium, and high load conditions based on the number and complexity of arithmetic and geometric operations required to solve them. Workload estimation techniques based on opinion, spare mental capacity, primary task performance, and physiological measures were obtained and compared. This paper describes: (1) the ability of the techniques to discriminate statistically between the three levels of loading conditions, and (2) changes in primary task performance caused by introduction of the workload technique procedures and equipment.

Schneider, W., Vidulich, M. A., and Yeh, Y. -Y. (1982). Training spatial skills for air-traffic control. In R. E. Edwards (Ed.), *Proceedings of the Human Factors Society 26th Annual Meeting* (pp. 10-14). Santa Monica, CA: Human Factors Society, Inc.

Guidelines for microprocessor based skill trainers are presented. A training program for air traffic control (ATC) of rendezvous for inflight refueling is described. The program seeks to optimize practice for developing automatic component skills. The program sequences the trainee through 10 stages to develop spatial skills for ATC. The resulting training program can develop fast, accurate, and reliable performance on the individual components with only a few hours' training per component. The proposed approach is contrasted with current training methods. The general applicability of the guidelines to microprocessor based skill trainers is described.

Vidulich, M. A. and Wickens, C. D. (1982). The influence of S-C-R compatibility and response competition on performance of threat-evaluation and fault diagnosis. In R. E. Edwards (Ed.), *Proceedings of the Human Factors Society 26th Annual Meeting* (pp. 223-226). Santa Monica, CA: Human Factors Society, Inc.

Stimulus/central-processing/response compatibility defines the optimum assignment of tasks to input modalities (auditory, A and visual, V) and output modalities (manual, M and speech, S). Spatial tasks are S-C-R compatible with visual/manual assignments. Verbal tasks are compatible with auditory/speech assignments. Ten subjects time-shared a spatial task of aerial threat evaluation with a verbal task of fault diagnosis. All four i/o

modality combinations of the threat task were performed while the fault task was performed with A/M and V/M assignments. The joint effects of compatibility, and competition between tasks for input and output modalities were demonstrated. When resource competition was held constant, the effects of compatibility were found to be enhanced in dual task conditions. When both influences varied they were demonstrated to counteract in certain conditions and balance each other's effect.

Wickens, C. D. and Vidulich, M. A. (1982). *S-C-R Compatibility and Dual Task Performance in Two Complex Information Processing Tasks: Threat Evaluation and Fault Diagnosis* (Tech. Rep. No. EPL-82-3/ONR-82-3). Champaign: University of Illinois, Engineering-Psychology Research Laboratory.

This experiment was conducted to extend the principles of stimulus/central-processing/response or S-C-R compatibility, described in an earlier report by Sandry and Wickens, to a more complex environment. The principle states that tasks with verbal central-processing demands will be best served by voice input and output channels. Tasks with spatial demands will be best served by visual/manual channels. A verbal task requiring subjects to evaluate the relative velocity vector of two aircraft for the likelihood of interception. In different conditions each of these were served by both input and output modalities, in single and dual task configurations.

The general results indicated that anticipated compatibility effects were obtained and often enhanced under dual task conditions. In particular, in some circumstances compatibility effects dominated those of resource competition. That is, performance on both tasks in a dual task pair was better when they shared different channels, but one was incompatibly displayed. The practical implications of these results to the interfacing of tasks with voice recognition and synthesis technology are discussed.

Wickens, C. D. and Yeh, Y. -Y. (1982). The dissociation of subjective ratings and performance. *Proceedings of the 1982 IEEE/SMC Meeting* (pp. 584-587). Santa Monica, CA: Human Factors Society.

This investigation provides three demonstrations of the manner in which subjective measures of task workload and performance dissociate. (1) The number of tasks performed concurrently influences subjective measures more than performance. (2) The extent to which tasks demand common resources influences performance relatively more than subjective measures. (3) The control order of single axis tracking influences subjective measures relatively more than performance, in contrast to the bandwidth of a single axis task. These results suggest caution in the interpretation of subjective measures as a ubiquitous measure of task difficulty.

Wierwille, W. W. (1982). Determination of sensitive measures of pilot workload as a function of the type of piloting task. *Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics (AFFTC-TR-82-5)*. (pp. 471-490). Edwards AFB, CA: Air Force Flight Test Center.

The purpose of our present work, sponsored by NASA Ames, is to examine the sensitivity, intrusion, and transferability of a variety of workload assessment techniques. The study

will use four different simulated piloting tasks, emphasizing psychomotor, perceptual, mediational, and communications aspects. Pilot loading levels will be systematically adjusted. Our simulation facility is a GAT-1B that has been modified and instrumented for workload estimation techniques measurement. The flight simulator itself has three degrees of physical motion and a full complement of IFR instruments.

Recently we completed the experiment emphasizing the psychomotor aspect of flight. Instrument-rated pilots flew instrument approaches under three combined settings of the independent variable: increasing turbulence and decreasing longitudinal stability. Twenty different workload measures were taken between the outer and middle markers, only five of which showed statistically reliable changes as a function of the independent variable. Included in the five were: two rating scales, one measure of control movement activity, pulse rate, and one measure of time estimation. The results of the experiment are to some extent surprising, for they indicate that several "accepted" measures of workload are not reliably sensitive to the kinds of psychomotor load which pilots encounter.

APPENDIX E:
RESEARCH PAPERS AND PUBLICATIONS

- 1983 -

- Acton, W. H., Crabtree, M., Simons, J. C., Gomer, F. E. and Eckel, J. S. (1983). Quantification of crew workload imposed by communications-related tasks in commercial transport aircraft. *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 239-243). Santa Monica, CA: Human Factors Society.

Information theoretical analysis and subjective paired-comparison and task ranking techniques were employed in order to scale the workload of 20 communications-related tasks frequently performed by the captain and first officer of transport category aircraft. Tasks were drawn from taped conversations between aircraft and air traffic controllers (ATC). Twenty crewmembers performed subjective message comparisons and task rankings on the basis of workload. Information theoretic results indicated a broad range of task difficulty levels, and substantial differences between captain and first officer workload levels. Preliminary subjective data tended to corroborate these results. A hybrid scale reflecting the results of both the analytical and the subjective techniques is currently being developed. The findings will be used to select representative sets of communications for use in high fidelity simulation.

- Casali, J. G. and Wierwille, W. W. (1983). A comparison of rating scale, secondary-task, physiological, and primary task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors*, 25 (6), 623-642.

Sixteen potential metrics of pilot mental workload were investigated regarding their sensitivity to communication load and their intrusion on primary-task performance. A moving-base flight simulator was used to present three cross-country flights. The flights varied only in the difficulty of the communications requirements. Rating scale measures were obtained immediately postflight; all others were taken over a 7-min. segment of the flight task. The results indicated that both the Modified Cooper-Harper Scale and the workload Multi-Descriptor Scale were sensitive to changes in communications load. The secondary-task measure of time estimation and the physiological measure of pupil diameter were also sensitive. As expected, those primary-task measures that were direct measures of communicative performance were also sensitive to load, whereas aircraft control primary-task measures were not, attesting to the task specificity of such measures. Finally, the intrusion analysis revealed no differential interference between workload measures.

- Casali, J. G. and Wierwille, W. W. (1983). Communications-imposed pilot workload: A comparison of sixteen estimation techniques. *Proceedings of the Second Symposium on Aviation Psychology* (pp. 223-234). Columbus: Ohio State University.

Sixteen potential metrics of mental workload were investigated in regard to their relative sensitivity to communications load and their differential intrusion on primary task performance. A moving-base flight simulator was used to present three cross-country flights to each of 30 subject pilots, each flight varying only in the difficulty of the inherent

communications obtained immediately post-flight, all measures were taken over a seven minute segment of the flight task. The results indicated that both the Modified Cooper-Harper and the workload Multi-Descriptor rating scales were reliably sensitive to changes in communications load. Also, the secondary task measure of time estimation and the physiological measure of pupil diameter yielded sensitivity. As expected, those primary task measures which were direct measures of communicative performance were also sensitive to load, while aircraft control primary task measures were not, attesting to the task-specificity of such measures. Finally, the intrusion analysis revealed no differential interference between workload measures.

Childress, M. E. (1983). Subjective scales for workload evaluation: Critical aspects and new directions for research. *Proceedings of the 19th Annual Conference on Manual Control* (pp. 1-2). Cambridge: Massachusetts Institute of Technology.

As aircraft and other mechanical systems increase in complexity and rely more heavily on computerization of function, and as the pilot or other operator assumes greater supervisory responsibility for system monitoring and control, need for evaluation of the workload associated with system changes increases. Many of the methods currently available, though helpful in specific situations and often necessary in promoting understanding of some basic processes, are often difficult and unwieldy to use in complex, practical situations. Subjective rating scales, however, are convenient instruments for evaluating this workload and for estimating the magnitude of changes in load as system changes occur. The use of such scales has historical precedence in the personnel literature, particularly in performance evaluation. Subjective scales also have been used to evaluate specific system characteristics, such as aircraft handling qualities. The utility of the method is clear; however, psychometric development of subjective scales for the evaluation of workload currently is in its infancy. Thus, though the literature is replete with examples of and recommendations for their use as well as with criticisms of their deficiencies, research directed towards examination of their properties, and evaluation of the conditions under which their use is appropriate and obtained results generalizable is just beginning. Several important works (e. g., Landy and Farr, 1980; Moray, 1979; Nisbett and Wilson, 1977; Wherry, 1950, 1952) have described the problems associated with subjective ratings, have detailed some of the situations in which they may be appropriate, and have recommended specific topics for future research. This paper presents a review of critical aspects of that literature which suggest directions for future research relative to self-ratings of subjective workload. It provides examples of some recent work at Ames Research Center which has suggested extending the basic input-processing-outcome model for examining workload to consider all input sources and the related outcomes, and it details current work based on that model.

Connor, S. A. and Wierwille, W. W. (1983). *Comparative Evaluation of Twenty Pilot Workload Assessment Measures Using a Psychomotor Task in a Moving-base Aircraft Simulator* (NASA CR-166457). Washington, DC: National Aeronautics and Space Administration.

A comparison of the sensitivity and intrusion of twenty pilot workload assessment techniques was conducted using a psychomotor loading task in a three degree of freedom moving-base aircraft simulator. The twenty techniques included opinion measures, and primary task performance measures. The primary task was an instrument landing system (ILS) approach and landing. All measures were recorded between the outer marker and

the middle marker on the approach. Three levels (low, medium, and high) of psychomotor load were obtained by the combined manipulation of windgust disturbance level and simulated aircraft pitch stability. Six instrument rated pilots participated in four sessions lasting approximately three hours each.

Eckel, J. S. and Crabtree, M. S. (1983). Analytic and subjective assessments of operator workload imposed by communications tasks in transport aircraft. *Proceedings of the Second Symposium on Aviation Psychology* (pp. 237-242). Columbus: Ohio State University.

The purpose of this project is to use analytical and subjective techniques to estimate the workload imposed on the aircrew by typical communications-related tasks performed during selected flight phases. Communications-related tasks are defined operationally to consist of sequences of verbal and discrete manual responses which are initiated when the crew receives and interprets radio techniques will be used to quantify communications-related workload. The first, an information theoretic technique, permits determination of bit values for perceptual and for verbal and manual action components of each task. The second is a paired comparison technique to obtain subjective estimates of the cognitive processing demands of individual communication requests. By combining the results of the paired comparison analysis with the results of the information theoretic analysis, we will derive a single hybrid scale of communications-related workload. The third technique relies on pilots' estimations of the overall workload associated with communications tasks. Recommendations for future research include an examination of communications-induced workload among the air crew and the development of simulation scenarios which impose distinctly different levels of communications-related workload. This work was performed under Contract NAS2-11562 for the National Aeronautics and Space Administration, Ames Research Center, Moffett Field, California.

Goguen, J. A. and Linde, C. (1983). *Linguistic Methodology for the Analysis of Aviation Accidents* (NASA CR-3741). Washington, DC: National Aeronautics and Space Administration.

This research develops a linguistic methodology for the analysis of small group discourse, and demonstrates the use of this methodology on transcripts of commercial air transport accidents. The methodology first identifies the discourse types that occur (these include planning, explanation, and command and control) and determines their linguistic structure; it then identifies significant linguistic variables based upon these structures or other linguistic concepts such as speech act and topic; next, it tests hypotheses that support the significance and reliability of these variables; and finally, it indicates the implications of the validated hypotheses. These implications fall into three categories: (1) training crews to use more nearly optimal communication patterns; (2) using linguistic variables as indices for aspects of crew performance such as attention; and (3) providing guidelines for the design of aviation procedures and equipment, especially those that involve speech.

Gopher, D. and Braune, R. (1983). On the psychophysics of workload: Why bother with with subjective measures? *Proceedings of the Second Symposium on Aviation Psychology* (pp. 253-268). Columbus: Ohio State University.

Psychophysical functions describe the relationship between variations in the amplitude of a defined physical quantity and the psychological perception of these changes. Examples are brightness, loudness, and pain. The regularities of these relationships have been recognized since the early days of experimental psychology, and have been formulated into psychophysical laws. The Harvard group, led by S. S. Stevens, proposed a power function as a general form for such laws. The main argument of the present paper is that a similar scaling approach can be adapted to the measurement of workload and task demands based upon subjective estimates given by subjects. The rationale is that these estimates, like other psychophysical judgments, express the individual's perception of the demands imposed on him by the surrounding environment. This approach was successfully applied to the assessment of 21 experimental conditions given to a group of 60 subjects. The paper discusses the main results of this effort and their implication to theory and application in human performance.

Hart, S. G. (1983). Effect of VFR aircraft on approach traffic with and without cockpit displays of traffic information. *Proceedings of the 18th Annual Conference on Manual Control (AFWAL-TR-83-3021)*. (pp. 522-544). Wright-Patterson AFB, OH: Air Force Wright Aeronautical Laboratories.

This study investigated the impact of cockpit displays of traffic information (CDTI) on the flow of approach traffic. A mix of aircraft type, CDTI- equipage, and type of air traffic control (ATC) were included in the simulation. In addition, the practical issue of simulator fidelity in conducting such experiments was studied. Seven piloted simulators that represented a mix of general aviation-type and transport-type aircraft were simulated with two levels of control fidelity. They were flown by four teams of seven pilots each under ATC. A computer-generated target flying a predetermined flight path was also included to represent aircraft not in contact with ATC. The results indicate that aircraft type and not simulator fidelity influenced pilot and system performance. The frequency and content of communications, several measures of system performance, and pilot ratings also reflected pilot willingness to accept closer spacing and clearances to follow aircraft seen on a CDTI.

Hart, S. G. and Bortolussi, M. R. (1983). Pilot errors as a source of workload. *Proceedings of the Second Symposium on Aviation Psychology* (pp. 269-278). Columbus: Ohio State University.

A pilot opinion survey was conducted to develop a data base for creating simulation scenarios that impose predetermined levels of pilot workload. Twelve pilots estimated the effect of 163 different events and activities on their performance, effort, workload and stress. The events included routine control, navigation and communications activities, aircraft and system failures, and pilot errors. Predicted changes in workload, stress, and effort were significantly correlated with each other but not with performance. When events were coupled with high workload flight segments, the predicted impact on workload, stress, and performance was proportionally greater than it was for less demanding segments. Effort ratings did not vary with flight phase. Workload ratings were highest for weather-related events, systems failures, and approach- and departure-related problems and lowest for routine activities, although there was considerable range within each category. Errors were found to be a significant source of pilot workload, stress, and performance decrements, suggesting that errors should be conceptualized as a *cause* of workload rather than as a *symptom*.

Hart, S. G. and Chappell, S. L. (1983). Influence of pilot workload and traffic information on pilot's situation awareness. *Proceedings of the 19th Annual Conference on Manual Control* (pp. 4-26). Cambridge: Massachusetts Institute of Technology.

Although it seems intuitively obvious that the addition of a cockpit display of traffic information (CDTI) should enhance a pilot's awareness of the current and projected situation of other aircraft, it has not been empirically determined that such is the case. Furthermore, there is some question about the utility of CDTI under conditions of relatively high pilot workload; when pilots become busy they may ignore the CDTI or take unilateral actions based on incompletely understood information. The current simulation was designed to determine how much information pilots could recall about eight aircraft simultaneously participating in a simulated approach task. A stop-action technique was used, so that each approach sequence was terminated at some point and the participants completed a written debriefing describing their recall of aircrafts' positions, situations, and intentions. The experimental variables included: (1) presence or absence of CDTI; (2) CDTI quality; and (3) level of concurrent workload. Four groups each consisting of three transport pilots, four instrument-rated general aviation pilots and one controller participated in the experiment. Concurrent workload but not CDTI quality or presence significantly affected the type and amount of information remembered. Rated workload and several types of communications were increased by the addition of CDTI and by the experimental manipulations intended to increase workload. The pilots reported feeling that CDTI afforded them a better understanding of the traffic situation, but this subjective impression was not supported by an improvement in the amount of information recalled.

Hart, S. G. and Chappell, S. (1983). Pilot communications as a source and indicator of workload. Paper presented at the Meeting of the IEEE/SMC. Santa Monica, CA: Human Factors Society.

A simulation was conducted with four groups of seven pilots to evaluate subjective ratings and communications as measures of pilot workload. Each group consisted of three airline and four general aviation pilots flying simulated transport and general-aviation aircraft and an air traffic controller. Six approaches were flown under either low- or high-workload levels. Workload was manipulated by introducing ATC-system and aircraft problems, and by imposing additional tasks on the pilots. For half of the High and Low-Workload conditions, a visual display of the traffic situation was superimposed on the primary flight instrument. Workload was assessed at 2-min intervals inflight with a 10-point rating scale (POSWAT) and with 19 rating scales post-flight. The types and frequencies of communications were tabulated from transcripts. Both inflight and post-flight ratings increased significantly between the High and Low Workload conditions and from the beginning to the end of each approach. In addition, the presence of a cockpit display of traffic information also contributed to an increase in subjective workload. The primary types of communications were clearances, reports, readbacks, and acknowledgements. The frequency of traffic advisory and holding instruction communications differed significantly between experimental conditions; both occurred more often in the High Workload conditions. Communications rate increased significantly between Low and High Workload conditions when pilots did not have visual traffic situation displays. However, fewer communications occurred in the High Workload conditions when pilots did have traffic situation displays. This suggests that traffic situation displays can reduce the need for verbal communications during demanding phases of flight, although this can result in higher pilot workload (as they increase information-processing demands).

Hartzell, E. J., Gopher, D., Hart, S., Lee, E., and Dunbar, S. (1983). The Fittsberg Law: The joint impact of memory load and movement difficulty. Paper presented at the Meeting of the IEEE/SMCS. Santa Monica, CA: Human Factors Society.

In a typical dual-task paradigm, two different tasks are performed within the same time period (thereby competing for an operator's limited resources), yet the component tasks are unrelated either functionally or subjectively. An alternative paradigm would be one in which component tasks are functionally related; the output or response to one initiates or provides information for the other. This type of task is common in operational environments where the decision to initiate a change in system state requires preliminary information gathering, processing, and decision making followed by a control action. The source of information, processing requirements, response modality, and workload levels of the first stage are independent of those of the second stage. Nevertheless, the two tasks are functionally related and some processing stages may be performed in parallel or the activities required for one may simultaneously satisfy some of the requirements of the other. A task was designed for the current study that combined a target acquisition task based on FITTS Law with a SternBERG memory search task ("FITTSBERG"). Two identical targets are displayed equi-distant from a centered probe stimulus. Subject acquired the target on the right if the probe was a member of the memory set and the target on the left if it was not. It was found that reaction time, but not movement time, increased as the difficulty of the memory search task was increased. Movement time, but not reaction time, increased as the difficulty of the target acquisition task was increased. Subjects rated the workload of the combined "FITTSBERG" task as slightly greater than the workload of the response selection task by itself. In comparison to the traditional dual-task paradigms, performance decrements for the response selection or response execution components were not found as the difficulty of the other component was increased. Rather, the two components appeared to impose relatively independent (or at least parallel) demands that did not interfere with each other's performance, although the response to the first task component simultaneously initiated the second task component, and the combined task was performed with less workload than would be predicted from the sum of the single task levels.

Hauser, J. R., Childress, M. E. and Hart, S. G. (1983). Rating consistency and component salience in subjective workload estimation. *Proceedings of the 18th Annual Conference on Manual Control* (AFWAL-TR-83-3021). (pp. 127-149). Wright-Patterson AFB, OH: Air Force Wright Aeronautical Laboratories.

Twelve general aviation pilots participated in a two-day experiment performing four tasks intended to load on different cognitive, perceptual, and motor dimensions. The tasks were varied in apparent difficulty level so that each pilot performed a total of sixteen tasks counter-balanced for task and level. Subjective ratings of factors contributing to workload were made immediately following each level of each task using a 15 bipolar adjective scale. Results indicated that the subjective perception of workload was not related to actual performance measures; however, the subjective ratings were generally consistent with the demands made by the levels of each task. Although only two of the rating scale items, own Performance and Task Difficulty, demonstrated significant within-task differences for all four tasks, the majority of rating scales showed within-task differences for those tasks that imposed higher cognitive demands. Strong relationships were found between Overall Workload, Stress Level, and Task Difficulty ratings on all tasks.

Hauser, J. R. and Hart, S. G. (1983). The effect of feedback on subjective and objective measures of workload and performance. *Proceedings of the Human Factors Society 27th Annual Meeting* (p. 144). Santa Monica, CA: Human Factors Society.

Thirty subjects were employed in a mixed experimental design that examined five levels of feedback and two levels of difficulty for two tasks, with repeated measures on the difficulty and task variables. The amount and type of feedback was varied so that it provided information about performance of the task on the objective measures within a block of trials, or provided the same information at the end of the task (simply providing knowledge of results), and was also varied in quality, either as a comparison to the subject's own average performance of the task, or in comparison to an experimentally determined 'figure of merit.' Two tasks, each with two levels of difficulty, were used: (1) a task that primarily imposed cognitive demands, a version of the Sternberg memory task, and (2) a task that primarily imposed psychomotor demands, a target acquisition task modeled on the Fitts' Law paradigm. Both objective and subjective measures demonstrated reliable and predictable effects for the difficulty levels of the two tasks, however the tasks were differentially affected by the feedback conditions, but differences between and within tasks were generally small. The relationships between objective measures, and subjective ratings of workload and performance rarely reached significant levels.

Hauser, J. R. and Hart, S. G. (1983). Subjective workload experienced during pursuit tracking as a function of available information. *Proceedings of the 19th Annual Conference on Manual Control* (p. 3). Cambridge: Massachusetts Institute of Technology.

Twelve general aviation pilots performed a pursuit tracking task where the objective was to "pilot" the pursuit vehicle (a simplified delta wing aircraft) after a target (represented by a cross). Successful acquisition of the target was always displayed on the screen. Each subject experienced twenty 10-min experimental runs in a partially counterbalanced order. The experimental variables included: (1) number of dimensions (2 or 3), (2) target path complexity (low or high), and (3) availability of information (both target and pursuit vehicle were displayed for either 100%, 50%, 25% of the time, or at subject command). Subjects controlled the vehicle by pressing rocker arm switches for the functions of yaw (left and right), roll (left and right), speed (acceleration and deceleration), pitch (up and down), and screen illumination (for the one condition where subject control was given to display time). After every experimental trial, subjects rated the preceding experience using a set of 15 bipolar adjective scales. Significant differences were found for their difficulty and display time variables on the majority of the scales, and for the dimension variable on eight of the scales. Strong relationships were found between actual time-on-target with ratings of overall workload, performance, and difficulty for almost all subjects, who were also able to estimate their time-on-target with a high degree of accuracy.

Kantowitz, B. H., Hart, S. G. and Bortolussi, M. R. (1983). Measuring pilot workload in a moving-base simulator: I. Asynchronous secondary choice-reaction task. *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 319-322). Santa Monica, CA: Human Factors Society.

The de facto method for measuring airplane pilot workload is based upon subjective ratings. While researchers agree that such subjective data should be bolstered by using objective behavioral measures, results to date have been mixed. No clear objective technique has surfaced as the metric of choice. We believe this difficulty is in part due to

neglect of theoretical work in psychology that predicts some of the difficulties that are inherent in a futile search for the one and only best secondary task to measure workload. An initial study that used both subjective ratings and an asynchronous choice-reaction secondary task was conducted to determine if such a secondary task could indeed meet the methodological constraints imposed by current theories of attention. Two variants of a flight scenario were combined with two levels of the secondary task. Appropriate single-task control conditions were also included. Results give grounds for cautious optimism but indicate that future research should use synchronous secondary tasks where possible.

Kreifeldt, J. G. (1983). *Simulation Fidelity and Numerosity Effects in CDTI Experimentation* (NASA-CR-166459). Washington, DC: National Aeronautics and Space Administration.

A comparison of twenty pilot workload assessment techniques was performed using a simulated flying task in which three levels of psychomotor workload were imposed. The experiment was conducted in a three-degree of freedom motion-base simulator. The twenty techniques evaluated included opinion measures, spare mental capacity measures, physiological measures, eye movement behavior and primary task performance measures. The primary task was an instrument landing system (ILS) approach and landing. All measures were recorded between the outer and middle markers on the approach. Three levels of psychomotor load were obtained by the combined manipulation of wind gust disturbance level and simulated aircraft pitch stability. Six instrument-rated general aviation pilots participated in the experiment.

Two opinion measures, one spare mental capacity measure, one physiological measure, and one primary task measure demonstrated sensitivity to psychomotor load in this experiment. These measures were: Cooper-Harper ratings, WCI/TE ratings, time estimation standard deviation, pulse rate mean, and control movements per unit time. No intrusion into primary task performance was found for the physiological spare mental capacity measures. The results of this experiment demonstrate that the sensitivities of workload estimation techniques vary widely, and that only a few techniques appear to be sensitive to psychomotor load.

Madni, A. (1983). Integrated modeling approaches in advanced cockpit automation. *Proceedings of the Second Aerospace Behavioral Engineering Technology Conference* (SAE Technical Paper Series No. 831543). Warrendale, PA: Society of Automotive Engineers.

With advances in display and control methods and recent developments in sensor and microelectronic technologies, the term automation, especially as it pertains to the cockpit of a tactical aircraft, has taken on a totally new dimension. No longer are we restricted to automation as it pertains to solely flight management functions. Functions such as realtime situation assessment, tactics selection and trajectory control are all candidates for partial or total automation. In addition, adaptive vehicle subsystem reconfigurations as a function of tactical posture, onboard faults and ongoing emergencies are all within the preview of onboard automation. In order to realize these rather ambitious goals it is suggested that no one class of models is adequate in providing the necessary onboard intelligence to enhance overall performance and reduce workload. Rather, a multi-model integrated approach that relies on a compendium of models from such diverse fields as artificial intelligence (AI) and expert systems, decision analysis, control theory and simulation is suggested as a basis for introducing onboard automation. This approach

relies on the selective use of one or more of these models depending on the specific tactical function and mission requirement being addressed at the time.

- Madni, A. M. and Lyman, J. (1983). Model-based estimation and prediction of task-imposed mental workload. *Proceedings of the 1983 IEEE Systems, Man, and Cybernetics Symposium* (pp. 314-317). Santa Monica, CA: Human Factors Society.

Mental workload has been an area of intensive research for better than a decade. One specific area of interest in aircrew related workload is operational terms. The suggested modeling framework is based on an interpreted Petri net characterization of a task in which "places" are equated to specific task-related activities and "transitions" are viewed as internal or external forcing events. It is shown that within this framework quantitative assessments can be made of both cumulative and instantaneous workload associated with the performance of a task and its individual component subtasks. It is suggested that insights gained from analyzing task-specific workload within this modeling paradigm can suggest plausible explanations for reconciling discrepancies between subjectively elicited workload estimates and behavioral/performance measures.

- McDonald, G. (1983). *Multi-flight Simulator System* (NASA-CR-166449). Washington, DC: National Aeronautics and Space Administration.

A prototype Air Traffic Control facility and multi-man flight simulator facility was designed and one of the component simulators fabricated as a proof of concept. The facility was designed to provide a number of independent simple simulator cabs that would have the capability of some local, stand-alone processing that would in turn interface with a larger host computer. The system was designed to accommodate up to eight flight simulators (commercially available instrument trainers) which could be operated stand-alone if no graphics were required or could operate in common simulated airspace if connected to the host computer. A proposed addition to the original design is the additional capability of inputting pilot inputs and quantities displayed on the flight and navigation instruments to the microcomputer when the simulator operates in the stand-alone mode to allow independent use of these commercially available instrument trainers for research. This document describes the conceptual design of the system and progress made to date on its implementation.

- Rieger, C. A. (1983). *Analysis of Decision Tree Rating Techniques for the Assessment of Pilot Mental Workload in a Simulated Flight Task Emphasizing Mediatlional Behavior*. Unpublished M.S. Thesis. Blacksburg: Virginia Polytechnic Institute and State University.

The purpose of this study was to improve the sensitivity of the Modified Cooper-Harper (MCH) Scale and to try to identify what aspects of the scale contribute to its effectiveness. A simulated flight task emphasizing mediational (cognitive) behavior was used to present low, medium, and high levels of loading to 6 student and thirty licensed pilots. In a Singer-Link GAT-1B flight simulator, the pilots performed three counterbalanced load level flights. After each simulated flight, a rating scale and questionnaire was administered. The results indicated that the paper rating scale having 15 response alternatives and the original decision tree was the most sensitive to load. Both 10-point modifications, the

computerized version of the MCH Scale and the version with the decision tree format removed, were somewhat superior to the original MCH Scale, which was also sensitive to load. These findings, however, are not consistent with those obtained in a companion study of communications tasks, indicating that these rating scale measures are task dependent. Use of the MCH Scale is recommended since it alone has consistently demonstrated sensitivity to load across tasks and across studies.

Skipper, J. H. (1983, August). *The Effects of Modification of a Decision Tree Rating Scale Used for Mental Workload Estimation in a Communications Task*. Unpublished M.S. Thesis. Blacksburg: Virginia Polytechnic Institute and State University.

Six rating scale designs emphasizing major characteristics which might cause the MCH scale to be a sensitive measure of mental workload were used in this study. The aims of the research were to discover what modifications of the MCH might make it even more sensitive.

A communications task developed by Casali and Wierwille (1983) was manipulated to present 36 subject pilots, both private and student, with three communications loading levels. The pilots were distributed into the six rating scales by experience level. Six different experience levels were represented in each of the rating scale groupings. Using the communications loading, the presence of a decision tree in the scales appeared to improve the scale's ability to discriminate among loading levels. The expansion of the MCH scale to 15 categories decreased the sensitivity of the MCH rating scale. The standard 10-point MCH rating scale was the most consistent of the six rating scales and attained a high ability to discriminate among loading levels.

Tanaka, K., Buharali, A. and Sheridan, T. B. (1983). Mental workload in supervisory control of automated aircraft. *Proceedings of the 19th Annual Conference on Manual Control* (pp. 40-58). Cambridge: Massachusetts Institute of Technology.

The purpose of this study is to investigate the nature of pilot mental workload in highly automated aircraft. On the basis of Rasmussen's model where human behavior is divided as skill, rule and knowledge-based, we hypothesize that mental workload is multi-dimensional, and that different aspects of workload are associated with each level of human behavior. In order to examine these hypotheses, a laboratory flight simulator was developed, functions of which included dynamics of a general aviation aircraft, autopilots, and navigational aids, as well as artificial air traffic controllers. Terminal-area approaches were simulated based on several scenarios where pilot tasks included aircraft guidance, navigation, aircraft configuration changes, and communication with air traffic control. In each case workload was measured by employing subjective rating scales and the number of pilot actions. It was shown that the level of automation available affects only the workload of skill-based behavior, whereas the abnormality of the situation resulted in an increase in workload for rule- and knowledge-based behavior.

Vidulich, M. A. and Wickens, C. D. (1983). *Processing Phenomena and the Dissociation between Subjective and Objective Workload Measures* (Tech. Rep. No. EPL-83-2/ONR-83-2). Champaign: University of Illinois, Engineering-Psychology Research Laboratory.

Causes of dissociation between subjective workload assessments and objective performance were investigated. A Sternberg memory search task was utilized. Sternberg task configurations varied in the automaticity of performance, stimulus presentation rate, discernability of stimuli, and the value of good performance. Automaticity in Sternberg task performance was manipulated by using two independent sets of stimuli, one of which was consistently mapped (i.e., targets were always the same) while the other was inconsistently mapped (i.e., targets changed over trials). Also, all Sternberg configurations were performed both as single tasks and as part of dual-task combinations (with a manual control task). During testing subjects rated all trials on eight typical bipolar rating scales.

Analysis of the results detected three major differences (i. e. dissociations) between what the ratings of workload would predict and, the actual performance which occurred. Subjects' ratings: (1) did not reflect the dual-task advantage of the consistently mapped Sternberg, (2) predicted an advantage for the slower presentation rate in which performance was degraded, and (3) indicated a higher level of workload was associated with the performance gain in a bonus-available condition. All of these dissociations identified could potentially contaminate subjective assessments in the field. The results were interpreted as supporting cognitive-processing-based experimentation in subjective workload assessment aimed at identifying differences between the cognitive processing accounting for subjective assessments and those processes that produce performance.

Vidulich, M. A., Yeh, Y. -Y., and Schneider, W. (1983). Time-compressed components for air-intercept control skills. *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 161-164). Santa Monica, CA: Human Factors Society.

The study tested guidelines for the use of microprocessors in training spatial skills for air traffic control. The central issue was the use of time-compressed simulation to aid the development of skill in identifying turn points and rollout headings for aircraft. Two groups of subjects were used. One group trained with a real-time simulation of the task, while the second group trained with a time-compressed version of the task running about 20 times as fast as real-time trials. Both groups were then tested in real-time trials. The results indicate that time compression can be a useful technique for increasing the efficiency of training.

Wickens, C. D., Sandry, D. L., and Vidulich, M. A. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25, 227-248.

Synthesized auditory displays and speech recognizers were used in two experiments to develop guidelines for their implementation in military aircraft. In the first experiment, the competition between encoding and response modalities of concurrent tasks was examined. The memory-search task was more susceptible to competition for visual encoding, whereas the tracking task bore the greater impact from shared manual responding. The second experiment examined competition between tasks for encoding and response modalities and the optimum assignment of modalities to a given task. A simulated flight task was performed concurrently with either a spatial task (target

acquisition) or a verbal task (memory). Best performance and least interference with the flight task were obtained when the spatial task was displayed visually and responded to manually, and also when the verbal task was displayed auditorily and responded to with speech.

Wickens, C. D., Vidulich, M. A., and Sandry-Garza, D. (1983). Principles of S-C-R compatibility with spatial and verbal tasks. *Proceedings of the Second Symposium on Aviation Psychology* (pp. 299-306). Columbus: Ohio State University.

A pilot's tasks may be categorized into those that demand predominantly verbal operations and those that are spatial. We describe two experiments that define two principles of compatibility of interfacing such tasks with displays and controls. The first defines compatibility according to display-location and response hand; the second according to the modality of display (auditory and visual) and response (manual and speech). In both experiments, these principles of compatibility are confirmed under dual task conditions. We describe their implications for cockpit design.

Wickens, C. D. and Yeh, Y. -Y. (1983). The dissociation between subjective workload and performance: A multiple resources approach. *Proceedings of the Human Factors Society 27th Annual Meeting* (pp.244-247). Santa Monica, CA: Human Factors Society.

A theory of the dissociation between subjective measures of mental workload and performance is described. The theory proposes that subjective measures are heavily driven by the *number* of tasks or task elements that a subject must perform concurrently. However they are relatively less sensitive to whether these tasks compete for common or separate resources, and to the difficulty of a single task, particularly if this difficulty is related to response factors. Performance, on the other hand, is particularly influenced by single task difficulty of both a perceptual and response nature and by resource competition between tasks. A set of three experiments are described to examine the dissociation between subjective difficulty measures and performance. These experiments employ different combinations of three tasks: tracking, memory search, and a simulated air traffic control problem. The results supported all forms of dissociation predicted by the theory and the implications of results to workload measurement are discussed.

Wierwille, W. W. (1983). *Comparative Evaluation of Workload Estimation Techniques in Piloting Tasks* (NASA CR-166496). Washington, DC: National Aeronautics and Space Administration.

In January 1980, NASA Ames Research Center awarded a research grant to Virginia Polytechnic Institute and State University (Virginia Tech). The objective of this research was to examine the sensitivity and intrusion of a wide variety of workload-assessment techniques in simulated piloting tasks. The study employed four different piloting tasks emphasizing psychomotor, perceptual, mediational, and communications aspects of piloting behaviors. An instrumented moving-base general aviation aircraft simulator was used for the study. This document provides a summary of the research.

Wierwille, W. W. and Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 129-133). Santa Monica, CA: Human Factors Society.

The Cooper-Harper (1969) scale has been extensively used for evaluation of aircraft handling qualities and associated mental workload. The scale is a 10-point scale with a decision tree. A modified version of the scale, called the MCH scale, has been devised for the purpose of assessing workload in systems other than those where the human operator performs motor tasks; namely, where perceptual, mediational, and communications activity is present. The MCH scale has been validated in three different experiments. The scale is recommended for applications in which overall mental workload is to be assessed.

Wierwille, W. W. and Connor, S. A. (1983). Evaluation of twenty workload assessment measures using a psychomotor task in a motion-base aircraft simulator. *Human Factors*, 25 (1), 1-16.

The sensitivity and intrusion of 20 pilot workload assessment techniques were compared using a psychomotor loading task in a three degree-of-freedom moving-base aircraft simulator. The primary task was an instrument landing system approach and landing, with measures taken between the outer and middle markers. Three levels of psychomotor load were obtained by combined manipulation of random wind-gust disturbance level and pitch stability. Two rating scale measures and one control movement measure demonstrated sensitivity to all levels of load. Additionally, one time-estimation measure and one pulse-rate measure demonstrated sensitivity to some levels of load. No intrusion was found. The results of this experiment indicate that the sensitivities of workload estimation techniques vary widely, and that only a few techniques appear sensitive to psychomotor load.

Wierwille, W. W. and Connor, S. A. (1983). The sensitivity of twenty measures of pilot mental workload in a simulated ILS task. *Proceedings of the 18th Annual Conference on Manual Control (AFWAL-TR-83-3021)*. (pp. 150-162). Wright-Patterson AFB, OH: Air Force Wright Aeronautical Laboratories.

The sensitivity and intrusion of 20 pilot workload assessment techniques were compared using a psychomotor loading task in a three degree-of-freedom moving-base aircraft simulator. The primary task was an instrument landing system approach and landing, with measures taken between the outer and middle markers. Three levels of psychomotor load were obtained by combined manipulation of random wind-gust disturbance level and pitch stability. Two rating scale measures and one control movement measure demonstrated sensitivity to all levels of load. Additionally, one time-estimation measure and one pulse-rate measure demonstrated sensitivity to some levels of load. No intrusion was found. The results of this experiment indicate that the sensitivities of workload estimation techniques vary widely, and that only a few techniques appear sensitive to psychomotor load.

ORIGINAL PAGE IS
OF POOR QUALITY

APPENDIX F:
RESEARCH PAPERS AND PUBLICATIONS

-1984-

Berg, S. L. and Sheridan, T. B. (1984). Measuring workload differences between short-term memory and long-term memory scenarios in a simulated flight environment. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (pp. 397-416). Washington, DC: National Aeronautics and Space Administration.

Four highly experienced Air Force pilots each flew four simulated flight scenarios. Two scenarios involved less maneuvering, but required remembering a number of items. All scenarios were designed to be equally challenging. Pilot's subjective ratings for activity-level, complexity, difficulty, stress, and workload were higher for the maneuvering scenarios than the memory scenarios. At a moderate workload level, keeping the pilots active resulted in better aircraft control. When required to monitor and remember items, aircraft control tended to decrease. Pilots tended to weigh information about the spatial positioning and performance of their aircraft more heavily than other items.

Damos, D. (1984). Classification schemes for individual differences in multiple-task performance and subjective estimates. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (pp. 97-104). Washington, DC: National Aeronautics and Space Administration.

Human factors practitioners often are concerned with mental workload in multiple-task situations. Investigations of these situations have demonstrated repeatedly that individuals differ in their subjective estimates of workload. These differences may be attributed in part to individual differences in definitions of workload (Hart, Childress, and Hauser, 1982). However, after allowing for differences in the definition of workload, there are still unexplained individual differences in workload ratings. The general purpose of the two studies reported in this paper was to examine the relation between individual differences in multiple-task performance, subjective estimates of workload, information processing abilities, and the Type A personality trait.

Damos, D. L. (1984). *Examining the Relation between Subjective Estimates of Workload and Individual Differences in Performance* (NASA CR-2341). Washington, DC: National Aeronautics and Space Administration.

The primary purpose of this 2-year grant was to examine the relation between subjective estimates of workload, personality measures, and individual differences in single- and multiple-task performance. As specified in the grant proposal and second-year revision, four experiments were completed during the course of the grant examining these relations.

Damos, D. L. (1984). Individual differences in multiple-task performance and subjective estimates of workload. *Perceptual and Motor Skills*, 59, 567-580.

This experiment examined the relation between individual differences in multiple-task performance and subjective estimates of workload. Thirty female subjects performed various complex tasks alone and together and rated each task and task combination on ten bipolar adjective scales describing different dimensions of workload. The subjects also completed tests of field dependence, memory span, and time estimation. Two classification schemes were used to identify each subject. One was based on the subject's dual-task response strategy; the other, on the subject's performance on a complex monitoring task. However, the data showed little evidence of consistent individual differences on the monitoring task and this classification system was subsequently dropped. Between-response strategy group differences were found on two of the workload scales. Additionally, some between-group trends were found on the time estimation and memory span tasks, suggesting additional topics for investigation.

Damos, D. (1984). Subjective workload and individual differences in information processing abilities. *Proceedings of the Behavioral Engineering Technology Conference* (pp. 71-74). Warrendale, PA: Society of Automotive Engineers.

This paper describes several experiments examining the source of individual differences in the experience of mental workload. Three sources of such differences were examined: information processing abilities, timesharing abilities, and personality traits/behavior patterns. On the whole, there was little evidence that individual differences in information processing abilities or timesharing abilities are related to perceived differences in mental workload. However, individuals with strong Type A coronary prone behavior patterns differed in both single- and multiple-task performance from individuals who showed little evidence of such a pattern. Additionally, individuals with a strong Type A pattern showed some dissociation between objective performance and the experience of mental workload.

Goguen, J. A., Linde, C. and Murphy, M. (1984). Crew communications as a factor in aviation accidents. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (pp. 217-248). Washington, DC: National Aeronautics and Space Administration.

The basic motivation for the research reported here is to reduce the incidence of those air transport accidents caused wholly or in part by problems in crew communication and coordination. A major objective is to determine those communication patterns which actually are most effective in specific situations; this requires developing methods for assessing the effectiveness of crew communication patterns. It is hoped that these results will lead to the development of new methods for training crews to communicate more effectively, and in addition will provide guidelines for the design of aviation procedures and equipment.

This paper presents a number of results of a study on crew communication patterns in emergency situations, based on linguistic analysis applied to cockpit voice recorder (CVR) transcripts of commercial air transport accidents. The most important discourse types present are planning, explanation, and the command and control speech act chain.

Gopher, D. (1984). Measurement of Workload: Physics, Psychophysics, and Metaphysics. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (p. 55). Washington, DC: National Aeronautics and Space Administration.

The measurement of operator workload is an issue of great concern in the design and evaluation of modern engineering systems. This concern has led to the development of a wide arsenal of measurement techniques, all intended to quantify the phenomena accompanying the behavior of the human-processing system when its capacity to meet task demands has been exceeded. Three general categories of measurement approaches are performance-based measures, physiological indices, and subjective scales. In theory, the three approaches should constitute alternative strategies to expose the hidden limitations of internal processors. In practice, there is only a sparse knowledge on the relationship between workload measures obtained under different approaches. Moreover, there appears to be a debate among proponents of these approaches on the validity, comprehensiveness, and exclusiveness of different measures. The present paper reviews the results of two experiments in which workload analysis was conducted based upon performance measures, brain evoked potentials, and magnitude estimations of subjective load. The three types of measures were jointly applied to the description of the behavior of subjects in a wide battery of experimental tasks. Data analysis shows both instances of association and dissociation between types of measures. A general conceptual framework and methodological guidelines are proposed to account for these findings.

Gopher, D. (1984). *Workload Book: Assessment of Operator Workload in Engineering Systems* (NASA CR-166596). Washington, DC: National Aeronautics and Space Administration.

The report describes the structure and initial work performed toward the creation of a handbook for workload analysis directed at the operational community of engineers and human-factors psychologists. The goal of the report, when complete, will be to make accessible to such individuals the results of theoretically-based research that are of practical interest and utility in the analysis and prediction of operator workload in advanced and existing systems. In addition, the results of a laboratory study focused on the development of a subjective rating technique for workload that is based on psychophysical scaling techniques are described.

Gopher, D. and Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors*, 26 (5), 519-532.

Psychophysical functions describe the relationship between variations in the amplitude of a defined physical quantity and the psychological perception of these changes. Examples are brightness, loudness, and pain. The regularities of these relationships have been formulated into psychophysical laws. The measurement methodology of psychophysical scaling has been refined by the Harvard group led by S. S. Stevens, who proposed a power function as a general form for such laws. The main argument of the present article is that a similar scaling approach can be adapted to the measurement of workload and task demands based upon subjective estimates. The rationale is that these estimates, like other psychophysical judgements, reflect the individual's perception of the amount of processing

resources that the subject invests to meet the demand imposed by a task. This approach was successfully applied to the assessment of 21 experimental conditions given to a group of 60 subjects. The paper discusses the main results of this effort and their implications to theory and application in human performance.

- Hart, S. G., Battiste, V. and Lester, P. T. (1984). POPCORN: A supervisory control simulation for workload and performance research. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341) (pp. 431-454). Washington, DC: National Aeronautics and Space Administration.

A multitask simulation of a semi-automatic supervisory control system was developed to provide an environment in which training, operator strategy development, failure detection and resolution, levels of automation, and operator workload can be investigated. The goal was to develop a well-defined, but realistically complex, task that would lend itself to model-based analysis. The name of the task (POPCORN) reflects the visual display that depicts different task elements milling around waiting to be released and "pop" out to be performed. The operator's task was to complete each of 100 task elements that were represented by different symbols, by selecting a target task and entering the desired command. The simulated automatic system then completed the selected function automatically. Task difficulty, operator behavior, and experienced workload were varied by manipulating; (1) the number of elements per task; (2) the number of discrete tasks; (3) the penalties for lagging behind the system; (4) task schedule; and (5) payoff structure for performing or failing to perform task elements. Highly significant differences in performance, strategy, and rated workload were found as a function of all experimental manipulations (except reward/penalty). In addition, a proposed technique for reducing the between-subject variability of workload ratings was described and applied successfully. The first simulation conducted with this task defined a range of scenarios that imposed distinctly different levels of workload on operators and resulted in different levels of performance and operator strategies.

- Hart, S. G. and Bortolussi, M. R. (1984). Pilot errors as a source of workload. *Human Factors*, 26 (5), 545-556.

A pilot-opinion survey was conducted to develop a database for creating simulation scenarios that impose predetermined levels of pilot workload. Twelve pilots estimated the effect of 163 events and activities (which they had encountered during their previous flying experiences) on performance, effort, workload, and stress. The events, described in the context of flight scenario segments, included control, navigation and communications activities, aircraft and system failures, and pilot errors. In general, workload, stress, and effort ratings were significantly correlated with each other but not with performance ratings; however, some different response patterns were found as a function of flight segment (e. g., workload, stress, and performance, but not effort, ratings varied with flight phase) and type of event. Errors were rated as a significant source of change for workload, stress, and performance, suggesting that errors could be conceptualized as a cause of workload rather than as a symptom.

Hart, S. G., Hauser, J. R., and Lester, P. T. (1984). Inflight evaluation of four measures of pilot workload. *Proceedings of the 28th Annual Meeting of the Human Factors Society* (pp. 945-949). Santa Monica, CA: Human Factors Society.

Four measures of pilot workload were tested in the NASA C-141 Kuiper Airborne Observatory. The measures included a communications analysis, subjective ratings of workload, subjective ratings of additional factors related to workload, and heart rate. Data were collected for 11 flights, each of which lasted approximately 7 hours. Heart rate was found to be significantly higher for the pilot who was flying than for the pilots who were not flying and it varied significantly across flight segments, peaking during landing and take-off, particularly for the pilot in the left seat who was responsible for aircraft control. For both left and right seats, the subjective assessment of stress rather than the subjective assessment of workload was significantly correlated with variation in heart rate. Frequencies of different types of communications varied significantly across segments of flight, however, they were not correlated with subjective ratings of workload. There was a significant difference between the left and right seats in the types of activities that contributed to their workload, however, workload was considered to be equivalent for the two.

Hart, S. G., Sellers, J. J., and Guthart, G. (1984). The impact of response selection and response execution difficulty on the subjective experience of workload. *Proceedings of the 28th Annual Meeting of the Human Factors Society* (pp. 732-736). Santa Monica, CA: Human Factors Society.

The influence of variations in response selection and response execution difficulty on the workload and performance of 11 experimental subjects was investigated. The 20 laboratory tasks they performed involved a binary response selection that required different levels of mental processing (e.g., choice reaction time, prediction, memory search, etc.). A target-acquisition task was added following response selection on half of the trials. A weighted combination of bipolar ratings on nine workload-related dimensions was used to evaluate the workload experienced by the subjects. In addition, subjects rank-ordered the tasks with respect to workload before (a prediction) and again after (a retrospective comparison) performing them. Apparently minor variations in stimulus presentation resulted significantly increased reaction times and workload ratings, as did the more obvious manipulations of response selection load. The addition of the target-acquisition task increased workload ratings and reaction times, however the "cost" of performing the two-stage task (as indicated by measures of speed, accuracy, and subjective opinion) was considerably less than would be expected by combining measures for the component tasks. Movement times for the target acquisition tasks increased significantly as a function of the index of difficulty of the target, but were not affected by the difficulty of the response selection task.

Hart, S. G. and Sheridan, T. B. (1984). Pilot workload, performance, and aircraft control automation. *Proceedings of the AGARD Symposium on Human Factors Considerations in High Performance Aircraft - Conference Proceedings No. 371* (pp. 18.1 - 18.12). Neuilly sur Seine, France: NATO-Advisory Group for Aerospace Research and Development.

This report reviews conceptual and practical issues associated with the design, operation, and performance of advanced systems and the impact of such systems on the human operators. The development of highly automated systems has been driven by the

availability of new technology and the requirement that operators safely and economically perform more and more activities in increasingly difficult and hostile environments. It has become obvious that the workload of the operators, particularly their mental workload, may become a major area of concern in future design considerations. There has been, however, little research to determine how automation and workload relate to each other, although it is assumed that the abstract, supervisory, or management roles that are assumed by operators of highly automated systems will impose increased mental workload. The relationship between performance and workload, which is poorly understood at best for relatively simple tasks, will be discussed in relation to highly complex and automated environments.

Jensen, R. S. and Chappell, S. (1984). *Pilot Performance and Workload Assessment: An Analysis of Pilot Errors* (Final Report for NAS 2-184). Moffett Field, CA: NASA Ames Research Center

The preceding taxonomy of pilot errors provides a useful tool for the human factors investigator seeking answers to basic and applied problems in a real world aviation environment. The design of simulation scenarios that impose predictable and objectively determined levels of workload on pilots is essential in analyzing aircraft systems and procedures in applied environments as well as in developing metrics of pilot workload and performance in the laboratory. The occurrence of unplanned events (such as pilot errors) during the execution of the most carefully designed simulation scenario can result in the loss of costly and important data in such experiments.

By considering errors as a source of workload rather than as a symptom or product of workload, errors may be analytically and theoretically related to experimentally controlled variations in input load. Thus, the contribution of errors to flight task scenario workload can be computed and added to the original prediction of imposed load (Hart, 1983). Whenever pilots slip, blunder, err, or even hesitate, additional workload may be created because this forces them out of well-learned, automatic sequences of actions, and requires additional effort to discover, diagnose, and resolve the consequences of the error.

The belief that increased errors reflect increased workload is often expressed but less often supported by experimental results and needs clarification. The categorization of pilot-related behaviors with respect to impact on pilot workload provides a useful organizational scheme for a taxonomy of pilot behavior, with a particular emphasis on pilot errors. Such a taxonomy could be used to structure summarization and analysis of errors that are observed in flight-related research, and in reporting them in a standardized way. Errors observed under a variety of well-defined experimental situations and summarized in a common format as shown above, provide an understanding of the degree to which variation in imposed task demands and pilot effort cause errors.

Kantowitz, B. H., Hart, S. G., Bortolussi, M. R., Shively, R. J., and Kantowitz, S. C. (1984). Measuring pilot workload in a moving-base simulator: II. Building levels of load. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (pp. 359-372). Washington, DC: National Aeronautics and Space Administration.

Studies of mental workload conducted in flight simulators usually regard flying as a unitary task. Workload is varied by changing the mission and/or turbulence and little attempt is made to evaluate the individual workload required by a specific flight sub-task. As a first

effort in this direction, we chose three levels of flight sub-task complexity and measured the mental workload associated with each by an asynchronous secondary reaction-time task and by subjective ratings.

The base level of complexity was the simplest, requiring elementary maneuvers that do not utilize all the degrees of freedom of which an aircraft, or moving-base simulator, is capable. A base task would be maintaining constant airspeed, heading or altitude. A Paired level task required two base tasks performed simultaneously. A complex level task required three base tasks performed simultaneously.

Primary task (flying) performance was not adversely affected by the addition of the auditory secondary reaction-time task. Rate of transmitted information (bits/sec) on the secondary task was able to discriminate among all three levels of the flight task for dual-task conditions. Furthermore, single-stimulation transmitted information rate was reliably greater than any dual-task performance, indicating that even the elementary maneuvers of the base level imposed some mental workload. Subjective ratings also discriminated among the three levels of the flight task and in addition were sometimes able to discriminate between tasks within the same level of complexity.

Miller, R. C. and Hart, S. G. (1984). Assessing the subjective workload of directional orientation tasks. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (pp.85-96). Washington, DC: National Aeronautics and Space Administration.

An experiment was conducted to investigate the impact of various flight-related tasks on the workload imposed by the requirement to compute new headings, course changes and reciprocal headings. Eight instrument-rated pilots were presented with a series of heading-change tasks in a laboratory setting. Two levels of difficulty of each of three tasks were presented verbally (numeric values imbedded in simple commands) and spatially (headings were depicted on a graphically drawn compass). Performance was measured by evaluating the speed (response times) and accuracy (percent correct and time outs) of the responses. The workload experienced by the pilots under each experimental condition was determined by responses to a standard set of bipolar rating scales. The subjective responses and objective measures of performance reflected a strong association between subjective experience and objective behavior. The reciprocal calculations were performed quickly and accurately throughout and were considered to be minimally loading. Subjective workload, percent correct, and response times for the two course-change tasks varied significantly as a function of level of difficulty and display format, with no discernable speed/accuracy trade off. The results of this study will be used to predict the workload that is imposed on pilots of actual and simulated flights by course corrections and computations in conjunction with previously obtained estimates of control and communications workload.

Moray, N. (1984). Recent research in mental workload. Presented at the International Conference on Occupational Ergonomics. Toronto, Canada.

Abstract not available.

Moray, N. and King, B. (1984). Error as a cause and effect of workload: Mental workload as a closed loop system. Working Paper 84-11. Canada: University of Toronto, Department of Industrial Engineering.

Participants performed mental arithmetic tasks under conditions with various kinds of feedback, and rated the tasks for subjective effort and difficulty. The greatest increase in ratings occurred when errors resulted immediately in more work to be done. But even in a condition where the occurrence of an error made no difference to the work to be done, knowing that an error had been made increased the subjective difficulty and effort of the task. It appears that error can be a cause, as well as an effect of workload, and hence there are problems in using performance degradation as a dependent measure of load.

Moray, N., King, B., Turksen, B., and Waterton, K. (1984). A closed-loop causal model of workload based on a comparison of fuzzy and crisp measurement techniques. Working Paper 84-12. Canada: University of Toronto, Department of Industrial Engineering.

Fuzzy and crisp measurement of workload are compared for a tracking task which varied in bandwidth and order of control. Fuzzy measures are as powerful as crisp measurement, and can under certain conditions give extra insights into workload causality. Both methods suggest that workload arises in a system in which effort, performance, difficulty and task variables are linked in a closed-loop. Marked individual differences were found. Future work on the fuzzy measurement of workload is justified.

Murphy, M. R. (1984). Space station application of simulator-developed aircrew coordination and performance measures. *Proceedings of the Workshop on Advances in NASA-Relevant Minimally Invasive Instrumentation*. Pasadena, CA: Jet Propulsion Laboratory.

This paper discusses some ongoing work at NASA Ames Research Center to develop linguistic and video-derived measures of aircrew interaction factors and to relate these factors to flight task performance. Results of prior research are summarized, and a study in progress that measures interpersonal interaction factors within a full mission simulator environment is presented. The possible application of similar methodology to space station crew performance research is also discussed.

In the current study, three-man airline crews fly a full-mission scenario designed to elicit a high level of verbal interaction during instances of critical decision-making and resources management. The scenario is implemented in a flight-training simulator augmented to record simulator state data, voice communications data, and video-taped images of individual and crew (context) performance. Following the simulator run, each crew member and each of two observer-raters independently view the four video recordings presented on a quartile split screen, and make interpretive comments at viewer-selected stopping points (times within the scenario). The instructions solicit comments on events judged to be important in fostering, or recovering from, problematic crew coordination and task performance. Time and interpretive concordance results are calculated from this data base.

A linguistic analysis of voice transcripts is made to identify other variables that provide quantitative measures of crew coordination. Crew coordination factors as assessed by video peer review, linguistic analyses, and observer ratings are then correlated with crew and system performance measures.

Murphy, M. R., Randle, R. J., Tanner, T. A., Frankel, R. M., Goguen, J. A., and Linde, C. (1984). The measurement of crew coordination factors and their relationships to flight task performance. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (pp. 249-262). Washington, DC: National Aeronautics and Space Administration.

Sixteen three-man crews flew a full-mission scenario in an airline flight simulator. The scenario was designed to elicit a high level of verbal interaction during instances of critical decision-making. Each crew flew the scenario only once, without prior knowledge of the scenario problem. Following a simulator run and in accord with formal instructions, each of the three crewmembers independently viewed and commented on a videotape of their performance. Two check-pilot observers rated pilot performance across all crews and, following each run, also commented on the video tape of that crew's performance. A linguistic analysis of voice transcripts is being made to provide added assessment of crew coordination and decision-making qualities. Measures of crew coordination and decision-making factors are being correlated with flight task performance measures. Some results and conclusions from observational data are presented.

Sheridan, T. B. and Berg, S. (1984). Supervisory workload: Monitoring of overlapping tasks. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (pp. 397-316). Washington, DC: National Aeronautics and Space Administration.

It is hypothesized that significant causes of mental workload in supervisory control are the requirements (1) to keep track of multiple overlapping task schedules, (2) to cope with time delays in knowledge of results, and (3) to tolerate unexpected interruptions and forced modifications of plan. Experiments are reported in which experienced pilot subjects fly terminal area let-down scenarios on a fixed-base simulator with varying degrees of overlapping mental tasks, time delays in feedback, and interruptions. Other experiments are reported using an abstract multitask computer game where subjects have to keep track of overlapping tasks and make correct responses at the appropriate time. In both experiments, both subjective workload ratings and objective performances are correlated with various task variables.

Silverstein, L. D., Gomer, F. E., Crabtree, M. S. and Acton, W. H. (1984). *A Comparison of Analytic and Subjective Techniques for Estimating Communications-Related Workload During Commercial Transport Flight Preparations* (NASA CR-2341). Washington, DC: National Aeronautics and Space Administration.

The objectives of this research contract were to develop a classification scheme for categorizing commercial transport communications and to apply analytic and subjective estimation techniques to quantify the workload imposed by these communications tasks. A communications task was defined as the sequence of perceptual, cognitive, motor, and verbal responses initiated by the aircrew immediately following transmission of a message

or instruction from ATC. Four techniques were used to quantify the workload: (1) an information-theoretic analysis, (2) paired-comparison technique for obtaining the opinions of current line pilots, (3) a combined hybrid scale that combined information from the other two techniques, and (4) a subjective rank-order scale. Highly significant agreement was found among the different methods of estimating communications workload. The results of this research provided a basis for the selection of standard sets of communications tasks with variable loading characteristics. Such a standard task repertoire can be used to control communications-related demands in future simulation research and should serve as input to a data-base of "workload calibrated" flight tasks.

Vidulich, M. A. and Wickens, C. D. (1984). Subjective workload assessment and voluntary control of effort in a tracking task. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (pp. 57-72). Washington, DC: National Aeronautics and Space Administration.

A manual-control tracking task was manipulated along two dimensions: (1) control order, and (2) forcing function bandwidth. In the first phase of the experiment subjective workload assessments were collected. It was found that subjective assessments of workload were closely associated with performance in the case of increasing control order, but not in the case of increasing bandwidth. This was interpreted as indicating that subjective workload assessments are most appropriate for the study of increasing difficulty centered in response-selection processes as opposed to response execution processes. In the second phase of the experiment the subjects were asked to voluntarily limit the effort they applied in the performance of the tracking task. The results indicate that the subjects were quite facile in doing this. However, comparison of this data to the findings of other studies that manipulated effort via dual-task biasing indicate that effort manipulation is much more potent in a single-task configuration. This finding is discussed in terms of multiple-resource theories of attentional capacity. Also, the utility of an analysis of covariance (ANACOVA) procedure in studying the relationships between subjective ratings and performance is highlighted.

White, S. A., Mackinnon, D. P. and Lyman, J. (1984). *Structuring Modified Petri Net Model-Based Assessment of Workload Components*. Research Report to NASA Ames Research Center. Los Angeles: University of California, Los Angeles.

The current research has been undertaken to investigate a novel approach to a model that partitions workload into events and activities for individual and concurrent tasks. The approach, which utilizes Modified Petri Nets to operationalize the workload model, is intended for generalizing to a large class of supervisory control tasks (e.g. supervisory tasks in a modern cockpit).

The experimental vehicle is the Supervisory Control Training Simulation developed at NASA Ames Research Center. The SCTS has been modeled with a MPN representation. The principle reasons for selecting this approach are that MPNs are able to model concurrent tasks and can also objectively model tasks with a large mental workload component. By manipulating parameters of the SCTS that affect workload (time stress, number of concurrent tasks, and task payoffs), we can test the model's sensitivity to workload changes.

After the model has been tested and refined, workload values specific to the event and activity classifications can be derived. The resulting system can then be used to model other supervisory workload situations and make prescriptive workload predictions.

Wickens, C. D., Vidulich, M. A., and Sandry-Garza, D. (1984). Principles of S-C-R compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. *Human Factors*, 26 (5), 533-544.

A pilot's tasks may be categorized into those that demand predominantly verbal operations and those that are spatial. We describe two experiments that define two principles of compatibility of interfacing such tasks with displays and controls. The first, based upon hemispheric laterality effects, defines compatibility according to the display location and the response hand; the second defines compatibility according to the modality of display (auditory and visual) and response (manual and speech). Verbal tasks are best served by auditory inputs and speech response, whereas spatial tasks are best served by visual-manual channels. In both experiments, these principles of compatibility are confirmed under dual-task conditions. We describe their implications for cockpit design.

Wierwille, W. W., Skipper, J. H., and Rieger, C. A. (1984). Decision tree rating scales for workload estimation -- Theme and variations. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (pp. 73-84). Washington, DC: National Aeronautics and Space Administration.

The Modified Cooper-Harper (MCH) scale has been shown to be a sensitive indicator of workload in several different types of aircrew tasks (Wierwille and Casali, 1983). The study to be described in this paper was undertaken to determine (1) if certain variations of the scale might provide even greater sensitivity and (2) the reasons for the sensitivity of the scale. The MCH scale, which is a 10 point scale, and five newly devised scales were examined in two different aircraft simulator experiments in which pilot loading was treated as an independent variable. The five scales included a 15 point scale, computerized versions of the MCH and 15 point scales, a scale in which the decision tree was removed, and one in which a 15 point left-to-right format was used.

Yeh, Y. -Y. and Wickens, C. D. (1984). *The Dissociation of Subjective Measures of Mental Workload and Performance* (NASA CR-2341). Washington, DC: National Aeronautics and Space Administration.

This report describes research conducted during the first years under a contract from NASA Ames Research Center; Dr. Sandra Hart was the technical monitor. The report addresses the *dissociation* between subjective measures of mental workload and performance. Three generic factors are identified that will drive subjective workload ratings upward more than driving performance downward: Perceptual (vs response) load, and increased number of tasks, and better data quality. One factor, resource competition, is assumed to drive performance more than subjective workload. The theory of dissociation is tested in three experiments that employ different variations and combinations of three different tasks. Predictions of the theory are generally supported by the data. In addition, various subjective scales of mental workload are tested across the experiments. The correlations between these scales and multidimensional scaling data are used to help interpret the hidden cognitive structure of task difficulty.

Yeh, Y. -Y. and Wickens, C. D. (1984). Why do performance and subjective workload measures dissociate? *Proceedings of the Human Factors Society 28th Annual Meeting* (pp. 504-508). Santa Monica, CA: Human Factors Society.

A set of three experiments is described that examine the sources of information processing that produce a dissociation between subjective workload measures and performance. The experimental results support a theory of the dissociation. Subjective measures are driven more by the number of tasks currently performed and are also less sensitive to resource competition than are performance measures. Factors that demand more resource investment improve performance, but these factors also increase subjective ratings of workload.

Zaleski, M. and Sanderson, P. (1984). Hitts' Law? A test of the relationship between information load and movement precision. *Proceedings of the 20th Annual Conference on Manual Control* (NASA CP-2341). (pp. 575-584). Washington, DC: National Aeronautics and Space Administration.

An experiment was run to test the independence of information load (Hick's Law) and movement precision (Fitts' Law) using additive factors methodology. There were two elements to the subjects' task. Subjects were required to classify stimuli according to a decision rule with a variable entropy. The stimuli were presented in the center of the CRT screen. In response, subjects had to move a cursor from a starting point near the stimulus to the appropriate target. The targets were arranged in an annular pattern around the central point. The precision of the response movement was varied by manipulating the ratio of the radius of the annulus to the width of the target area. The dependent measure was elapsed time between onset of the stimulus and completion of the response movement. Independence of the Hick's Law and Fitts' Law components of the reaction time was tested with an analysis of variance. Presence of an interaction would suggest that a decision stage and a response stage are dependent, and cannot be considered discrete steps in a serial process.

ORIGINAL PAGE IS
OF POOR QUALITY

APPENDIX G:
RESEARCH PAPERS AND PUBLICATIONS

- 1985 -

- Allen, E. M. R. and Discepola, M. N. (1985). *Mental Workload in Rule-Based Problem Solving*. Unpublished M.S. Thesis. Toronto, Canada: University of Toronto.

Rasmussen's taxonomy of human information, specifically rule-based behavior, acted as a basis for this NASA sponsored experimental research in mental workload determination. The integration of both qualitative subjective analysis and quantitative physiological measures was used in order to develop a more accurate representative modeling of the human automated motor/information processing systems. Particular attention was paid to experimental design and procedure in order to justify significant findings. Three levels of difficulty (easy, moderate, and difficult) rules were found that were statistically distinguishable at a 99.5% level of confidence using both subjective and physiological data.

- Battiste, V. and Hart, S. G. (1985). Predicted versus experienced workload and performance on a supervisory control task. *Proceedings of the Third Biannual Symposium on Aviation Psychology* (pp. 255-262). Columbus: Ohio State University.

Some of the stated goals in workload research are to provide information about operator workload in existing systems and predict the impact of modifications of existing systems, to provide designers with an accurate estimate of the expected workload of new systems at the inception of the design stage. This was the second in a series of studies conducted with a multi-task simulation of a supervisory control system. The operators' task was to complete a number of task elements (represented by different symbols), by selecting a target task and entering the desired command. Task difficulty and experienced workload were varied by manipulating the number of elements per task, the number of tasks, task schedule and availability of task elements for performance. The goal was to investigate operators' abilities to predict the workload and performance impact of unfamiliar task features and configurations from their basic knowledge of the system and the specific information provided before each scenario. Significant differences in performance and workload were found as a function of all experimental manipulations. In addition, different relationships between workload predictions and workload ratings were found due to the similarity of the task modifications to familiar levels and to task complexity.

- Berg, S. L. and Sheridan, T. B. (1985). *Effect of Time Span and Task Load on Pilot Mental Workload* (Final Report for NASA Grant NAG 2-227, CR-177388). Washington, DC: National Aeronautics and Space Administration.

Two sets of simulations are described that were designed to examine how a pilot's mental workload would be affected by continuous manual-control activity versus discrete mental tasks that included the length of time between receiving an assignment and executing it. A fixed-base flight simulator was used that consisted of a control box (joystick, throttle, switches for operating electronic and mechanical systems) and a high resolution CRT. Aircraft dynamics were modeled on a Lockheed Jetstar business jet. The CRT display

consisted of a forward "out-the-window" perspective view and a cockpit instrument/indicator presentation. The first experiment evaluated two types of measures: objective performance indicators and subjective ratings. Pilots flew two missions: a high-workload manual control mission and a high-workload mission that emphasized mental activities. Subjective ratings for the two missions were different, but the objective performance measures (altitude deviations) were similar. In the second set of experiments, workload levels were increased and a second performance measure was taken (e.g., airspeed deviations). Mental workload had no influence on either performance-based workload measure. Subjective ratings discriminated among the scenarios and correlated with performance measures for high-workload flights. The number of mental tasks performed did not influence error rates, although high manual workloads did increase errors.

Biferno, M. A. (1985). *Mental Workload Measurement: Event-Related Potentials and Ratings of Workload and Fatigue* (NASA CR-177354). Washington, DC: National Aeronautics and Space Administration.

Event-related potentials were elicited when a digitized work representing a pilot's call-sign was presented. This auditory probe was presented during 27 workload conditions in a 3 x 3 x 3 design where the following variables were manipulated: short-term memory load, tracking task difficulty, and time-on-task. Ratings of workload and fatigue were obtained between each trial of a 2.5-hour test. The data of each subject were analyzed individually to determine whether significant correlations existed between subjective ratings and ERP component measures. Results indicated that a significant number of subjects had positive correlations between: (1) ratings of workload and P300 amplitude, (2) ratings of workload and N400 amplitude, and (3) ratings of fatigue and P300 amplitude. These data are the first to show correlations between ratings of workload or fatigue and ERP components thereby reinforcing their validity as measures of mental workload and fatigue. Since ratings of fatigue and workload were significantly correlated for 16 of 20 subjects, future studies of workload would benefit from examining the relationship between them.

Bloem, K. A. and Damos, S. L. (1985). Individual differences in secondary task performance and subjective estimation of workload. *Psychological Report*, 56, 311-322.

This experiment had two purposes. First, it attempted to replicate the easy-to-hard prediction for residual capacity described by Lansman and Hunt (1982) for two complex task combinations. Second, it examined the relation between individual differences in resource capacity, as indicated by the easy-to-hard prediction, and the subjective experience of workload. One task combination involved a verbal memory task paired with a vowel-consonant classification task. The other combination involved a paired associate task with a name classification task. The easy-to-hard prediction was not replicated for either task combination: easy primary task performance provided a better prediction of hard primary task performance than did secondary task performance. Measures of residual capacity were not related to subjective ratings of workload, however, the workload scales were sensitive to between-task differences.

ORIGINAL PAGE IS
OF POOR QUALITY

Bortolussi, M. R., Kantowitz, B. H., and Hart, S. G. (1985). Measuring pilot workload in a motion base trainer: A comparison of four techniques. *Proceedings of the Third Biannual Symposium on Aviation Psychology* (pp. 263-270). Columbus: Ohio State University.

Various techniques have been developed to predict and measure pilot workload. This simulation was conducted in order to compare four widely used methods: A visual two- and four-choice reaction time task, time production, retrospective multidimensional subjective ratings and in-flight verbal workload estimates. Two scenarios with different levels of difficulty as determined by preliminary research were designed to test these techniques. The insertion of the secondary tasks did not significantly affect flight performance. All four techniques were able to distinguish among levels of scenario complexity. In addition, the three secondary tasks and workload ratings obtained in-flight were generally able to distinguish among levels of difficulty for different segments within the scenarios.

Casper, P. A. and Kantowitz, B. H. (1985). Seeing tones and hearing rectangles: Attending to simultaneous auditory and visual events. In R. Eberts and C. G. Eberts (Eds.), *Trends in Ergonomics: Human Factors, Volume II* (pp. 41-49). Amsterdam: Elsevier Science Publications, North Holland Press.

The allocation of attention of dual-task situations depends on both the overall and the momentary demands associated with both tasks. Subjects in an inclusive-or reaction-time task responded to changes in simultaneous sequences of discrete auditory and visual stimuli. Performance on individual trials was affected by (1) the ratio of stimuli in the two tasks, (2) response demands of the two tasks, and (3) patterns inherent in the demands of one task.

Chan, G. and Krushelncky, E. (1985). *Mental Workload in Knowledge-Based Problem Solving*. Unpublished M.S. Thesis. Toronto, Canada: University of Toronto.

Rasmussen's taxonomy of human information acted as a basis for this NASA sponsored experimental research in mental workload determination. The integration of both qualitative subjective analysis and quantitative physiological measures was used in order to develop a more accurate representative modeling of the human automated motor/information processing/decision-making systems. Particular attention was paid to experimental design and procedure in order to justify significant findings. A dialectic relationship was found to exist with respect to greater difficulty of an assigned task versus perceived effort. The greater the "difficulty," the more pronounced the disparity became. Heuristic effects were also found to play an important role in operator behavior and preconceived notions of operator performance were experimentally validated. Three levels of difficulties were found that were statistically distinguishable at a 99.5% level of confidence using both subjective and physiological data. Three distinct levels of human behavior were found: skill-based, rule-based, and knowledge-based, which could be relatively accurately modeled by the human process control model developed.

Damos, D. L. (1985). The relation between the type A behavior pattern, pacing, and subjective workload in single- and dual-task conditions. *Human Factors*, 27(6), 675-680.

Twenty Type A and twenty Type B subjects performed two discrete tasks alone and together. Half of the subjects performed paced versions of both tasks; half, unpaced versions. Workload ratings were obtained for all subjects under single- and dual-task conditions using eight bipolar objective scales. Under single-task conditions there was a significant interaction between behavior pattern and pacing on one of the tasks. This interaction indicated that Type A subjects responded more rapidly under unpaced conditions than did Type B subjects, although there was little difference between the groups under paced conditions. Under dual-task conditions, Type A subjects responded more rapidly than did Type B subjects regardless of pacing. There was one significant interaction between behavior pattern and task on one of the workload scales.

Damos, D. (1985). Type A behavior pattern, multiple-task performance, and subjective estimation of workload. *Bulletin of the Psychonomic Society*, 23, 53-56.

This paper examines the relation between the Type A behavior pattern, individual differences in multiple-task performance, and the dissociation between performance and subjective estimates of mental workload. Sixteen females completed the Jenkins Activity Survey and performed a variety of information-processing tasks under single- and dual-task conditions. After each task, subjects rated the workload they experienced on eight bipolar adjective scales. The slope of the memory-search task was the only single-task performance measure that showed a significant difference between Type As (36 ms) and Type Bs (68 ms). However, on three of the four dual-task combinations, Type As had faster response times than Type Bs. Dissociations between performance and subjective estimates of workload were apparent between Type A and Type B individuals on the frustration and fatigue scales. Type As reported less frustration and more fatigue under single-task than under dual-task conditions with Type Bs reporting the opposite pattern.

Frankel, R. M. (1985). "Captain, I was trying to bring up the fact that you made a mistake earlier:" Deference and demeanor at 30,000 feet. *Proceedings of the Third Biannual Symposium on Aviation Psychology* (pp. 403-410). Columbus: Ohio State University.

Preliminary evidence from the analysis of a single error suggests that there may be some practical utility to viewing cockpit communication as a microinteractional process. Qualitative and quantitative studies of interactional complexity, deference, and demeanor will increase our understanding of the dynamic group processes involved in communication breakdowns in the cockpit. In addition the use of a video-based research paradigm may enhance the development and impact of training programs in communication skills.

Gopher, D., Chillag, N., and Arzi, N. (1985). *The Influence of Voluntary Effort, Context, and Anchor Task on the Subjective Estimate of Load* (Final Report for NASA Grant NAGW-494). Haifa, Israel: Technion.

Subjects were given in three separate experimental sessions, a size-matching and a letter-typing task under six levels of emphasis (priorities). One session included a mixture of

trials of the two tasks both in single- and in dual-task conditions. Two sessions, one for each task, comprised only single-task trials of one task with emphasis manipulation. Subjects received individualized, continuous, visual feedback on their performance. It included a desired performance line and a moving bar graph (the height of which displayed the difference between actual and desired performance). As emphasis levels on a task were increased it became increasingly more difficult to raise its bar graph so as to match the height of the desired performance line. Following each trial, subjects were asked to give a magnitude estimate of the subjective load imposed by that trial. Estimates were given relative to a predefined reference condition. For one half of the sample, medium level size matching served as the reference task. For the other half, letter typing was used as a reference. Subjective measures were highly sensitive to priority change under all conditions and also increased in the transition from single- to dual-task conditions. They were not affected by the context of the surrounding tasks. Performance was not sensitive to priority changes, except for secondary task performance under dual-task conditions. Average performance levels were lower on dual as compared to single-task conditions, and on a task when it was secondary. The type of reference had a strong impact on both subjective ratings and performance. Higher load estimates were given to a task when it served as reference and performance on it deteriorated.

Gopher, D., Chillag, N., and Arzi, N. (1985). The psychophysics of workload -- A second look at the relationship between subjective measures and performance. *Proceedings of the of the Human Factors Society 29th Annual Meeting* (pp. 640-644). Santa Monica, CA: Human Factors Society.

Load estimates based upon subjective and performance indices were compared for subjects performing size matching and letter typing tasks under six levels of priorities, in single and dual task conditions. Each half of the group used a different task as reference in their subjective judgement. The results are interpreted to indicate that subjective measures are especially sensitive to voluntary allocation of attention and to the load on working memory. Association with performance is expected whenever these two factors are main determinants of performance efficiency, otherwise the two are likely to dissociate.

Hart, S. G. (1985). *Workload: Definition, Prediction and Assessment*. Invited Presentation for the U. S. Army MANPRINT course, Rosslyn, VA.

A comprehensive review of the field of workload assessment and prediction was provided for the participants in the MANPRINT course. The topics covered included: (1) a conceptual model of human performance and workload; (2) workload definitions; (3) types of workload measures--subjective, performance, secondary task, and physiological; (4) operational applications of different measures; (5) methods of predicting workload; (6) evaluating the results of workload assessment predictive efforts; (7) why measure workload.

Jacrew, M. and Vincente, K. (1985). *An Investigation of the Mental Workload Associated with Skill-Based Behavior*. Unpublished M.S. Thesis. Toronto, Canada: University of Toronto.

A pursuit-tracking task with preview has been used to study the mental workload associated with skill-based behavior. The experiment consisted of sixteen 1-hour sessions, each of which was composed of nine trials on the task. The track width and turbulence

**ORIGINAL PAGE IS
OF POOR QUALITY**

distribution were manipulated in a 3 x 3 factorial design. The measures taken were: a) time to complete a run; b) total number of wall hits; c) subjective estimates of difficulty; d) subjective estimates of effort; e) the amplitude of the 0.1 Hz component of sinusarrhythmia. As subjects became more skilled with practice their performance stabilized. The subjective ratings consistently decreased with practice, indicating a reduction in subjective load. Mulder's physiological measure of workload also showed a decrease in load with practice across all configurations. Although the technique was found to be nonintrusive, the large amount of noise present in the signal meant that the technique was not sensitive enough to discriminate between configurations. Subjects expended more effort on the easier trials, while on the more difficult ones they reduced both their goals and their effort. Because of the large amount of noise in the heart rate data, it was not possible to reliably test whether or not skill-base behavior implies zero load.

Kantowitz, B. H. (1985). Channels and stages in human information processing: A limited review. *Journal of Mathematical Psychology*, 29, 135-174.

This article reviews the status of the theoretical construct of capacity. Four basic questions are discussed: (1) What is capacity? (2) How is capacity measured? (3) Is capacity limited? (4) If so, where is it limited? It is claimed that empirical answers to these questions have been unsatisfactory due to theoretical and methodological issues that need be resolved. Data are presented to illustrate such difficulties. It is concluded that the construct of capacity has become more and more vacuous and that caution is required whenever capacity is invoked to explain behavior.

Kantowitz, B. H. and Weldon, M. (1985). On scaling performance operating characteristics: Caveat emptor. *Human Factors*, 27(5), 531-548.

Problems associated with scaling and normalizing empirical Performance Operating Characteristics (POCs) are examined. Normalization methods proposed by Wickens (1980) and by Mountford and North (1980) are critically evaluated. Computer simulations are used to generate raw-score and normalized POCs. Wickens' interpretation of transformed empirical POCs is shown to contain inconsistencies. The normalization techniques reviewed fail to resolve POC scaling problems. Caution must be exercised when interpreting transformed POCs.

Klapp, S. T., Kelly, P. A., Battiste, V., and Dunbar, S. (1985). Types of tracking errors induced by concurrent secondary manual tasks. *Proceedings of the 20th Annual Conference on Manual Control (NASA CP-2341)*. (pp. 299-304). Washington, DC: National Aeronautics and Space Administration.

Future one-man helicopters may require the pilot to control flight with one hand, and simultaneously manipulate other instruments using the other hand. This report of work in progress examines the nature of errors induced in a right hand tracking task (simulating flight control) when responses are required by the left hand. The present experiment focused on detection of hesitations in which the tracking joy stick remained motionless for 1.3 s or longer.

Moray, N., Turksen, I. B., & Thornton, C. (1985). Some conceptual problems in the measurement of workload. Presentation for the Human Factors Association of Canada.

Abstract not available.

Morris, N. M. and Rouse, W. B. (1985). *Human Error Tolerance in Complex Systems: Design Concepts and Research Approach*. Paper presented at NASA Ames Research Center, Aerospace Human Factors Research Division, Moffett Field NAS, CA.

Various surveys and compilations have led to conclusions that "human error" is a primary cause of most major accidents in aviation, power production, and process control. A strategy that is likely to be successful in ameliorating the problem of human error is one that seeks to tolerate the consequences of errors when they occur. Three complementary approaches to error tolerance will be discussed, and the conceptual design of a human error tolerant interface will be presented. An important feature of the proposed interface involves online error diagnosis and remediation in a manner appropriate to the error. If implementation of this feature is to be possible, a greater understanding of both the causes of error and contributing factors is necessary. Further, the effects of various interface characteristics upon subsequent human performance must be determined. Research directed at increasing this understanding is being conducted within the context of a process control task, PLANT, and results to date will be presented.

Morris, N. M. and Rouse, W. B. (1985). An experimental approach to validating a theory of human error in complex systems. *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 333-337). Santa Monica, CA: Human Factors Society.

The problem of "human error" is pervasive in engineering systems in which the human is involved. In contrast to the common engineering approach of dealing with error probabilistically, the present research seeks to alleviate problems associated with error by gaining a greater understanding of causes and contributing factors from a human information-processing perspective. The general approach involves identifying conditions which are hypothesized to contribute to errors, and experimentally creating the conditions in order to verify the hypotheses. The conceptual framework which serves as a basis for this research is discussed briefly, followed by a description of upcoming research. Finally, the potential relevance of this research to design, training, and aiding issues is discussed.

Shively, R. J. (1985). *Evaluation of Data Entry Devices*. Unpublished M.S. Thesis. West Lafayette, IN: Purdue University.

Evaluations of mental workload has been used in the aerospace community for some time. Human-computer interactions have many of the same properties that led to widespread usage in the complex environments of flight and space exploration. Mental workload evaluation brings with it a strong theoretical base for interpretation. This, as well as the validation of these techniques in operational environments argues for the inclusion of mental workload analysis in human-computer interaction research. The present experiment applies a subjective rating technique to a computer interaction laboratory task. The use of a laboratory task allows a comparison to the subjective ratings. The subjective

rating technique used provided essentially the same information as the performance measures. It is argued that the inclusion of mental workload assessment in human-computer research will lead to a fuller understanding of the performance of the system as a whole and a more complete understanding of the cost to the operator of completing the task.

Thornton, D. C. (1985). An investigation of the "von Restorff" phenomenon in post-test workload ratings. *Proceedings of the Human Factors Society 29th Annual Meeting*. Santa Monica, CA: Human Factors Society.

The present experiment was designed to examine the possibility of a "von Restorff" effect occurring during post-task ratings of difficulty, effort and workload as a result of performing tasks containing an isolated period of high workload. The task employed was a hovercraft simulation which combined elements of skill-based tracking and rule- and knowledge-based process control. The subjects performed missions which presented a relatively low level of workload from throughout while the second presented a high level of workload from throughout. The final three conditions were designed to produce isolated peaks in workload which occurred either early, midway or late in the mission. Results demonstrated that ratings on the three scales increased steadily as the peak approached the end of the mission. Further, the ratings of the early increase were only slightly greater than those reported for the task containing a low level of workload throughout. An analysis of error rates demonstrated that there was evidence for the dissociation of performance and workload which was especially apparent when the increase occurred at the beginning of the mission.

Tsang, P. S., Hartzell, E. J., and Rothschild, R. A. (1985). To speak or not to speak: A multiple resource perspective. *Proceedings of the Human Factors Society 29th Annual Meeting*, Vol. 1. (pp. 76-80). Santa Monica, CA: Human Factors Society.

The desirability of employing speech response in a dynamic dual task situation was discussed from a multiple resource perspective. A secondary task technique was employed to examine the time-sharing performance of five dual tasks with various degrees of resource overlap according to the structure-specific resource model (Wickens, 1980). The primary task was a visual manual tracking task which required spatial processing. The secondary task was with another tracking task (or a spatial transformation task) with one of four input (visual or auditory) and output (manual or speech) configurations. The results show that the dual-task performance was best when the primary tracking task was paired with the visual/speech transformation task. This finding was explained by an interaction of the S-C-R compatibility of the transformation task and the degree of resource competition between the time-shared tasks. Implications on the utility of speech response were discussed.

Tsang, P. S. and Wickens, C. D. (1985). The effects of task structures on time-sharing efficiency and resource allocation optimality. *Proceedings of the 20th Annual Conference on Manual Control*, 2, (NASA CP-2341). (pp. 305-317). Washington, DC: National Aeronautics and Space Administration.

A distinction was made between two aspects of time-sharing performance: time-sharing efficiency and attention allocation optimality. The first is concerned with the level of joint performance of the time-shared tasks. The second is concerned with the consistency of protecting the performance of a high priority task from varying with changes in task demand. A secondary task technique was employed to evaluate the effects of the structures of the component time-shared tasks on both aspects of the time-sharing performance. Five pairs of dual tasks differing in their structural configurations were investigated. The primary task was a visual/manual tracking task which requires spatial processing. The secondary task was either another tracking task or a verbal memory task with one of four different input/output configurations. Congruent to a common finding, time-sharing efficiency was observed to decrease with an increasing overlap of resources utilized by the time-shared tasks. Results also tend to support the hypothesis that resource allocation is more optimal when the time-shared tasks placed heavy demands on common processing resources than when they utilized separate resources. These data suggest that careful consideration of the tradeoff between time-sharing efficiency and resource allocation optimality is necessary in making multitask design decisions.

Turksen, I. B., Moray, N., and Fuller, K. (1985). A linguistic rule-based expert system for mental workload. In H. J. Bullinger and H. J. Warnecke (Eds.), *Toward the Factory of the Future*. (pp. 865-875). The Netherlands: Springer-Verlag.

Although the conventional mathematical techniques have been and will continue to be applied to the analysis of humanistic systems, it is clear that the great complexity of such systems calls for approaches that are significantly different in spirit as well as in substance from traditional methods... experts like to describe workload phenomena which historically have been found to be too complex or too ill-defined to be susceptible of characterization in precise quantitative terms with any degree of ease. In the design of expert workload system analysis, our intention is to incorporate linguistic rules of the form: "If bandwidth is *high* and the operator uses *moderate* effort, the task will seem *moderately* difficult and performance will be *poor*." It appears to us that it is more natural for operators and experts to express workloads, whether mental or physical in "imprecise" verbal terms such as "poor," "low," "moderate," "high," etcetera, than with arbitrary numerical scales.

Vidulich, M. A. and Tsang, P. S. (1985). Techniques of subjective workload assessment: A comparison of two methodologies. *Proceedings of the Third Biannual Symposium on Aviation Psychology* (pp. 239-246). Columbus: Ohio State University.

Subjective assessments of workload are becoming increasingly important to the assessment of new systems. Over the years a number of methodologies have been suggested for collecting these assessments. Two methods were compared in this investigation: The first method, the Subjective Workload Assessment Technique (SWAT), has developed around the use of conjoint analysis to create true interval scales. The second method, under development at NASA, utilizes subject-generated weights in creating a weighted overall workload score from a set of bipolar ratings. Both methods were used in a laboratory experiment involving rating a number of single- and dual-task trials of compensatory tracking and/or a spatial transformation task. The preliminary results comparing the two techniques' overall correlation and responsiveness to single-task difficulty manipulations were discussed. A striking similarity was found between the two techniques' performance and was interpreted as evidence for the robustness of the subjective experience of workload.

Vidulich, M. A. and Tsang, P. S. (1985). Assessing subjective workload assessment: A comparison of SWAT and the NASA-Bipolar methods. *Proceedings of the of the Human Factors Society 29th Annual Meeting* (pp. 71-75). Santa Monica, CA: Human Factors Society.

Subjective assessments of workload are becoming increasingly important in the evaluation of new systems. Two popular methods were compared in the present investigation: (1) the Subjective Workload Assessment Technique (SWAT) which was developed around the use of conjoint analysis to create interval scales, and (2) a technique under development at NASA that utilizes an individually weighted workload score from a set of nine bipolar ratings. Both methods were applied in a laboratory experiment that required rating a number of single- and dual-task trials of tracking and/or a spatial transformation task. The dual transformation-tracking task results were reviewed. The results for the two assessment techniques were remarkably similar, indicating that the subjective experience of workload is sufficiently robust to be resistant to variations in the measuring technique. Also, both subjective assessment techniques were successful in measuring the differences in task difficulty as indicated by a multivariate analysis of performance. Finally, the specific strengths and weaknesses of each assessment technique were reviewed.

Vidulich, M. A. and Wickens, C. D. (1985). Causes of dissociation between subjective workload measures and performance: Caveats for the use of subjective assessments. *Proceedings of the Third Biannual Symposium on Aviation Psychology* (pp. 223-230). Columbus: Ohio State University.

Dissociations between subjective workload assessments and performance were investigated. The difficulty of a Sternberg memory search task was manipulated by varying stimulus presentation rate, stimulus discernibility, value of good performance, and automaticity of performance. All Sternberg task conditions were performed both alone and concurrently with a tracking task. Bipolar subjective workload assessments were collected. Dissociations between workload and performance were found related to automaticity, presentation rate, and motivation level. The results were interpreted as supporting the hypothesis that the specific cognitive processes responsible for subjective assessments can differ from those responsible for performance. The potential contamination these dissociations could inflict on operational workload assessments is discussed.

Vidulich M. A. and Wickens, C. D. (1985). Stimulus-Central-Processing-Response compatibility guidelines for the optimal use of speech technology. *Behavior, Research Methods, Instruments, & Computers*, 17(2), 243-249.

With the emergence of speech technology as a viable display/control alternative, the question of guidelines is of importance. Stimulus-central processing-response (S-C-R) compatibility is proposed as a preliminary set of guidelines. S-C-R compatibility makes a two-part set of predictions about the best input/output (I/O) configuration for a task on the basis of the type of central processing that the task requires. For tasks with predominantly spatial central processing demands, the best I/O configuration is predicted to be visual/manual. For tasks with predominantly verbal central processing demands, the best I/O configuration is expected to be auditory/speech. A series of three experiments testing these predictions is reviewed. The results are interpreted as supporting the concept of S-C-R compatibility.

Vincente, K. J., Jarcew, M., and Moray, N. (1985). *An Investigation of the Mental Workload Associated with Skill-Based Behavior*. Working paper, # 85-3. Toronto, Canada: University of Toronto.

A pursuit tracking task with preview was used to study the mental workload associated with skill-based behavior. The task was designed to simulate a hovercraft traveling down a river. The track width, representing the permissible error, and the noise distribution, representing the turbulence, were manipulated in a 3 x 3 factorial design. Performance, subjective, and physiological workload measures were adopted. The measures taken were the time to complete a run, total number of times the bank of the river was hit, subjective estimates of difficulty, subjective estimates of effort, and the amplitude of the 0.1 Hz component of sinusarrhythmia. Subjects became more skilled with practice, to a point where their performance stabilized. The subjective ratings consistently decreased with practice, indicating a reduction in subjective load. The physiological measure of workload indicated a decrease in effort with practice across all configurations as well as suggesting an inverted "U" relationship between workload and performance. However, the subjective ratings indicted a linear relationship between workload and performance. A second study was proposed to resolve this conflict, as well as to further test the sensitivity of the physiological workload measure.

Wickens, D. D., Moody, M. J., and Vidulich, M. A. (1985). Retrieval time as a function of memory set size, type of probes, and interference in recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11(1), 154-164.

This research extends the investigation of Wickens, Moody, and Dow (1981) on retrieval time and its characteristics using an adoption of the Donders-Sternberg paradigm in primary and secondary memory. The two experiments were centered around the earlier finding that retrieval time (primary memory [PM] performance subtracted from secondary memory [SM] performance) was independent of memory set size. Experiment 1 repeated Wickens et al.'s previous research but added a negative probe of a taxonomic category different from that of the other negative probe and from the categorically homogeneous memory set itself. Although the out-of-category probe produced a much flatter slope than the other probes, retrieval time (SM-PM) and retrieval characteristics did not differ. As in the Wickens et al. (1981) experiment, interference effects were found only in secondary memory. Experiment 2 used memory sets of one, two, and four items with a consonant vocabulary and again found retrieval time to be independent of set size, retrieval time of the one-item set being approximately equal to that of the four-item set. This implies that a single-item set is retrieved like a plural-item set--namely, by first retrieving a pointer to the list, rather than by direct access to the item itself.

Wickens, C. D. and Yeh, Y. -Y. (1985). POCs and performance decrements: A reply to Kantowitz and Weldon. *Human Factors*, 27(5), 549-554.

This paper responds to some of the criticisms presented by Kantowitz and Weldon (1985) that have been directed toward the methodology used in a 1981 article by Wickens, Mountford, and Schreiner. We state here that some of their criticisms are valid. A performance operating characteristic (POC) cannot be derived from a single point in a POC space, and therefore resource competition cannot be separated from concurrence cost as a source of task interference. However, we also note that the primary issue of importance to system designers--how to compare interference between different tasks--is

not answered by Kantowitz and Weldon's critique. That issue requires that some technique for standardizing performance decrements across tasks be assumed. Two alternate techniques for standardizing are described in the present paper.

Wickens, C. D., Yeh, Y. -Y., Fuld, R., and Hodel, S. (1985). A comparison of operator performance in manual and automated versions of a dynamic decision-making task. *Proceedings of the of the Human Factors Society 29th Annual Meeting* (pp. 1089-1091). Santa Monica, CA: Human Factors Society.

Two dynamic decision tasks have been designed to investigate operator behavior in manual and automated systems. Rationale for the study and the nature of the tasks are detailed.

Wierwille, W. W., Rahimi, M., Casali, J. G. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors*, 27(5), 489-502.

As aircraft and other systems become more automated, a shift is occurring in human operator participation in these systems. This shift is away from manual control and toward activities that tap the higher mental functioning of human operators. Therefore, an experiment was performed in a moving-base flight simulator to assess mediational (cognitive) workload measurement. Specifically, 16 workload-estimation techniques were evaluated as to their sensitivity and intrusion in a flight task emphasizing mediational behavior. Task loading, using navigation problems presented on a display, was treated as an independent variable, and workload-measure values were treated as dependent variables. Results indicate that two mediational task measures, two rating scale measures, time estimation, and two eye behavior measures were reliably sensitive to mediational loading. The time estimation measure did, however, intrude on mediational task performance. Several of the remaining measures were completely insensitive to mediational load.

Yeh, Y. -Y. and Wickens, C. D. (1985). *An Investigation of the Dissociation Between Subjective Measures of Mental Workload and Performance* (Technical Report EPL-84-1/NASA-84-1). Urbana-Champaign: University of Illinois, Engineering-Psychology Research Laboratory.

This report describes research conducted during the first years under a contract from NASA Ames Research Center; Dr. Sandra Hart was the technical monitor. The report addresses the *dissociation* between subjective measures of mental workload and performance. Three generic factors are identified that will drive subjective workload upward more than drive performance downward: perceptual (versus response) load, and increased number of tasks, and better data quality. One factor, resource competition, is assumed to drive performance more than subjective workload. The theory of dissociation is tested in three experiments that employ different variations and combinations of three different tasks (tracking, memory search, and a simulated air traffic control task). The predictions of the theory are generally supported by the data. In addition, various subjective scales of mental workload are tested across the experiments. The correlations between these scales and multi-dimensional scaling data are used to help interpret the hidden cognitive structure of task difficulty.

Yeh, Y. -Y. and Wickens, C. D. (1985). *The Dissociation of Subjective Measures of Mental Workload and Performance* (Technical Report EPL-84-2/NASA-84-2). Urbana-Champaign: University of Illinois, Engineering-Psychology Research Laboratory.

A dissociation between performance and subjective workload measures occurs when two task configurations are compared and one shows better performance, but is perceived as subjectively more difficult than the other. The dissociation phenomenon was investigated in the theoretical framework of the multiple-resources model. Even though the underlying structure of subjective workload strongly corresponds with the structure of processing resources, subjective measures do not preserve the vector characteristics in the multidimensional space described by the model. A theory of dissociation (Wickens and Yeh, 1983) was proposed to locate the sources that may produce dissociation between the two workload measures. According to the theory, performance is effected by every aspect of processing whereas subjective workload is sensitive to the amount of aggregate resource investment and is dominated by the demands on the perceptual/central resources. The proposed theory was tested in three experiments, employing different combinations of a tracking task and a Sternberg memory search task.

In support of the theory, the results showed that performance improved but subjective workload was elevated with an increasing amount of resource investment. Furthermore, subjective workload, being affected by the aggregate demands, was not as sensitive as was performance to differences in the amount of resource competition between two tasks. The demand on perceptual/central resources was found to be the most salient component of subjective workload from both the multidimensional analysis of hidden structure and the regression analysis of the underlying components. Dissociation occurred when the demand on this component was increased by the number of concurrent tasks or by the number of display elements. However, in contrast to the prediction, demands on response resources were weighted in subjective introspection as much as demands on perceptual/central resources. The implications of these results for workload practitioners are described.

Yeh, Y. -Y., Wickens, C. D., and Hart, S. G. (1985). The effect of varying task difficulty on subjective workload. *Proceedings of the of the Human Factors Society 29th Annual Meeting* (pp. 765-769). Santa Monica, CA: Human Factors Society.

The goal of the present study was to determine whether or not retrospective workload ratings would reflect the average demands of the entire block of trials or whether one segment within the block would have more weight in determining the magnitudes of ratings than another. Performance data within a block of trials almost perfectly reflected the different task difficulty manipulations: reaction times (but not movement times) reflected variations in the difficulty of the more cognitive response selection component whereas movement times (but not reaction times) reflected variations in the difficulty of the response execution component. Subjective ratings consistently reflected the combined demands of both task components averaged across levels of difficulty even when their levels of difficulty were varied within the block of trials. In every case, it appeared that all of the trials within a block were given equal weight in the composite subjective evaluation. These results suggest that subjective workload is not a specific retrieval of experiences heeded in working memory. Rather, it may reflect the experiences of an ongoing integration process.

APPENDIX H:
RESEARCH PAPERS AND PUBLICATIONS

- 1986 -

Adie, P. and Drascic, D. (1986). *Validation of a Mental Workload Measurement Device*. Unpublished M.S. Thesis. Toronto, Canada: University of Toronto.

The main objective of our thesis was to determine whether the Heart Rate Variability 0.1 Hz Power Spectrum Analyzer can be used as an effective measure of human mental workload. A strong qualitative relationship, and a fair quantitative relationship was found between expected mental workload and the output of the device.

Berg, S. L. and Sheridan, T. S. (1986). The impact of physical and mental tasks on pilot mental workload. *Proceedings of the 21st Annual Conference on Manual Control (NASA CP-2428)*. (pp. 6.1-6.26). Washington, DC: National Aeronautics and Space Administration.

Seven instrument-rated pilots with a wide range of backgrounds and experience levels flew four different scenarios on a fixed-base simulator. The Baseline scenario was the simplest of the four and had few mental and physical tasks. An Activity scenario had many physical but few mental tasks. The Planning scenario had few physical and many mental tasks. A Combined scenario had high mental and physical task loads. The magnitude of each pilot's altitude and airspeed deviations was measured, subjective workload ratings were recorded, and the degree of pilot compliance with assigned memory/planning tasks was noted. Mental and physical performance was a strong function of the manual activity level, but not influenced by the mental task load. High manual task loads resulted in a large percentage of mental errors even under low mental task loads. Although all the pilots gave similar subjective ratings when the manual task load was high, subjective ratings showed greater individual differences with high mental task loads. Altitude or airspeed deviations and subjective ratings were most closely correlated when the total task load was very high. Although airspeed deviations, altitude deviations, and subjective workload ratings were similar for both low experience and high experience pilots, at very high total task loads, mental performance was much lower for the low-experience pilots.

Bortolussi, M. R., Kantowitz, B. H. and Hart, S. G. (1986). Measuring pilot workload in a motion base trainer. *Applied Ergonomics*, 17 (4), 278-283.

Various techniques have been developed to predict and measure pilot workload. This simulation was conducted in order to compare four widely used methods: a visual two- and four-choice reaction time task, time production, retrospective multidimensional subjective ratings, and in-flight verbal workload estimates. Two scenarios with different levels of difficulty determined by preliminary research were designed to test these techniques. The insertion of the secondary tasks did not significantly affect flight performance. All four techniques were able to distinguish between the overall levels of scenario complexity. In addition, the three secondary tasks and workload ratings obtained in-flight were generally able to distinguish among levels of difficulty for different segments within the scenarios.

Casali, J. G. and Wierwille, W. W. (1986). On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics*, 27 (10), 1033-1050.

A flight-simulator-based study was conducted to examine fourteen distinct mental workload estimation measures, including opinion, secondary task, physiological, and primary task measures. Both the relative sensitivity of the measures to changes in mental workload and the differential intrusion of the changes on primary tasks performance were assessed. The flight task was varied in difficulty by manipulation of the presentation rate and complexity of a hazard-perception task that required each of 48 licensed pilots to rely heavily on their perceptual abilities. Three rating scales (Modified Cooper-Harper, Multi-descriptor, and Workload-Compensation-Interference/Technical Effectiveness), two secondary task measures (time estimation and tapping regularity), one physiological measure (danger-condition response time) were reliable indicants of workload changes. Recommendations for applying the workload measures are presented.

Casper, P. A., Shively, R. J., and Hart, S. G. (1986). Workload Consultant: A microprocessor-based system for selecting workload assessment procedures. *Proceedings of the IEEE/Meeting on Systems, Man and Cybernetics*, International Conference (pp. 1054-1059). Piscataway, New Jersey: IEEE Service Center.

Recent years have seen a deepening interest in the measurement of human operator workload. However, not all persons involved in the design and production of human-machine systems are educated in the rigors of workload measurement and the currently available techniques. Furthermore, as in most areas of expertise, there aren't enough human "experts" to go around. The present paper describes an "expert" system, created at the NASA Ames Research Center, that was designed to provide decision support for persons interested in assessing operator workload. The system is based on current research in the field of workload measurement and is flexible enough to allow for incorporation of new knowledge as it is empirically validated.

Chignell, M. H., Hancock, P. A., Smith, P. J., and Shute, S. J. (1986). Information retrieval: An intelligent interface perspective. *Proceedings of the IEEE Meeting on Systems, Man, and Cybernetics*, Vol. 1. (pp. 372-377). Piscataway, NJ: IEEE Service Center.

The intelligent interface is seen as a third entity mediating communication between human and machine. While there is general agreement on what the goals of an intelligent interface should be, detailed specification of how to build and operate such an interface is lacking at present. Information retrieval represents a compelling illustration of the problem of translation between human (end user) and machine (database). A human search intermediary often acts as an intelligent interface between the end user and the database. Using information retrieval as a prototypical example, we outline a model of the intelligent interface based on an analysis of the role of the human search intermediary.

Eisen, P. and Money, L. (1986). *Fuzzy Set Analysis of Mental Workload*. Unpublished M.S. Thesis. Toronto, Canada: University of Toronto.

Subjective measurements of task difficulty, in the form of fuzzy set membership estimates were gathered for a range of tasks. Rasmussen's model of human behavior was used to determine the basis of the tasks. Specifically, several levels of skill based and rule based tasks were employed. Results indicate that the perceived task difficulty of the combination of one skill based and one rule based task can be predicted, knowing the perceived difficulty of the tasks individually. Further investigations, using fuzzy set calculus to evaluate mental workload, are recommended.

Fuld, R. and Wickens, C. D. (1986). *An Investigation of Operator Performance in Manual and Automated Versions of a Visual Monitoring Task*. (EPL-86-/NASA-86-1). Champaign: University of Illinois, Engineering-Psychology Research Laboratory.

This experiment compared 7 subjects' ability to detect violations of optimal performance, and respond to infrequent malfunctions, when either performing a "customer assignment task." or monitoring an automated system performing the same task. In the task, randomly arriving customers were to be assigned to one of three cues with the shortest expected wait. Malfunctions occurred when a cue stopped processing customers, and the subject was then required to reassign the customers. This event occurred only once in each mode of participation. The results revealed that subjects were more accurate at detecting non-optimal cue assignments in the automatic mode. In the manual mode they were conservative in reporting their own departures from optimal assignments. However, subjects intervened more rapidly following a failure in the manual mode than in the automatic mode. The results are discussed in terms of models of manual-automatic differences, and in terms of shortenings of the present paradigm.

Goguen, J., Linde, C. and Murphy, M. (1986) *Crew Communications as a Factor in Aviation Accidents*. (NASA TM-88354). Washington, DC: National Aeronautics and Space Administration.

A method for the detailed analysis of within-crew communications is developed and applied in formulating and testing several hypotheses about the basic structure of the aircrew communication process. Planning and explanation are shown to be well-structured discourse types, described by formal rules. These formal rules are integrated with those describing the other most important discourse type within the cockpit: the command-and-control speech act chain. Command-and-control discourse is described as a sequence of speech acts for making requests (including orders and suggestions), for making reports, for supporting or challenging statements, and for acknowledging previous speech acts. Mitigation level, a linguistic indication of indirectness and tentativeness in speech, was an important variable in several hypotheses. Testing these hypotheses showed that the speech of subordinates is more mitigated than the speech of superiors, that the speech of all crewmembers is less mitigated when they know that they are in either a problem or emergency situation, and that mitigation is a factor in failures of crewmembers to initiate discussion of new topics or have suggestions ratified by the captain. The test results also indicated that planning and explanation are more frequently performed by captains than by other crewmembers, are done more during crew-recognized problems, and are done less during crew-recognized emergencies.

Gopher, D. (1986). *Assessment of Workload in Engineering Systems*. (Final scientific report on NASA Grant NAGW - 494) Haifa, Israel: The Technion - Israel Institute of Technology.

This document summarizes the scientific work conducted under NASA grant NAGW - 494, entitled "Assessment of workload in engineering systems" awarded to Daniel Gopher, at the Technion - Israel Institute of Technology. The objectives of this work were twofold: 1) to review the theoretical and empirical work in the problem area of workload with an attempt to develop a theoretical framework that can serve workers in this field; and 2) to conduct experimental work to enhance our understanding of the nature of subjective measures of workload, provide methods for their measurement and recommendations for their application.

Gopher, D. (1986). In defense of resources: On structures, energies, pools, and the allocation of attention. In R. Hockey, A. Gaillard, and M. Coles (Eds.), *Energetics and Human Information Processing* (pp. 353-372). The Netherlands: Nijhoff.

Current theoretical thinking in cognitive psychology is dominated by the computer metaphor which tends to emphasize a detailed analysis of computational processes and neglects energetical considerations. A typical example is the debate between structural and resource interpretations of the limitations of the human processing system. The present chapter reviews the main aspects of this debate and defends the theoretical view that resources are hypothetical constructs representing aggregates of elementary processing units. The approach developed is of multiple resources, within which energetical sources are linked to and influence the efficiency of specific processing structures. While the influence of structural factors is not denied, resource availability and modulations in the intensity of processing are argued to be major contributors to the efficiency of the system. When contrasting structural and energetical interpretations, a critical test is the substantiation of the resource scarcity assumption. The chapter discusses different senses of resource scarcity, along with evidence in their support and experimental paradigms to test them. Special emphasis is placed on the role of collaborative efforts from physiological and behavioral research in uncovering the operation rules of the human processing system.

Gopher, D. and Donchin, E. (1986). Workload -- An examination of the concept. In K. Boff and L. Kauffman (Eds.), *Handbook of Perception and Human Performance* (pp. 41-1 - 41-49). New York: Wiley and Sons.

This chapter represents a theoretical examination of the multidimensional, multifaceted concept of workload. Due to the complexity of the construct, no single measure is capable of capturing all relevant aspects, nor may multiple measures covary within a single task. The discussion was concerned with clarifying the nature of the dimensions along which workload varies to explicate the attributes that should be considered in the selection of a measurement procedure. The primary thesis is that workload assessment focuses on measuring the processing and response limitations of the human information processing system which are revealed through the interactions between an operator and the assigned tasks. The nature of the limitations were considered on two levels: (1) the more theoretical level (in which the invariant, open loop properties of the human processing system were examined), and (2) a more practical level (in which workload was characterized, at any instant, as the joint, closed loop property of the human and the assigned task). In general, the focus of the theoretical discussions emphasized the close affinity between the study of

workload and attention, with an additional discussion of the energetical and structural characteristics of the central processor. The recommendation was made that measurement procedures should encompass both conscious and nonconscious processing activities; a detailed task analysis should be performed to uncover the major components of the task, followed by a battery of performance-based measures designed to evaluate the load on each component.

Gopher, D., Weil, M., and Siegel, D. (1986). Is it only a game? Using videogames as surrogate instructors for the training of complex skills. *Proceedings of the 1986 IEEE International Conference on Man, Machine, and Cybernetics* (pp. 1060-1064). Piscataway, New Jersey: IEEE Service Center.

Modern computer games are complex, interesting, and demanding. The present work investigates the possibility of using them for the training of complex skills. To do so, learning strategies should be formalized and incorporated in the game routines. The characteristics of expert performance are discussed and a training approach based upon emphasis manipulation of task components is proposed. This approach has been applied to the training of subjects in a highly demanding computer game. It led, in a short period, to a substantial improvement in the performance ability of trained subjects, as compared with a group which played the game for an equal duration without training.

Hancock, P. A. (1986). On the use of time: The irreplaceable resource. In H. Hendrick and O. Brown (Eds.), *Proceedings of the Second International Symposium on Organizational Design and Management* (pp. 83-89). The Netherlands: North Holland Press.

This paper reviews differing aspects of the role of time in the operation of organizational systems. Such facets can range from the understanding of time as a unique functional resource of the organization to the personal use by key management individuals. It is this latter behavioral perspective that is the concern of this work. Our present operational perception of time as an immalleable and uncontrollable progression is questioned. Personal time utility analysis is addressed as a potential avenue through which to maximize temporal efficiency.

Hancock, P. A. (1986). The role of temporal factors in workload prediction. *Proceedings of the IEEE Meeting on Systems, Man, and Cybernetics*, Vol. 2 (pp. 1049-1053). Piscataway, NJ: IEEE Service Center.

In examining the role of time in mental workload, this paper presents a different perspective from which to view the problem of assessment. Workload is plotted in three dimensions, whose axes represent effective *time* for action, perceived *distance* from desired goal state, and level of *effort* required to achieve such a goal. This representation allows the generation of *isodynamic workload contours* which incorporate the factor of operator competence. A simple physical analogy for this representation indicates an avenue toward quantification and, subsequently, the potential for useful workload prediction.

Hancock, P. A. (1986). Stress and adaptability. In R. Hockey, A. Gaillard, and M. Coles (Eds.) *Energetics and Information Processing*. The Netherlands: Martinus Nijhoff.

One of the mandates given to the Workshop C group was the exploration of definitions of the concept of stress and how stress may act to impact various human capabilities. Due to time constraints, it was not possible to address this important and broad issue in detail, although an initial consensus was found in support of the position adopted by Lazarus. The purpose of the present paper is to elaborate upon this little-explored theme through examination of a recent position which bears upon this problem and, specifically, to indicate how insights gained during the meeting have acted to enhance this latter perspective.

Hancock, P. A. (1986). Stress, information-flow, and adaptability in individuals and collective organizational systems. In H. Hendrick and O. Brown (Eds.), *Proceedings of the Second International Symposium on Organizational Design and Management*. (pp. 293-296). The Netherlands: North Holland Press.

The central theme of this brief paper is the comparison of the commonalities between the characteristics of individuals and the collective organizational structures within which they operate. Each entity collects, filters, and sequentially transduces information in order to effect optimal adaptive action. Information in this context is distinguished along two axes which represent flow-rate and utility. Each entity seeks to locate itself within this two-dimensional information space at a point which maximizes task-related output at the least energetical cost consistent with successful performance. The transition between normal and failure modes of operation are compared across the human and the organization and can be represented as either gradual degradation or rapid dissolution of adaptability that can be described through the tenets of Catastrophe theory. A compromise between hierarchical, heterarchical, and holarchical structures is posed as one which optimizes response to stress intrinsic to environmental inputs. Manners in which the human and the organization utilize such structures are explored briefly.

Hancock, P. A. and Chignell, M. H. (1986). Toward a theory of mental workload: Stress and adaptability in human-machine systems. *Proceedings of the IEEE Meeting on Systems, Man, and Cybernetics*, Vol. 1 (pp. 378-383). Piscataway, NJ: IEEE Service Center.

In light of the present difficulties in assessment, there is a pressing need for a general theory of mental workload (MWL). This paper explores a view of the task as a stress and highlights the commonalities between mental workload and the psychological and physiological reactions of an operator to stress in general. In the absence of a normative theory of mental work, mental workload is defined as an organismic response to the requirements of the task. The role of mental workload assessment within an adaptive human-machine system is outlined, and the use of changes in cognitive functioning to predict subsequent failure in human performance is recommended.

Hart, S. G. (1986). Background, description, and application of the NASA Task Load Index (TLX). *Proceedings of the Department of Defense Human Factors Engineering Technical Advisory Group*. Dayton, OH, September.

The NASA-Task Load Index (NASA-TLX) is the product of a multi-year research effort devoted to identifying the dimensions of subjective workload experiences, developing methods of quantifying such experiences, and accounting for differences in the sources of workload that are relevant to different raters and for different tasks. Each stage of the scale development process are reviewed and the results of validation studies are presented.

Hart, S. G. (1986). The relationship between workload and training: An introduction. *Proceeding of the Human Factors Society 30th Annual Meeting*. Vol. 2 (pp. 1116-1120). Santa Monica, CA: Human Factors Society.

This paper reviews the relationships among workload, performance, and training. It is intended to serve as an introduction for the remaining papers in this symposium. Its goal is to introduce the concepts of workload and training and to suggest how they may be related. It suggests some of the practical and theoretical benefits to be derived from their joint consideration: training effectiveness can be improved by monitoring trainee workload and the reliability of workload predictions, and measures can be improved by identifying and controlling the training levels of experimental subjects.

Hart, S. G. (1986). Theory and measurement of human workload. In J. Zeidner (Ed.), *Human Productivity Enhancement: Training and Human Factors in System Design*, Vol. 1 (pp. 396-456). New York: Praeger.

The goal of this chapter is to define human workload, what influences it, how it is measured, and why it is of theoretical and practical concern. The first section reviews typical definitions and motives for measuring and predicting workload. A structure is proposed to relate and integrate many of the factors that create or influence it (e. g., the demands imposed in a man-machine system, its response to them, and the subjective experiences of operators). A third section describes five types of assessment and predictive methodologies: (1) subjective ratings, (2) primary task performance, (3) secondary task performance, (4) physiological recordings, and (5) analytic procedures. Finally, the selection and application of appropriate tools to predict or assess imposed workload, system performance and behavior, or operator experience are considered.

Hart, S. G. (1986) Workload Studies at NASA Ames Research Center. *Practical Assessment of Pilot Workload*. Presented at the Meeting sponsored by RAE, AGARD, and Cranfield Institute of Technology. Cranfield, England.

In 1981, NASA formed a Workload Assessment Program to address the many unresolved issues in this increasingly important field. The goal was to merge the theoretical information available from academia with the practical requirements of industrial and government organizations to develop a comprehensive workload definition and a set of practically useful measures and predictors. Throughout the program, well-controlled laboratory experiments provided answers to specific questions and theoretical issues while

simulation and inflight research provided verification that the results were valid and meaningful in an operational environment. The first phase of the program was devoted to understanding the factors that influence pilot workload, evaluating existing assessment techniques, and developing new techniques. The work was accomplished by an active interaction between government laboratories, industry research groups, and universities. The second phase of the program, which is underway, is devoted to completing a computer model for workload prediction, developing workload criteria (e.g., how much is "too much" or "too little"), and investigating the relationship between workload, training, and performance. On a continuing basis, the methods and theories developed by participants in the program have been applied to specific operational and design problems at the request of other government agencies and industry. This report summarizes the research conducted during the first phase of the program and describes the results obtained in several simulator and inflight applications.

Hart, S. G. (1986). *Workload in Complex Systems*. Presentation prepared for the Symposium on the U.S. Army Key Operational Capabilities. Carlisle, PA: The U.S. Army War College, May 12-15.

Workload is an important, integrative concept that determines the ability of the human users of the advanced systems to accomplish mission requirements. Factors contributing to the workload imposed on an individual include task demands, temporal constraints and schedules, equipment provided, environmental factors, and operator skills and training. Workload can be measured with some degree of accuracy when the appropriate method of measurement is selected in order to achieve the goal of the analysis. By describing available information into more predictive models, the workload-impact decisions can be included in the design process to insure acceptable workload margins. When performance failures do occur within systems, however, workload modifications may be called upon for solutions. In these ways systems may be designed which have workload levels that are acceptable to the human operators and which achieve the target levels of performance.

Hart, S. G., Shively, R. J., Vidulich, M. A., and Miller, R. C. (1986). The effects of stimulus modality and task integrality: Predicting dual-task performance and workload from single task levels. *Proceedings of the 21st Annual Conference on Manual Control* (NASA CP-2428). (pp. 5.1-5.18). Washington, DC: National Aeronautics and Space Administration.

The influence of stimulus modality and task difficulty on workload and performance was investigated in the current study. The goal was to quantify the "cost" (in terms of response time and experienced workload) incurred when essentially serial task components shared common elements (e. g., the response to one initiated the other) which could be accomplished in parallel. The experimental tasks were based on the "Fittsberg" paradigm; the solution to a SternBERG-type memory task determines which of two identical FITTS targets are acquired. Previous research suggested that such functionally integrated "dual" tasks are performed with substantially less workload and faster response times than would be predicted by summing single-task components when both are presented in the same stimulus modality (visual). In the current study, the physical integration of task elements was varied (although their functional relationship remained the same) to determine whether dual-task facilitation would persist if task components were presented in different sensory modalities. Again, it was found that the cost of performing the two-stage task was considerably less than the sum of component single-task levels when both were presented visually. Less facilitation was found when task elements were presented in different

sensory modalities. These results suggest the importance of distinguishing between concurrent tasks that compete for limited resources from those that beneficially share common resources when selecting the stimulus modalities for information displays.

Haworth, L. A., Bivens, C. C., and Shively, R. J. (1986). An investigation of single-piloted advanced cockpit and control configurations for nap-of-the-earth helicopter combat mission tasks. *Proceedings of the 1986 Meeting of the American Helicopter Society* (pp. 657-682). Washington, DC.

A two-phase handling qualities and pilot workload investigation of single-pilot operation in the combat nap-of-the-Earth (NOE) environment was started by the Aeroflightdynamics Directorate in October 1985. Phase one of the investigation was conducted in cooperation with the NASA Ames Research Center on the NASA Vertical Motion System (VMS) simulator, using the Advanced Digital Optical Control System (ADOCS) laws and a glass cockpit. Handling Quality Ratings (HQR) and workload ratings were recorded for NOE flight-task maneuvers during single-pilot and "dual"-pilot operation. Control automation/augmentation was varied to record differences between configurations for dual and single-pilot operation. Only one control system configuration investigated was rated satisfactory for single-pilot NOE flight due to increased attentional demands placed on the pilot.

Kantowitz, B. H. (1986). *A Theoretical Approach to Measuring Pilot Workload* (Final Report for NASA Grant NCC 2-228). Washington, DC: National Aeronautics and Space Administration.

This final report for Cooperative Agreement NCC 2-228 covers the period January 1, 1983, through December 31, 1985. The NASA Technical Officer was S. G. Hart, Ames Research Center, Man-Vehicle Systems Research Division. The work accomplished during this period can be grouped into three categories. First, and most important, are theoretical advances aimed at integrating the concepts of attention and workload. Second, are empirical studies, primarily performed at Ames Research Center, that studied objective measures of pilot workload. Third, are systems software written in West Lafayette to allow data collection and analysis of workload experiments. This Cooperative Agreement has produced nine publications, three book chapters, one technical report, and three C-language software systems. A list of publications and book chapters is attached and the Appendix includes copies of most of this work.

Linde, C., Goguen, J., and Devenish, L. (1986). *Evaluation Criteria and Initial Review of Communications Training Program* (Progress Report for NASA NAS2-12379). Moffett Field, CA: NASA Ames Research Center, Aerospace Human Factors Research Division.

This report is the first of three in a project studying communications training for civilian and military aviation personnel, including multiperson crews, single-pilot fixed-wing aircraft teams, commercial aviation crews, and helicopter teams. It is well known that a high percentage of aviation accidents are caused wholly or in part by problems of communications and human resources management. In a number of commercial aviation accidents, the NTSB has recommended assertiveness training for crew members as one way to reduce the number of such accidents. Existing and ongoing research at NASA attempts

to determine more exactly the nature of communication problems which lead to accidents. The current project focuses on available training programs and techniques that would help apply the results of such research to the practical problem of training crews to communicate better. This report offers lists of criteria for evaluating the applicability of given training programs in the aviation context. It then applies these criteria to United Airlines Resources Management Training, and to a number of commercially available general purpose training programs. Finally, it discusses a range of existing theories of communication which appear to have some relevance to effective training. Later reports will consider the most immediately applicable communications and training theories in more depth, and provide a critical assessment of their actual applicability to the aviation context.

Lyman, J. (1986). *Modified Petri Net Model Sensitivity to Workload Manipulations* (Final Technical Report for NASA Grant NAG 2-216). Washington, DC: National Aeronautics and Space Administration.

The purpose of this research is to investigate modified Petri nets (MPNs) as a workload modeling tool. This paper describes the results of an exploratory study of the sensitivity of MPNs to workload manipulations in a dual task. The results of the canonical correlation indicated that MPN model of the experimental task represented the task components that influenced subjective workload. Thus, the goal of this experiment was achieved by this demonstration that the MPN model was sensitive to workload changes. The next stage of this research will involve generating a classification scheme that will group events and activities that are similar in their contribution to task workload. Workload values for each class of events and activities can then be derived. This will allow testing of MPN model simulations for their prediction capability of the workload of a task.

Mane, A. and Wickens, C. D. (1986). The effects of task difficulty and workload on training. *Proceedings of the Human Factors Society 30th Annual Meeting*, Vol. 2 (pp. 1124-1127). Santa Monica, CA: Human Factors Society.

We propose four hypotheses regarding the possible effect of workload and task difficulty on training: (1) increased levels of task difficulty will facilitate learning to the extent that these increases are (a) resource loading and (b) intrinsic to the component task to be learned. (2) Decrease of task difficulty will facilitate learning to the extent that these decreases (a) reduce the resource load and (b) are extrinsic of the component task to be learned. (3) The learner's tendency to conserve resources may lead to the adoption of undesirable, short-term, low resource strategies early in training. (4) The effect of changes in resource demand on learning will depend upon the similarity of the resource whose demand is changed to the resource involved in learning.

Miller, R. C., Bortolussi, M. R., and Hart, S. G. (1986). Evaluating the subjective workload of directional orientation tasks with varying displays. *Fifth Aerospace Behavioral Engineering Technology Conference Proceedings "Human Integration Technology: The Cornerstone for Enhancing Human Performance"*. (pp. 135-138). Warrendale, PA: Society of Automotive Engineers.

An experiment was conducted to investigate the impact of various flight-related tasks on the workload imposed by the requirement to compute new headings, course changes and reciprocal headings. Nine instrument-rated pilots were presented with a series of heading-change tasks in a laboratory setting and in a single-place instrument trainer. Two levels of difficulty of each of three tasks were presented verbally (numeric values embedded in simple commands), spatially (headings were depicted on a graphically drawn compass) and combined (each of the previous displays were given simultaneously). In the instrument-trainer setting problems were presented orally by one of the experimenters and no effort was made to manipulate display types. Performance was measured by evaluating the speed (response times) and accuracy (percent correct and time outs) of the responses. The workload experienced by the pilots under each experimental condition was determined by responses to a standard set of bipolar rating scales. These subjective measures reflected the differences between levels of difficulty and types of tasks, but were generally insensitive to the manipulation of display type. The performance measures, however, displayed significant differences for all manipulations. Problems presented in the combined and alpha display formats, were done significantly faster and with significantly greater accuracy than problems in the compass format alone suggesting that the pilots were primarily using the alpha information contained in the combined display to perform the calculations. Workload ratings for the compass-only laboratory condition and the instrument trainer portion of the study were virtually identical across all conditions.

Moray, N., Eisen, P., Greco, G., Krushelnycky, E., Money, L., Muir, B., Noy, I., Shein, F., Turksen, B., and Waldon, L. (1986). Fuzzy and vector measurement of workload. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (pp. 1040-1043) Piscataway, NJ: IEEE Service Center.

This paper reports two approaches to workload estimation. In the first, membership functions for fuzzy estimates of task difficulty were obtained for skill-based and rule-based tasks and for their combination. The difficulty of the combined task was successfully predicted from the single tasks by an equation combining the membership functions of single-task difficulty. The second approach explored the consequences of the widely held (but practically neglected) belief that workload should be measured as a vector rather than a scalar quantity. The results suggest that while it is difficult to estimate workload it may nonetheless be possible using a vector-matrix measure to match people to tasks.

Moray, N., Turksen, B., Adie, P., Drascic, D., Eisen, P., Krushelnycky, E., Money, L., Schonert, H., and Thornton, C. (1986). Progress in mental workload measurement. *Proceedings of the Human Factors Society 30th Annual Meeting*, Vol. 2. (pp. 1121-1124). Santa Monica, CA: Human Factors Society.

Two new techniques are described, one using subjective, the other physiological data for the measurement of workload in complex tasks. The subjective approach uses fuzzy measurement to analyze and predict the difficulty of combinations of skill-based and rule-based behavior from the difficulty of skill-based behavior and rule-based behavior measured separately. The physiological technique offers an on-line real-time filter for measuring the Mulder signal at 0.1 Hz in the heart rate variability spectrum.

Morris, N. M. and Rouse, W. B. (1986). *Human Operator Response to Error-Likely Situations in Complex Engineering Systems* (Interim Report for Contract No. NAS2-12048). Moffett Field, CA: NASA Ames Research Center.

The experiment reported in this paper is part of a research effort directed at understanding the causes of human error in complex systems. First, a conceptual framework is provided, in which two broad categories of error are discussed: errors of action, or slips, and errors of intention, or mistakes. Conditions in which slips and mistakes might be expected to occur are identified, based on existing theories of human error. Then, the results of an experiment designed to evaluate relationships in the conceptual framework are presented. Subjects in the experiment controlled PLANT under a variety of conditions. Three independent variables were manipulated in the experiment: 1) compatible vs. incompatible keyboard arrangement (expected to affect the occurrence of slips); 2) simple vs. complex PLANT failures (expected to affect the occurrence of mistakes); and 3) self- vs. forced-pacing (a manipulation of imposed load). Ratings of subjective mental effort were obtained from subjects every ten iterations (approximately every 31-40 secs) as they controlled PLANT. A rather complex pattern of results was obtained, in that the three independent variables interacted in a variety of ways in their effects upon subjects' behavior and performance. It was concluded that subjects responded to situations in which errors were likely by trying to reduce the likelihood of error in those situations. Two approaches were taken: 1) controlling the situation by altering their strategies, and 2) controlling themselves by being more careful. The implications of these results for future research are considered.

Mosier, K. L. and Hart, S. G. (1986). Levels of information processing in a Fitts Law task. *Proceedings of the 21st Annual Conference on Manual Control* (NASA CP-2428). (pp. 4.1-4.15). Washington, DC: National Aeronautics and Space Administration.

State-of-the-art flight technology has restructured the task of human operators, decreasing the need for physical and sensory resources, and increasing the quantity of cognitive effort required, changing it qualitatively. Recent technological advances have the most potential for impacting the contemporary pilot in two areas: performance and mental workload. In an environment in which timing is critical, additional cognitive processing can cause performance decrements, and increase a pilot's perception of the mental workload involved. The effects of stimulus processing demands on motor response performance and subjective mental workload are examined in the current study, using different combinations of response selection and target acquisition tasks. The information processing demands of the response selection were varied (e. g., Sternberg memory set tasks, math equations, pattern matching), as was the difficulty of the response execution. Response latency as well as subjective workload ratings varied in accordance with the cognitive complexity of the task. Movement times varied according to the difficulty of the response execution task. Implications in terms of real-world flight situations are discussed.

Murphy, M. and Awe, C. A. (1986). Aircrew coordination and decision-making: A peer review approach. *Proceedings of the 21st Annual Conference on Manual Control* (NASA CP-2428). (pp. 23.1-23.33). Washington DC: National Aeronautics and Space Administration.

Six professionally active, retired captains rated the coordination and decision-making performances of sixteen aircrews while viewing videotapes of a simulated commercial air

transport operation. The videotapes displayed a composite of four views of crewmembers and the cockpit from cameras located inside the simulator. The scenario featured a required diversion and a probable minimum fuel situation. Seven point Likert-type scales were used in rating variables on the basis of a model of crew coordination and decision-making. The variables were based on concepts of, for example, decision difficulty, efficiency, and outcome quality; and leader-subordinate concepts such as person- and task-oriented leader behavior, and competency motivation of subordinate crewmembers. Five front-end variables of the model were, in turn, dependent variables for a hierarchical regression procedure. The variance in safety performance was explained 46% by decision efficiency, command reversal, and decision quality. The variance of decision quality, and alternative substantive dependent variable to safety performance was explained 60% by decision efficiency and the captain's quality of within-crew communications. The variance of decision efficiency, crew coordination, and command reversal were, in turn, explained 78%, 80%, and 60% by small numbers of preceding independent variables. A principle component varimax factor analysis supported the model structure suggested by regression analyses. Crewmembers for this study were diverse with respect to airline origin and recency, or currency on the Boeing 707--the aircraft simulated. Some retired personnel were used. The results should be interpreted accordingly.

NASA Task Load Index (TLX): Computerized Version (1986). Moffett Field, CA: NASA Ames Research Center.

This booklet and the accompanying diskette contain the materials necessary to collect subjective workload assessments with the Computerized Version NASA Task Load Index on IBM PC compatible microcomputers. This procedure for collecting workload ratings was developed by the Human Performance Group at NASA Ames Research Center during a three year research effort that involved more than 40 laboratory, simulation, and inflight experiments. Although the technique is still undergoing evaluation, this package is being distributed to allow other researchers to use it in their own experiments. Comments or suggestions about the procedure would be greatly appreciated. This package is intended to fill a "nuts and bolts" function of describing the procedure. A bibliography provides background information about previous empirical findings and the logic that supports the procedure.

NASA Task Load Index (TLX): Paper and Pencil Package (1986). Moffett Field, CA: NASA Ames Research Center.

This booklet contains the materials necessary to collect subjective workload assessments with the Paper and Pencil Package NASA Task Load Index. This procedure for collecting workload ratings was developed by the Human Performance Group at NASA Ames Research Center during a three year research effort that involved more than 40 laboratory, simulation, and inflight experiments. Although the technique is still undergoing evaluation, this booklet is being distributed to allow other researchers to use it in their own experiments. Comments or suggestions about the procedure would be greatly appreciated. This package is intended to fill a "nuts and bolts" function of describing the procedure. A bibliography provides background information about previous empirical findings and the logic that supports the procedure.

NASA Workload Consultant for Field Evaluation (WC FIELDE) (1986). Moffett Field, CA: NASA Ames Research Center.

WC FIELDE is a microprocessor based system designed to assist users in selecting appropriate workload assessment procedures. It suggests measures, in descending order of utility, based on the users' answers to a variety of questions concerning their specific application. The factors that it takes into account include: the focus of the research question, the research environment, and the facilities that are available. It draws from a data base of widely used measures in proposing alternatives, and provides specific instructions about how to apply many techniques. It was created with EXSYS, a commercially-available rule-based expert system development package. A copy protected version of the program is provided on the diskette. It runs on IBM/PC and IBM/PC compatible machines.

Shively, R. J. (1986). Application of mental workload methodology to human-computer interaction. *Proceedings of the IEEE Meeting on Systems, Man, and Cybernetics*, Vol. 2. (pp. 907-911). Piscataway, NJ: IEEE Service Center.

The evaluation of mental workload has been of major interest to the aerospace industry for some time. However, while human-computer interactions pose many of the same problems that have led to widespread usage of workload evaluation in aerospace, these workload evaluation techniques have not been applied to this environment. The aerospace community has found that the use of mental workload techniques greatly enhances understanding of human-system performance, and it appears that the understanding of human-computer interactions would also be enhanced. The present experiment applies two workload assessment techniques to human-computer interaction. The results lead us to conclude that the use of mental workload assessment techniques will provide additional information to computer system designers and enhance the understanding of the total human-computer environment.

Skipper, J. H., Rieger, C. A. and Wierwille, W. W. (1986). Evaluation of decision-tree rating scales for mental workload estimation. *Ergonomics*, 29 (4), 585-599.

Recent studies suggest that a decision-tree rating scale called the Modified Cooper-Harper (MCH) rating scale is a globally sensitive indicator of change in mental loading. The present study was directed at developing refinements in the scale and at obtaining additional background information. The MCH scale and five design variations of the scale were studied in two independent aircraft-simulator experiments. Aspects studied were the decision-tree structure, the number of categories, the decision sequence and the effects of computer implementation. Results using the rating scales indicate that the MCH scale and its computerized version are generally more consistent than the others. Attendant questionnaire results indicate that pilots base their ratings on the same factors that researchers believe are the important elements of the multidimensional construct of workload.

Staveland, L., Hart, S. G., and Yeh, Y. -Y. (1986). Memory and subjective work load assessment. *Proceedings of the 21st Annual Conference on Manual Control* (NASA CP-2428). (pp. 7.1-7.13). Washington, DC: National Aeronautics and Space Administration.

Recent research suggested subjective introspection of workload is not based upon specific retrieval of information from long-term memory, and only reflects the average workload that is imposed upon the human operator by a particular task. These findings are based upon global ratings of workload for the overall task, suggesting that subjective ratings are limited in ability to retrieve specific details of a task from long-term memory. To clarify the limits memory imposes on subjective workload assessment, the difficulty of task segments was varied and the workload of specified segments was retrospectively rated. The ratings were retrospectively collected on the manipulations of three levels of segment difficulty. Subjects were assigned to one of two memory groups. In the Before group, subjects knew before performing a block of trials which segment to rate. In the After group, subjects did not know which segment to rate until after performing the block of trials. The subjective ratings, RTs, and MTs were compared for within group, and between group differences. Performance measures and subjective evaluations of workload reflected the experimental manipulations. Subjects were sensitive to different difficulty levels, and recalled the average workload of task components. Cuing did not appear to help recall, and memory group differences possibly reflected variations in the groups of subjects, or an additional memory task.

Strayer, D. L. and Kramer, A. F. (1986). Psychophysiological Indices of Automaticity and Attentional Resources. Paper presented at the Society for Psychophysiological Research Conference, Montreal, Canada.

The present study examines the effects of practice and task structure on human performance in single- and dual-task conditions. The development of automatic processing through consistent stimulus-response mapping (CM) is contrasted with controlled processing obtained with variable stimulus-response mapping (VM). Seven subjects received ten sessions of CM and VM practice. Two tasks, one a Sternberg memory search task and the other a step-tracking task, were employed. In dual-task conditions, subjects were instructed to maintain single task performance in the step-tracking task at the expense of performance in the Sternberg task. Two levels of tracking difficulty (first and second order) were used. Three memory set sizes (2, 3, 4) were presented with a probe frame size of 2, resulting in memory loads of 4, 6, and 8. Only data from the first and tenth sessions of the Sternberg task are discussed.

RT increased with memory load for both CM and VM task in session 1. In session 10, RT increased with memory load for the VM task, but was not affected by memory load in the CM task, indicating that superior performance was obtained in the CM task after practice. Effects of memory load produced a pattern of results which suggests that P3 latency may be a more sensitive measure of the development of automaticity than RT or error rate. Because performance (both RT and error rate) was superior in the CM session 10 condition, it implies that less perceptual information is extracted in automatic tasks. This suggests that a type of perceptual automaticity develops during CM training. Evidence for parallel processing in the CM task was obtained by comparing the left and right stimuli presented in the probe frame. The results suggest that if a CM target is presented it "pops out" and is processed in parallel with other information in the visual field. The N2 component of the ERP was found to discriminate between target present and target absent trials in the Sternberg task. It is not influenced by task structure, practice, or concurrent

load, suggesting that a mismatch process operates over a wide range of practice in both automatic and non-automatic tasks.

Townsend, J. T. (1986). *Toward a Dynamic Mathematical Theory of Mental Workload in POPCORN* (Annual Report for NASA Grant NAG 2-307). Washington, DC: National Aeronautics and Space Administration.

The development of a time-dynamic stochastic theory of mental workload capable of describing and explaining performance in the POPCORN task are described. The principles learned from the POPCORN modeling process will be transferable to other skill/workload paradigms. The research strategy can be broken into sub-goal phases: development of a schematized performance model, development of a descriptive algorithm, programming the model on a computer with additional analytic modeling, and empirical testing. The first phase is complete and the second is underway. Initial experiments have been run on the IBM/AT version of POPCORN to provide an initial data base for the model.

Tsang, P. S. (1986). Can pilots time-share better than non-pilots? *Applied Ergonomics*, 17 (4), 284-290.

Time-sharing performance of a group of pilots was compared with that of a group of college students. In a secondary tasks paradigm, both groups were required to perform five dual tasks with various degrees of structural similarity. A higher degree of task interference was observed for the structurally more similar task pairs. The data were consistent with the results from previous research and support the concept of multiple resources. Although the pilots appeared to be more efficient in one of the dual-task conditions, evidence for a general difference in time-sharing ability between the students and pilots was not compelling. It was concluded that the degree by which time-sharing performance is structure-dependent is not easily alterable by training. The results suggested that laboratory findings on the structural determinants of time-sharing efficiency are generalizable to operational environments.

Tsang, P. S. (1986). Display/control integrality and time-sharing performance. *Proceedings of the Human Factors Society 30th Annual Meeting*, Vol. 1. (pp. 445-449). Santa Monica, CA: Human Factors Society.

Time-sharing performance was investigated as a function of the display and response integrality of the time-shared tasks. A manual step-tracking task was time-shared with a Stroop task that could be responded to manually or by speech. A secondary task technique was employed to manipulate the resource allocation between the two tasks. Display integrality was manipulated by: (1) contingent processing of the different dimensions of the Stroop task, and (2) the "objectness" of the dual-task display. Response integrality was manipulated by the number of responses required of the dual task and the response modality of the Stroop task. A prevalent resource-competition effect between the manual responses of the two tasks was observed, supporting the concept of multiple

resources. Results were also in concordance with Kahneman's object file model of attention; demonstrating that irrelevant elements within an object were difficult to ignore. The findings demonstrated the interactive effects of resource competition and task integrality on time-sharing performance.

Tsang, P. S. and Vidulich, M. A. (1986). Attentional processes and color object displays. *Proceedings of the International Scientific Conference: Work with Display Units* (pp. 557-560). Stockholm, Sweden.

The effects of integrated color displays on dual-task performance will be examined in a series of experiments employing a Stroop task and a Fitts' law target acquisition task. Also of interest is the effect of the display-control relationship on attention division between two concurrent tasks. In these experiments, the display integrality is manipulated by the temporal and spatial proximity of the stimuli for the two tasks and by the presence or absence of a Gestalt "good-figure" relationship among the elements. The control integrality is manipulated by the mode of response (manual vs. vocal) and the number of responses required to perform the two tasks concurrently. The effect of color is expected to interact with the stimulus integrality and predictions from two contemporary attention models will be tested. Results of the experiments are expected to provide some insights into the application of color in visually presented objects in a multitask environment.

Vidulich, M. A. (1986). Response modalities and time-sharing performance. *Proceedings of the Human Factors Society 30th Annual Meeting*, Vol. 1 (pp. 337-341). Santa Monica, CA: Human Factors Society.

The recent development of speech technology has provided an opportunity for new approaches in display/control design. Some researchers have proposed that the use of speech can reduce resource competition with manual controls and improve multitask performance. However, it has also been suggested that due to the heavy reliance on within-subject experimental designs, the research supporting the resource competition hypothesis was potentially contaminated by asymmetric transfer. The present study examined the value of speech responses as a control device in a dual-task experiment. The experimental design permitted the evaluation of asymmetric transfer effects. Despite numerous significant effects supporting the advantage of mixing manual and speech responses there was no statistically significant finding that suggested the occurrence of asymmetric transfer. Also, the value of speech output was demonstrated in between-subject analyses that were logically immune to asymmetric transfer effects. Therefore, although the possibility of asymmetric transfer remains a legitimate experimental design concern, it is not a sufficient explanation for the observed response modality effects. The present results supported the resource competition hypothesis of response modality effects, and suggested that in operational environments the judicious use of speech technology can enhance performance.

Vidulich, M. A. and Pandit, P. (1986). Training and subjective workload in a category search task. *Proceedings of the Human Factors Society 30th Annual Meeting*. Vol. 2 (pp. 1133-1136). Santa Monica, CA: Human Factors Society.

This study examined automaticity as a means by which training influences mental workload. Two groups were trained in a category search task. One group received a training paradigm designed to promote the development of automaticity; the other group received a training paradigm designed to prohibit it. Resultant performance data showed the expected improvement as a result of the development of automaticity. Subjective workload assessments mirrored the performance results in most respects. The results supported the position that subjective mental workload assessments may be sensitive to the effect of training when it produces a lower level of cognitive load.

Vidulich, M. A. and Tsang, P. S. (1986). *Collecting NASA Workload Ratings: A Paper and Pencil Package*. Working paper, Moffett Field, CA: NASA Ames Research Center.

This package is a collection of materials designed to collect subjective workload assessments. This procedure for collecting workload ratings has been developed by the Human Performance Group at NASA Ames Research Center as a result of three years and 25 experiments. Although the technique is still very much in the evaluation stage, this package is being distributed to allow other researchers to examine the techniques in their own experiments. This package is intended to fill strictly a "nuts and bolts" function of describing the procedure. A bibliography is provided at the end of these instructions for those researchers interested in the logic that supports the procedure and previous empirical findings. There are two main components to the procedure: one, the rating scales themselves are a set of six bipolar rating scales selected to give a good coverage of the subject's experiences in the different task conditions. Two, the sources-of-workload evaluation is designed to provide weights to adjust for individual biases in the use of the rating scales and to identify the specific sources of loading that were most influential for a given task.

Vidulich, M. A. and Tsang, P. S. (1986). Evaluation of two cognitive abilities tests in a dynamic dual-task environment. *Proceedings of the 21st Annual Conference on Manual Control* (NASA CP-2428). (pp. 12.1-12.10). Washington, DC: National Aeronautics and Space Administration.

Most real-world operators are required to perform multiple tasks simultaneously. In some cases, such as flying a high-performance aircraft or trouble-shooting a failing nuclear power plant, the operator's ability to "time-share" or "process in parallel" can be driven to extremes. This has created interest in selection tests of cognitive abilities. Two tests that have been suggested are the Dichotic Listening Task and the Cognitive Failures Questionnaire. Correlations between these test results and time-sharing performance were obtained and the validity of these tests were examined. The primary task was a tracking task with dynamically varying bandwidth. This was performed either alone or concurrently with either another tracking task or a spatial transformation task. The results were: (1) An unexpected negative correlation was detected between the two tests. (2) The lack of correlation between either test and task performance made the predictive utility of the tests scores appear questionable. (3) Pilots made more errors on the Dichotic Listening Task than did college students.

Vidulich, M. A. and Tsang, P. S. (1986). Techniques of subjective workload assessment: A comparison of SWAT and the NASA-Bipolar methods. *Ergonomics*, 29 (11), 1385-1398.

Assessment of subjective workload is becoming increasingly important in the evaluation of human-machine systems. Two popular methods were compared: (1) the Subjective Workload Assessment Technique (SWAT) that employed a conjoint measurement procedure to confer interval scale properties on the workload ratings, and (2) a technique under development at NASA that used an individually weighted workload score. Both methods were applied in a laboratory experiment that required rating a number of single- and dual-tracking and spatial transformation tasks. Both subjective assessment techniques displayed similar sensitivity to the different task manipulations. However, both techniques failed to detect the resource-competition effects in the dual-task performance, and were in general insensitive to response execution-processing demands. A notable difference between the two techniques was that the NASA-Bipolar ratings consistently had a smaller between-subject variability than the SWAT ratings. Discussion of the results is centered around the issue of the validity of assessment of subjective workload in general, and the construct and concurrent validity of the two techniques in particular.

Vidulich, M. A. and Wickens, C. D. (1986) Causes of dissociation between subjective workload measures and performance: Caveats for the use of subjective assessments. *Applied Ergonomics*, 17 (4), 291-296.

Dissociations between subjective workload assessments and performance were investigated. The difficulty of a Sternberg memory-search task was manipulated by varying stimulus presentation rate, stimulus discernibility, value of good performance, and automaticity of performance. All Sternberg task conditions were performed both alone and concurrently with a tracking task. Bipolar subjective workload assessments were collected. Dissociations between workload and performance were found related to automaticity, presentation rate and motivation. The results were interpreted as supporting the hypothesis that the specific cognitive processes responsible for subjective assessments can differ from those responsible for performance. The potential contamination these dissociations could inflict on operational assessments was discussed.

White, S. A., McKinnon, D. P., and Lyman, J. (1986). Modified petri net sensitivity to workload manipulations. *Proceedings of the 21st Annual Conference on Manual Control* (NASA CP-2428). (pp. 3.1-3.17). Washington, DC: National Aeronautics and Space Administration.

The purpose of this research is to investigate modified Petri nets (MPNs) as a workload-modeling tool. This paper describes the results of an exploratory study of the sensitivity of MPNs to workload manipulations in a dual task. Petri nets have been used to represent systems with asynchronous, concurrent, and parallel activities (Peterson, 1981). These characteristics led some researchers to suggest the use of Petri nets in workload modeling where concurrent and parallel activities are common. Petri nets are represented by places and transitions. In the workload application, places represent operator activities and transitions represent events. MPNs have been used to formally represent task events and activities of a human operator in a man-machine system. For example, Madni, Chu, Purcell, and Brenner (1983) used MPNs to model the tasks underlying the identification and reaction to a lube oil leak in a ship propulsion system. Madni and Lyman (1983) used

a MPN to model the checkout and start-up procedure for a Cessna 182 light aircraft. White, MacKinnon and Lyman (1984) formulated a MPN for POPCORN, a complex computer simulation at NASA Ames for workload research. These descriptive applications demonstrate the usefulness of MPNs in the formal representation of systems.

Wickens, C. D. (1986) Gain and energetics in information processing In R. Hockey, A. Gaillard, and M. Coles (Eds.), *Energetics and Human Information Processing*. (pp. 373-390). The Netherlands, Nijhoff.

The concept of gain, related to bias and to signal-to-noise ratio, is introduced as an element that should continuously modulate the components of information-processing models. The relation between this concept and different existing models, particularly in tracking, is described and the different sources and pathways of gain modulation in the human processing system are categorized. It is then explained how gain parameters have been useful in accounting for strategy choices in cognitive tasks, and for resource competition in dual-task situations. However, in the dual-task situation caution is prescribed to separate gain-related changes based on scarce resources from other sources of dual task interference. This separation must be based on converging evidence from performance analysis and from neurophysiology.

Wickens, C. D. and Yeh, Y. -Y. (1986). A multiple resources model of workload prediction and assessment. *Proceedings of the IEEE Meeting on Systems, Man, and Cybernetics*, Vol. 2 (pp. 1044-1048). Piscataway, NJ: IEEE Service Center.

The Multiple Resource Model defines different structural resources within the human processing system. This paper first describes how the model may be employed, early in the system design process, to predict performance in complex settings. Limitations of the model in this regard are also pointed out. The paper then describes how the model may be used for prescribing workload-assessment techniques late in the design process, and for interpreting the dissociations that are often observed between subjective workload and performance.

Wierwille, W. W., Casali, J. G., Connor, S. A. and Rahimi, M. (1986) Evaluation of the sensitivity and intrusion of mental workload estimation techniques. *Advances in Man-Machine Systems Research*, Vol. 2 (pp. 51-127).

The objective of the research reported in this paper was to examine the sensitivity and intrusion of a wide variety of workload-assessment techniques in simulated piloting tasks. The study employed four different piloting tasks emphasizing psychomotor, perceptual, mediational, and communications aspects of piloting activities. Techniques in the opinion, primary-task, secondary-task, and physiological categories were evaluated. An instrumented moving-base general aviation aircraft simulator was used for the study. This paper provides a summary of the research.

Zaleski, M. and Moray, N. (1986). Hitts' Law? A test of the relationship between information load and movement precision. *Proceedings of the 21st Annual Conference on Manual Control*, (NASA CP-2428). (pp. 22.1-22.21). Washington, DC: National Aeronautics and Space Administration.

Recent technological developments have made viable a man-machine interface heavily dependent on graphics and pointing devices. This has led to new interest in classical reaction and movement time work by Human Factors specialists.

Two experiments were designed and run to test the dependence of target capture time on information load (Hick's Law) and movement precision (Fitts' Law). The proposed model linearly combines Hick's and Fitts' results into a combination law which then might be called Hitts' Law. Subjects were required to react to stimuli by manipulating a joystick so as to cause a cursor to capture a target on a CRT screen. Response entropy and the relative precision of the capture movement were crossed in a factorial design and data obtained that were found to support the model.

APPENDIX I:
RESEARCH PAPERS AND PUBLICATIONS

-1987-

Battiste, V. (1987). *Part-Task vs. Whole-Task Training: Twenty Years Later*. Unpublished Master's Thesis. San Jose, CA: San Jose State University.

The primary aim of training is to improve performance. Part-task training may be the more economical method, because full-mission training simulators often cost more than the vehicles they simulate. The skills acquired with a Part-task approach can often be learned with devices that are less expensive, thus the cost of training may be reduced considerably, however the skills learned may not transfer effectively to performance of the complete task. This study investigated the effectness of Part-task training on the psychomotor portion of a supervisory control simulation. That is, specific training was provided to develop proficiency with the cursor-control device (a magnetic pen and pad). Prior to a transition to the Whole-task. Twelve subjects, which were divided into two groups based on their criterion task scores, served as paid participants. Subjects were seated in front of a video screen on which the simulation was displayed. The subject's job was to perform subtasks, represented by symbols, from each of five boxes as quickly as possible. Each trial consisted of one combination of the within-subject variables: interval between box refilling (30 or 60 sec) and element velocity (1.53 or 3.06 cm/sec). during a trial, each box was refilled four times with seven symbols. There were some distinct advantages of the initial Part-task training: 1) The Part-task group learned the task faster; 2) The Part-task group's scores and elapsed times continue to improve; and 3) The Part-task group experienced about the same reduction in workload during training. The primary focus of the present experiment was on a speed related aspect of the Popcorn task - cursor movement and control - and resulted in significant increases in speed of response for the Part-task group. These findings of improved performance due to Part-task training may be useful in designing training programs for other supervisory control environments (i.e., advanced aircraft and air traffic control).

Bortolussi, M. R., Hart, S. G., and Shively, R. J. (1987). Measuring moment-to-moment pilot workload using synchronous presentations of secondary tasks in a motion-base trainer. *Proceedings of the Fourth Symposium on Aviation Psychology*. (pp. 651-657). Columbus: Ohio State University.

A simulation was conducted to determine whether the sensitivity or secondary task measures of pilot workload could be improved by synchronizing their presentation to the occurrence of specific events or pilot actions. This synchronous method of presentation was compared to the more typical asynchronous method, where secondary task presentations are independent of pilot's flight-related activities. Twelve pilots flew Low- and High-Difficulty scenarios in a motion-base trainer with and without concurrent secondary tasks (e.g., choice was manipulated by the addition of 21 flight-related tasks superimposed on a standard approach landing sequence. The insertion of the secondary tasks did not affect primary flight performance. However, secondary task performance did reflect workload differences between scenarios and among flight segments within scenarios, replicating the results of an earlier study in which the secondary tasks were presented asynchronously. In addition, the choice reaction time secondary task was also sensitive to the workload of

specific activities within flight segments. Workload ratings were virtually identical between this and the earlier study.

Brumberg, R. and Wu, J. (1987). *Validation of the NMM Cogitometer*. Unpublished B.S. Thesis. Canada: University of Toronto, Department of Industrial Engineering.

The use of the 0.1 Hz component of the heart rate variability signal as a measure of mental workload has been validated in a number of laboratory experiments. In this experiment, the NMM Cogitometer, an on-line device which through digital filtering techniques isolates the 0.1 Hz component of the heart rate variability, was used to measure the mental workload of subjects performing mental arithmetic. In performing the appropriate statistics on the results obtained we feel that this device has potential but still requires further refinements before this device will be able to measure intervals of steady state load which are changing as rapidly as every ten to fifteen seconds, and to detect changes at shorter intervals.

Donchin, E., Hart, S. G., and Hartzell, E. J. (1987). *Executive Summary: Workshop on Workload and Training, an Examination of their Interactions*. (NASA TM-89459). Washington, DC: National Aeronautics and Space Administration.

This report provides an overview of the Workshop on Workload and Training: An Examination of their Interactions which was held in Carmel, California from January 5 to 10, 1986. The workshop was jointly sponsored by Ames Research Center's Aerospace Human Factors Research Division and the Army Aeroflightdynamics Directorate, and was organized and chaired by Dr. Emanuel Donchin. The goal of the workshop was to bring together experts in the fields of workload and training and representatives from the Department of Defense and industrial organizations who are responsible for specifying, building, and managing advanced, complex systems. The challenging environments and requirements imposed by military helicopter missions and space station operations were presented as the focus for the panel discussions. The workshop enabled a detailed examination of the theoretical foundations of the fields of training and workload, as well as their practical applications. Furthermore, it created a forum where government, industry and academic experts were able to examine each other's concepts, values, and goals. The discussions pointed out the necessity for a more efficient and effective flow of information among the groups represented. The executive summary describes the rationale of the meeting, summarizes the primary points of discussion, and lists the participants and some of their summary comments. A complete transcription of the tutorials and panel reports is being transcribed and will be published in their entirety.

Fuld, R., Liu, Y. and Wickens, C. D. (1987). *Computer Monitoring vs. Self Monitoring: The Impact of Automation on Error Detection*. (ARL-87-3/NASA-87-4). Urbana-Champaign: University of Illinois, Aviation Research Laboratory.

Nine subjects received 10 hours of training on a micro computer-based decision making task in which series of incoming customers were assigned to one of three queues with the shortest estimated wait. Two operating modes were then compared. In the manual mode, subjects monitored their own assignments for errors. In the automatic mode, the computer

made the assignments, while subjects continued to monitor for errors. Unknown to the subjects, the computer assignment stream was a playback of their earlier manual assignment performance. Fast and slow assignment paces were also compared as a workload manipulation.

Signal detection analysis showed subjects to be biased against declaring assignment errors in the manual mode, as well as less sensitive to misassignments in manual mode. These effects were coextant with higher subjective workload in manual. Results are discussed in terms of attentional resources, human decision making, and automation's impact on the operator.

Hancock, P. A. (1987). Arousal theory, stress and performance: Problems of incorporating energetic aspects of behavior into human-machine systems function. In L. S. Mark, J. S. Warm, and R. L. Huston (Eds) *Ergonomics and Human Factors: Recent Research*. (pp. 170-179). New York: Springer-Verlag.

This paper develops a critique of the unitary behavioral arousal theory of stress and human performance. The empirical, methodological, and theoretical shortcomings of this position are elaborated. The contemporary alternatives that have been generated to fill this theoretic vacuum are identified. Our limited understanding of the action of stress is taken as one example of why important *energetic aspects of performance* have yet to be incorporated into human-machine systems design and operation. Some steps directed toward such integration are developed.

Hancock, P. A. and Chignell, M. H. (1987). Adaptive control in human-machine systems. In P. A. Hancock (Ed.) *Human Factors Psychology*. (pp. 215-243). The Netherlands: North Holland.

In this chapter we review contemporary advances in the understanding of adaptive control as applied to systems which include the cooperative action of a machine and its operator. As an initial foundation we recognize that the prosthetics which can surround individuals and augment their capabilities allow human operators to traverse the traditional boundary constraints imposed by the environment. This freedom is granted only through harmonious and compatible interaction between operator and machine. The failure of *synchronization* between these two cooperative, intelligent, and goal-directed entities can result in sometimes serious, and occasionally irreversible violation of overall system integrity.

In our overview we begin by examining human adaptive response to stressful conditions and a particular expression of this capability in task-related mental workload. We then indicate examples of the growth of adaptive capability in automated machine systems. Finally, we examine the architecture of adaptive human-machine interfaces. These latter forms of interface use, among other inputs, estimates of operator mental workload to optimize the interactive articulation between human and machine in coping with task demands. In reviewing the progress in these areas, we indicate a number of promising avenues for future exploration. Prior to examining these developments in detail, we have summarized some of the forces involved in the changing nature of work that are driven by contemporary technological developments.

- Hart, S. G. (1987). Measurement of pilot workload. In A. Roscoe (Ed.), *AGARDograph on Pilot Workload Assessment* (AGARDograph No. 282). (pp. 116-122). Neuilly sur Seine, France: Advisory Group for Aerospace Research and Development.

A procedure for analyzing pilot workload is proposed and an example of how it might be used assess the workload encountered during the final five minutes of approach and landing for a transport aircraft is described. After specifying the research question that will be addressed by the evaluation, a simple time-line analysis is performed, performance criteria are established, and measures that can be used inflight that address the research question are selected: subjective ratings, heart rate and heart rate variability, communications analysis, measure of flight path control and time estimation. Finally, a summary of how the experimental results might be interpreted is provided.

- Hart, S. G. (1987). NASA Ames Workload Research Program. *Proceedings of the Space Station Human Factors Research Review*. (NASA CP-2426, vol. 4). Washington, DC: National Aeronautics and Space Administration.

The NASA Ames Research Center workload-assessment research program was described. A theoretical model of human performance and workload that serves as the unifying focus of the program was reviewed. Theoretical and applied research in support of the model were described, with particular emphasis on space-related application. The workload measures described included: subjective ratings, secondary and primary task measures, and a variety of physiological techniques. The development of a workload predictive model was described, with particular emphasis on its application to the RMS operations in the shuttle and space station.

- Hart, S. G. (1987). The prediction and measurement of mental workload during space operations. Presented at the Space Life Sciences Symposium. Washington, DC, June.

The field of workload assessment is reviewed briefly, and the results of a five-year NASA research effort are described. As a result of research performed at Ames and elsewhere, an alternative way to conceptualize operator workload is suggested. With this approach, workload is identified as one of the drivers in determining how operators perform complex tasks. This is contrast to the traditional definition of workload as the product of task demands. Finally, the goals and approach of a newly initiated research program focused on the workload and performance of Spation crewmembers are described.

- Hart, S. G., Battiste, V., Chesney, M. A., Ward, M. M., and McElroy, M. (1987). Responses of Type A and Type B individuals performing a supervisory control simulation. In B. Salvendy (Ed.) *Proceedings of the Second International Conference on Human-Computer Interaction*. (pp. 67-74). The Netherlands: Elsevier Science Publishers.

A supervisory-control simulation, presented with different levels of complexity and time pressure, was used to examine candidate behavioral, subjective, and physiological measures of mental workload. The predicted relationships were found among physiological and subjective workload measures, but their correlations with performance were low. Significant Difficulty x Personality Type interactions were found for heart rate, systolic,

and diastolic blood pressure: Type A men exhibited greater physiological responsiveness to difficulty manipulations than Type B men, suggesting that behavioral characteristics are particularly salient when interpreting physiological workload measures.

Hart, S. G. and Hauser, J. R. (1987). Inflight application of three pilot workload measurement techniques. *Aviation, Space, and Environmental Medicine*. 58 402-410.

Three measures of workload were tested during 11 routine missions conducted by the NASA Kuiper Airborne Observatory: communications performance, subjective ratings, and heart rate. The activities that contributed to crewmember workload varied; the commander was responsible for aircraft control and navigation whereas the copilot handled communications with ATC and the astronomers. For both crewmembers, rated workload, stress, and effort were equivalent and varied significantly across flight segments, peaking during takeoff and landing. Thus, subjective ratings appeared to reflect the overall experiences of the crew, rather the specific experiences of each pilot. For all flights, fatigue increased significantly from takeoff to landing, although the increase was significantly greater as landing times shifted from 10:00 pm and 9:00 am. The type, source, number, and frequency of communications tasks varied significantly across flight segments, providing an objective indicator of pilot workload. Heart rate was significantly higher when pilots were assigned to the left seat than the right. Although it peaked during takeoff and landing, for both positions, the change was significantly greater for the pilot-in-command. Subjective assessments of stress, workload, and mental effort were significantly correlated with heart rate and communications frequency but were unrelated to mission duration, rated fatigue, or pilot evaluation of performance.

Johnson, W. W., Tsang, P. S., Bennett, C. T., and Phatak, A. V. (1987). The visual control of simulated altitude. *Proceedings of the Fourth Symposium on Aviation Psychology*. (pp. 216-222). Columbus: Ohio State University.

The ability to control altitude as a function of changes in optical splay angles (the perspective angles formed by ground texture lines parallel to the direction of flight) and optical texture density (the number of texture units per degree of visual angle) was examined. Previous studies have shown people detect descent more accurately and quickly in flight simulations when just splay angles are present (i.e., only parallel ground texture is present) than when just optical texture density is present (i.e. only orthogonal ground texture is present). Two limitations of these studies were 1) they only involve passive observers and 2) subjects could perform the task in the splay condition by simply reponding when they detected movement in the intersections of the optical projection of the parallel ground texture lines with the edges of the display screen. To eliminate the first limitation the present experiment required subjects to actively control altitude while being buffeted by vertical winds (the active control task). To eliminate the second limitation uncontrollable lateral winds, which caused splay line/screen edge intersection to move in a manner uncorrelated with the vertical disturbance, were introduced. It was found that under these conditions people performed best with the orthogonal, and not the parallel, ground texture. Furthermore, it was found that during simulated flights over the parallel ground texture people would make inappropriate altitude corrections in response to the lateral winds. This supports the hypotheses that people inappropriately use movements in the screen/splay line intersections as information for altitude change.

Kramer, A. F., Sirevaag, E. J. and Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. *Human Factors*, 29 (2), 145-160.

Previous research has indicated that components of the event-related potential (ERP) may be used to quantify the resource requirements of complex cognitive tasks. The present study was designed to explore the degree to which these results could be generalized to complex, real-world tasks. The study also examined the relations among performance-based, subjective and psychophysiological measures of operator workload. Seven male volunteers, enrolled in an instrument flight rule (IFR) aviation course at the University of Illinois, participated in the study. The student pilots flew a series of IFR flight missions in a single-engine, fixed-base simulator. In dual-task conditions subjects were also required to discriminate between two tones differing in frequency and to make an occasional overt response. ERPs time-locked to the tones, subjective effort ratings, and overt performance measures were collected during two separate 45-min. flights differing in difficulty. The difficult flight was associated with high subjective effort ratings, as well as increased deviations from the command altitude, heading, and glide slope. The P300 component of the ERP discriminated among levels of task difficulty, decreasing in amplitude with increased task demands. Within-flight demands were examined by dividing each flight into four segments: takeoff, straight and level flight, holding patterns, and landings. The amplitude of the P300 was negatively correlated with deviations from command headings across the flight segments. In sum, the findings provide preliminary evidence for the assertion that ERP components can be employed as metrics of resource allocation in complex, real-world environments.

Krushelnicky, E. (1987). *Fuzzy Set Measurement of Mental Workload*. Unpublished M.A.Sc. Thesis. Canada: University of Toronto, Department of Industrial Engineering.

Mental workload is investigated in the context of Rasmussen's taxonomy of human behavior. Subjective measurements in the form of fuzzy set membership estimates were gathered for skill-based, rule-based, and knowledge-based behavior, as well as for interactions between them. Results indicated that models could be found to predict the difficulty of combinations of different kinds of behavior, knowing the rated difficulty of the component types. Implications suggesting a rethinking of Rasmussen's taxonomy were found. In this task, skill-based behavior was found to load a person, with rule- and knowledge-based behavior acting as moderating influences on the rated difficulty. Skill-based tasks were found to dominate. No particular interaction term was found to represent the best model for all cases considered, but the drastic intersection operator gave marginally better results for all combinations. Linguistic interpretation of the raw data supported the results obtained from regression modeling. Further investigations of mental workload evaluation using fuzzy set calculus are recommended.

Liang, K. S. and Szynda, A. (1987). *Validation of a 0.1 Hz Power Spectrum Analyzer of Mental Workload*. Unpublished B.S. Thesis. Canada: University of Toronto, Department of Industrial Engineering.

The determination of a viable method for measuring human mental workload has been sought after through various research. One method, in particular, which focuses on the 0.1 Hz power spectrum analyzer of the heart rate variability signal will be used to validate the measure of mental workload. The analyzer, NMM Cogitometer, was subjected to a number

of tests and the subsequent data was collected for analysis. Statistical, qualitative, and subjective measures of validating the Cogitometer were imposed. However, due to the amount of error sources in the experiment, the data set had manifested unreliable conclusions pertaining to the validity of the mental workload measuring device. In essence, these sources of errors should be corrected in order that further research warranted.

Lintern, G. and Wickens, C. D. (1987). *Attention Theory as a Basis for Training Research*. (ARL-87-2/NASA-87-3). Urbana-Champaign: University of Illinois, Institute of Aviation.

In this paper concepts of attention theory are used to develop an approach to training research. Multiple-resource theory and the time-sharing strategies of task alternation and task integration are outlined first; then hypotheses about the relationship of multiple resources and time-sharing to the acquisition of complex skills are developed. Data are reviewed from the experimental paradigms of guided training, adaptive training, and manipulations of task load during training, and also from experiments that have sought to examine the development of time-sharing strategies. Our review of data suggests a need to reduce resource loads in early learning without disrupting the acquisition of time-sharing skills. Resource-load effects were particularly evident where subjects were required to learn complex stimulus-response relationships or to learn predictable patterns of events. We concluded that this task domain is particularly promising as a focus for research related to special training methodologies.

Liu, Y. and Wickens, C. D. (1987) *Mental Workload and Cognitive Task Automation: An Evaluation of Subjective and Time Estimation Metrics*. (D-87-2/NASA-87-2). Urbana-Champaign: University of Illinois, Engineering-Psychology Research Laboratory.

The present study focused on two workload measurements: subjective assessment and time estimation, as applied to the task of decision and monitoring. The task required the assignment of a series of incoming customers to the shortest of three parallel service lines displayed on a CRT display. The subject was either in charge of the customer assignment (Manual Mode) or was monitoring an automated system performing the same task (Automatic Mode). In both cases, the subjects were required to detect the nonoptimal assignments that they or the computer had made, and to detect the very infrequent lane closure situations in which a lane stopped processing its "customers." Time pressure was manipulated by the experimenter to create fast and slow conditions. The results indicated that subjective workload was more influenced by the subject's participatory mode than by the variable of task speed. Workload in the manual mode was rated significantly higher than workload in the automatic mode. Only subscales of the NASA bipolar rating measures discriminated the workload caused by increased speed. The time estimation intervals produced while performing the decision and monitoring tasks had significantly greater length and larger variability than those produced while either performing no other tasks or automaticity. The experimental results were discussed in terms of mental representation and behavioral automaticity.

Moray, N., Eisen, P., Money, L., and Turksen, C. (1987). Fuzzy analysis of skill- and rule-based mental workload. Department of Industrial Engineering Working Paper 87-06.

Abstract not available.

Shively, R. J., Battiste, V., Matsumoto, J. H., Pepitone, D. D., Bortolussi, M. R., and Hart, S. G. (1987). Inflight evaluation of pilot workload measures for rotorcraft research. *Proceedings of the Fourth Symposium on Aviation Psychology*. (pp. 637-643). Columbus: Ohio State University.

The assessment of workload in aviation is accepted as an important aspect of aerospace research. Many techniques for assessing workload have been tested and (with varying degrees of success) have been shown to be useful. While the problem of assessment has yet to be solved, most researchers feel that methodology and understanding of workload have progressed to the point that it is possible to adequately assess workload in many diverse situations. A growing effort is now focused upon moving past workload assessment into workload prediction. Workload prediction has at least three important purposes: 1) predict workload peaks in proposed scenarios to be flown in existing aircraft, 2) Predict possible changes in workload due to modification of system design or manpower, 3) Predict the workload associated with new (or not yet developed) aircraft design. The present experiment was designed to provide data for a functionally-specified predictive model. In addition, this experiment compared a physiological workload measure (spectral analysis of heart-rate IPI) to an established validated technique (NASA-Task Load Index, TLX). If the physiological technique is shown to be useful, it may help provide a more finely grained (second by second) workload evaluation. A SH-3G helicopter was flown on two predetermined scenarios. The scenarios consisted of basic flight tasks such as hover, terrain following and contour flight. The heart-rate of the pilot was recorded throughout the flight. After each maneuver or segment of interest, the safety pilot took control of the aircraft and the evaluation pilot rated the workload of that segment on the NASA-Task Load Index (TLX). The workload ratings clearly distinguish between flight segments within each scenario. Additional analysis of the individual tasks revealed that those tasks that are functionally similar received similar workload ratings. The heart-rate data is still under analysis and when completed will be compared to the NASA-TLX ratings to determine the utility of this techniques for operational workload assessment.

Tsang, P. S., Hart, S. G., and Vidulich, M. A. (1987). Effects of display-control I/O, compatibility, and integrality on dual-task performance and subjective workload. *Proceedings of the AGARD Aerospace Medical Panel Symposium on Information Management and Decision Making in Advanced Airborne Weapon Systems* (Proceedings No. 414). Neuilly sur Seine, FR: Advisory Group for Aerospace Research and Development.

The utility of speech technology was evaluated in terms of three dual-task principles: resource competition between the time-shared tasks, stimulus-central processing-response compatibility, and task integrality. Empirical support for these principles was reviewed. Two studies investigating the interactive effects of the three principles were described. Objective performance and subjective workload ratings for both single- and dual-tasks were examined. It was found that the single-task measures were not necessarily good predictors

for the dual-task measures. It was shown that all three principles played an important role in determining an optimal task configuration. This was reflected in both the performance measures and the subjective measures. Therefore, consideration of all three principles is required to insure proper use of speech technology in a complex environment.

Tsang, P. S. and Johnson, W. W. (1987). Automation: Changes in cognitive demands and mental workload. *Proceedings of the Fourth Symposium on Aviation Psychology*. (pp. 616-622). Columbus: Ohio State University.

The cognitive demands of automated systems and implications for mental workload were examined. A variety of tasks were designed to tax different cognitive systems in a sufficiently complex scenario where automation would be necessary. The task battery included two manual control tasks and a decision making task. The manual control tasks were represented by a continuous 3-D flight path control task and a discrete Fitts' target acquisition task. The decision making task was represented by selectively turning off non-essential subsystems to conserve power during an engine failure. Three subjective workload assessment techniques were used: (a) an overall workload scale, (b) the NASA - Task Load Index (TLX) scale (Hart and Staveland, in press), and (c) the Bedford scale (Roscoe, 1984). The workload ratings suggested that the phenomenal experience of subjective workload was robust and assessable by a multitude of techniques. Results also demonstrated the value of understanding the cognitive demands in the process of function allocation between human and machine. For example, automating the lateral flight control task had significantly different effects on the target acquisition task and the engine failure task. These effects were reflected both in performance and in the phenomenal experience of the subjects.

Vidulich, M. A. and Pandit, P. (1987). Individual differences and subjective workload assessment: comparing pilots to nonpilots. *Proceedings of the Fourth Symposium on Aviation Psychology*. (pp. 630-636). Columbus: Ohio State University.

Two of the more popular subjective workload assessment techniques (SWAT and NASA-TLX) incorporate procedures intended to evaluate and adjust for individual differences in the perception and reporting of subjective workload. Two groups of subjects, pilots and nonpilots, filled out the individual differences portions of SWAT and NASA-TLX, along with several traditional personality tests. Although almost all the personality tests discriminated between the groups, the workload tests did not. Also, the workload tests did not correlate with the personality tests in any apparent pattern nor were the intercorrelations between the workload tests consistent with expectations. While the workload tests may provide useful information regarding the interaction between tasks and personality, the present results do not support their effectiveness as pure tests of individual differences.

APPENDIX J:
RESEARCH PAPERS AND PUBLICATIONS

- In Press -

Battiste, V. (in press). Part-task vs. whole-task training on a supervisory control task. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

The primary aim of training is to improve performance. Part-task training may be the more economical method, because full mission training simulators often cost more than the vehicles they simulate. However, the skills learned may not transfer effectively to performance of the complete task. This study investigated the effectiveness of Part-task training on the psychomotor portion of a supervisory control simulation. Twelve subjects were divided into Part-task and Whole-task groups and told to perform the task as quickly as possible. Part-task training was provided with the cursor-control device (a magnetic pen and pad), prior to transition to the Whole-task. Some distinct advantages of the Part-task training were: (1) The Part-task group learned the task faster; (2) The Part-task group's scores and task times continued to improve, while the Whole-task group's did not; and (3) A significant increase in speed for the Part-task group almost no improvement in speed for the Whole-task group.

Bennett, C. T., Graeber, R. C., and Voorhees, J. (in press). Army research psychology at NASA/Ames. Presented at the U. S. Army Medical Department Professional Postgraduate Short Course, *Research Psychology: Future Goals New Direction*. Washington, D.C., February 1987.

Currently there are three Army Research Psychologists at Ames Research Center, Moffett Field, CA. One (MAJ Voorhees) is assigned to the U. S. Army Aviation Research and Technology Activity (ARTA). The other two (LTC's Graeber and Bennett) are assigned to ODCSRDA and detailed to the Aerospace Human Factors Research Division.

MAJ Voorhees's recent efforts have concentrated on the development, acquisition, and program planning for the Crew Station Research and Development Facility (CSRDF). This is a \$15 million dollar R&D effort to establish ARTA as a focus for the development of cockpits for all types of rotorcraft. A main feature of the CSRDF is a fiber optic helmet mounted display system for the presentation of a virtual workload for flight evaluations. The system is also designed so that it can be quickly reconfigured for one or two man operations in multi-opponent scenarios.

LTC Graeber's research has focused on the effects of long- and short-haul flight operations on pilot performance based on physiological changes, as well as subjective reports. Much of this work was recently reported in a dedicated issue of *Aviation Space and Environmental Medicine* (V. 57, No. 12, Section II, 1986). LTC Graeber was attached to the Presidential Commission formed to investigate the Challenger accident, and responsible for the Human Factors portion of the report.

LTC Bennett is responsible for the helmet mounted display (HMD) research being

conducted under the rotorcraft program at Ames. His work is now concentrating on the development of virtual displays for the evaluation of three-dimensional tracking performance of pilots. This effort involves the investigation of how the presentation of optical/visual flow fields in an HMD influences spatial orientation.

Casper, P. A. and Kantowitz, B. H. (in press). Estimating the cost of mental loading in a bimodal divided-attention task: Combining reaction time, heart-rate variability and signal-detection theory. *Proceedings of the 1987 Mental-State Estimation Workshop*. Williamsburg, VA: NASA-Langley Research Center.

Mental workload is a multidimensional construct reflecting the interaction of several factors, including an operator's training and skill level, task demands, as well as the operator's physiological state, which itself is a function of manifold homeostatic systems. In order to better estimate the complex mental states produced by this multidimensional construct, several dependent variables need be jointly considered. The present research evaluates reaction time, heart-rate variability, and beta simultaneously.

Subjects are required to respond to simultaneous streams of binary auditory and visual events. Responses are manual and vocal. Performance measures reveal large workload effects due to (1) the modality of the stimulus stream, (2) the type of central processing required for each component task, and (3) the temporal compatibility of the two tasks. But such performance measures are difficult to implement in field settings, especially in aviation where obvious safety concerns override niceties of research design. More unobtrusive measures, such as heart rate, are preferred in operational settings. This presentation suggests how the use of cardiac measures of workload can help assess an operator's mental state in both laboratory and field settings. Disadvantages of cardiac measure, especially current disparities concerning the best way to characterize heart-rate variability, are also discussed.

Casper, P. A., Shively, R. J. and Hart, S. G. (in press). Decision Support for Workload Assessment: Introducing WC FIELDE. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

Currently there is a great demand for mental workload evaluation in the course of system design or modification. In light of this demand, a microprocessor-based decision support system has been created called WC FIELDE: Workload Consultant for FIELD Evaluation. The system helps the user select workload measures appropriate to his or her application from the wide pool of currently available techniques. Both novices and those with some workload experience may benefit from using WC FIELDE, since the systems operation is entirely transparent and all rules involved in the decision process are available for the user to examine. WC FIELDE recommends several assessment methodologies in decreasing order of appropriateness, and provides additional information on each measure at the end of the program in the form of text files.

Chignell, M. H. and Higgins, T. J. (in press). Intelligent warning systems for instrument landings. To appear in the *International Journal of Industrial Ergonomics*.

The requirements of maintaining supervisory control of a complex system may sometimes exceed the capability of the human operator for a variety of reasons ranging from temporary distraction from the main task to excessive mental workload. The ergonomic aspect of complex supervisory control can be improved by augmenting the interface with expert systems. This paper outlines the development of an intelligent warning system as one component of an augmented supervisory control system for complex aircraft. The system is designed to alert the pilot to potentially hazardous conditions during instrument landing approaches. The process of task analysis that precedes the implementation of the system is outlined, drawing on the experience of a former military pilot. The results and methods shown here can be generalized to other components of the pilot's task and, more generally, to other complex supervisory control tasks.

Connelly, Jr., J. G., Wickens, C. D., and Lintern, G. (in press). Attention theory and training research. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

This study used elements of attention theory as a methodological basis to decompose a complex training task in order to improve training efficiency. The complex task was a microcomputer flight simulation where subjects were required to control the stability of their own helicopter while acquiring and engaging enemy helicopters in a threat environment. Subjects were divided into whole-task, part-task, and part/open loop adaptive task groups in a transfer of training paradigm. The effect of reducing mental workload at the early stages of learning was examined with respect to the degree that subordinate elements of the complex task could be automated through practice of consistent, learnable stimulus-response relationships. Results revealed trends suggesting the benefit of isolating consistently mapped sub-tasks for part-task training and the presence of a time-sharing skill over and above the skill required for the separate subtasks.

Fuld, R. B., Liu, Y., and Wickens, C. D. (in press). The impact of automation on error detection: Some results from a visual discrimination task. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

Comparisons of operator performance in manual and automated systems have been in a large part limited to the manual tracking environment. The objective of the present study was to extend these investigations to a different task domain, one involving decision-making, selection, and monitoring. A dynamic decision-making task was developed that required subjects to assign serially incoming rectangular customers to one of three parallel queues depicted on a CRT. Service time was directly proportional to customer size, thus the waiting time of a customer entering the queue was proportional to the sum of the currently assigned customer areas. Subjects in the manual condition were then required to monitor their own assignment performance. In the automatic condition, those same subjects monitored the playback of their own performances, under the impression that the computer was now generating the the assignments (and errors). A fast and a slow pace were also compared as a workload manipulation. Analysis of the data indicated that operators in the automatic mode were significantly more sensitive to the occurrence of incorrect assignments. In addition, while operators showed a lack of significant response bias in the automatic mode, they showed a significantly conservative departure from the optimal response rate in the manual mode. This disinclination to revise an elected hypothesis is a well-documented feature of human decision-making. It is felt that this

study demonstrates that the judgment process is capable of interfering with even rudimentary perceptions in significant ways, such that humans are somewhat unlikely to catch their own mistakes. It has been suggested that improved decisions might be rendered by a decision-maker/decision-shadower team (Reason, 1985). This study lends empirical support to that idea.

Hancock, P. A. (in press). The effect of time of day upon the subjective estimate of mental workload during the performance of a simple task. *Proceedings of the Fourth Mid-Central Ergonomics/Human Factors Conference*.

One open question in mental workload assessment concerns the impact of diurnal variation upon the perceived load of constant task. To examine this question nineteen subjects performed a simple time estimation task by the production technique at four different times of day (0800, 1200, 1600, 2000h). Following each session subjects completed the NASA TLX workload assessment scale. Results indicated that female subjects had a greater intolerance for this repetitive task. Five of twelve female subjects failed to complete the series of exposures. There were no drop-outs among the male subjects. Analysis of the re-examining responses indicated that female subjects rated effort and frustration significantly higher and performance significantly lower than their male counterparts. There were no higher order interactions, and no significant effects were found for variation in time of testing. Caution concerning the ubiquitous application of these findings is advised in light of a number of potentially confounding influences.

Hancock, P. A. (in press) Incorporating the energetics of operator response in the human-machine-system function: A critique of the arousal theory of stress and performance. In L. Mark, J. Warm and R. Huston (Eds.) *Progress in Human Factors/Ergonomics*. Amsterdam: Springer-Verlag.

Useful human factors knowledge requires the integration of both symbolic and numeric reasoning as operationalized through both semantic and mathematical expression. While many traditional information-processing concepts use numeric bases and are easily incorporated into system design and operational specification, some key psychological constructs, particularly performance under operational stress, are expressed largely in semantic terms and have, consequently, suffered exclusion. A critique of the dominant recommendations for revision of theoretic approaches which emphasize the predictive capacity vital to meaningful integration into system function. A parallel effort by those entrenched in the numerical approaches to behavioral assessment is advocated, to include qualitative semantic reasoning into contemporary system configuration.

Hancock, P. A. and Chignell, M. H. (in press). The temporal dimension of mental workload and its application to adaptive interface design. Submitted to the *Special Issue of IEEE SMC on Human-Computer and Cognitive Engineering*. Piscataway, NJ: IEEE Service Center.

In examining the role of time in mental workload, this paper presents a different perspective from which to view the problem of assessment. Mental workload is plotted in three dimensions, whose axes represent effective *time* for action, perceived *distance* from desired goal state, and level of *effort* required to achieve such a goal. This representation allows the generation of *isodynamic workload contours* which incorporate the factor of

operator competence and equifinality of effort. An adaptive interface for dynamic task reallocation is then described which uses this mental workload assessment as an error signal for load leveling between operator and machine. In order to facilitate implementation and operation of this adaptive interface, previous formulations based on an attentional resource approach are augmented by the distinction between knowledge-based, rule-based, and skill-based behavior as distinguished by Rasmussen.

- Hancock, P. A. and Rosenberg, S. (in press) A model for evaluating stress effect of work with display units. In B. Knave and P. G. Widebäck (Eds.) *Selected Papers on Work with Display Units*. Amsterdam: North Holland.

With the introduction of new technologies comes stress. In the case of the display-unit operator, the action of many of these stresses are at once both subtle and complex. Our impoverished theoretical understanding of stress effects serves to limit the designers and managers in their attempts to provide safe, healthy, and productive work environments. The elaboration of a theoretical view of stressor interaction derived from the concepts of comfort and adaptability, as presented in this work, provides an initial direction toward the resolution of this important ergonomic problem.

- Hart, S. G. (in press). Helicopter human factors. In E. Wiener and D. Nagel (Eds.) *Human Factors in Aviation*. New York: Academic Press.

Helicopters are one of the most demanding and exciting challenges for human factors researchers. Not only has less attention been devoted to the problems faced by helicopter pilots than the pilots of other types of aircraft, but many of the problems they face are more severe. This chapter describes the range of task for which helicopters are used and the environments in which they fly, to provide a context for the description of current and advanced controls and displays. In addition, human factors problems typical of the cockpit environment (e.g., noise, vibration) and the primary sources of helicopter pilot workload (e.g., visual information processing and manual control) are discussed.

- Hart, S. G. (in press). NASA-Ames Research Center workload assessment program: Review of research. To appear in *Practical Assessment of Pilot Workload*.

NASA formed a Workload Program at Ames Research Center in 1982. During the first phase of the program, a number of methods of assessing workload were developed and tested in laboratory, simulation, and inflight research. Recently, the focus of the program has shifted to workload prediction. This paper reviews the measurement techniques that were developed and presents the results of several simulation and inflight experiments. In addition, the applicability of different measures for use in transport operations is discussed.

- Hart, S. G. (in press). NASA workload assessment research program: Theoretical foundation, assessment procedures, and applications. In *Workload in Transport Operations - Proceedings of the Meeting sponsored by the Community of European Countries*.

In 1981, NASA formed a Workload Assessment Program to address the many unresolved issues in this increasingly important field. The goal was to merge the theoretical information available from academia with the practical requirements of industrial and government organizations to develop a comprehensive workload definition and a set of practically useful measures and predictors. Throughout the program, well-controlled laboratory experiments provided answers to specific questions and theoretical issues while simulation and inflight research provided verification that the results were valid and meaningful in an operational environment. The first phase of the program was devoted to understanding the factors that influence pilot workload, evaluating existing assessment techniques, and developing new techniques. The work was accomplished by an active interaction between government laboratories, industry research groups, and universities. The second phase of the program, which is underway, is devoted to completing a computer model for workload prediction, developing workload criteria (e.g., how much is "too much" or "too little"), and investigating the relationship between workload, training, and performance. On a continuing basis, the methods and theories developed by participants in the program have been applied to specific operational and design problems at the request of other government agencies and industry. This report summarizes the research conducted during the first phase of the program and describes the results obtained in several simulator and inflight applications.

Hart, S. G. (in press) Review of subjective methods for measuring pilot workload. *Proceedings of the Workshop on the Assessment of Crew Workload Measurement Methods, Techniques, and Procedures: Preliminary Selection of Measures*. Wright-Patterson Air Force Base, OH.

This presentation reviewed current workload rating scales and evaluated their applicability in certifying transport aircraft. Workload is defined as the cost incurred by a human operator in achieving a particular level of performance. It reflects the interaction between an individual and the demands imposed by a particular task. Although different measures reflect the same global construct, they are each particularly sensitive to different aspects of workload. Subjective ratings can reflect remembered actions and conscious experiences, perceived task demands, estimated levels of performance, and opinions about the system used to perform the task. They do not directly represent objective task demands, unconscious cognitive processes, "automatic" behaviors, actual performance, reserve capacity, or physiological responses. Many methods of obtaining subjective ratings have been developed, however only limited information is available about their reliability, validity, sensitivity, and diagnosticity. Traditional measures of reliability are inappropriate because they were developed to measure trait variables (that are consistent within an individual), whereas workload represents a state variable (that changes from one situation to another). Nevertheless, several rating scales (such as SWAT and NASA-TLX) have demonstrated surprisingly high test-retest and alternate-forms reliability for the same raters and the pattern of ratings are similar for different raters. Because ratings depend directly on the rater's personal experiences, it is difficult to obtain objective evidence that they do, in fact, reflect these experiences accurately. Further, different people respond to (and thus perform) the same task in different ways; their skills, effort, and strategies vary, thus, their workload experiences are, in fact, different. Thus, the sensitivity and validity of candidate rating scales is usually determined by obtaining converging evidence from with other, more objective, measures, and by repeated use of a scale with many different tasks. If a scale is consistently sensitive to variations in task demands, then it is considered to be valid. Finally, diagnosticity is an important attribute. Since the sources of workload vary among tasks, ratings must suggest why workload is high (or

low) in a particular situation, so that system design or mission requirements can be adjusted to achieve an optimal level. Few rating scales are adequately diagnostic. For example, none of the modified Cooper-Harper-type or unidimensional scales provide any diagnostic information. SWAT and NASA-TLX provide information about psychological variables (e.g., stress, time pressure, mental demands), but not task-specific variables, because they were designed for general application. Other scales, such as that developed at Boeing, provides both psychological and task-specific information, but its range of application is limited. It is clear that subjective measures of workload should be included in evaluating new aircraft designs because they are uniquely able to reflect the pilots' experiences. However, since they can not represent all relevant aspects of workload, they should be accompanied by additional performance-based and physiological measures, as well.

Hart, S. G., Hartzell, E. J., Voorhees, J. W., Bucher, N. M. and Shively, R. J. (in press). An integrated approach to rotorcraft human factors research. *Proceedings of the 1987 NASA/Army Rotorcraft Technology Conference*. Moffett Field: NASA Ames Research Center.

As the potential of civil and military helicopters has increased, more complex and demanding missions in increasingly hostile environments have been required. Although new subsystems are being designed to meet these requirements, mission demands may have increased to the point that pilots will be overloaded during critical flight phases. Consequently, users, designers, and manufacturers have an urgent need for information about human behavior and function to create systems that take advantage of human capabilities, without overloading them. Because there is a large gap between what is known about human behavior and the information needed to predict pilot workload and performance in the complex missions projected for pilots of advanced helicopters, Army and NASA scientists are actively engaged in human factors research at Ames. The research ranges from laboratory experiments to computation modeling, simulation evaluation, and inflight testing. Information obtained in highly controlled, but simpler environments generate predictions which can be tested in more realistic situations. These results are used, in turn, to refine theoretical models, provide the focus for subsequent research, and ensure operational relevance, while maintaining the predictive advantages of a theoretical foundation. The goal of this paper is to describe the advantages and disadvantages of each type of research, provide examples of experimental results, and describe the Ames facilities with which such research is performed.

Hart, S. G. and Staveland, L. E. (in press). Development of a multidimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: Elsevier.

The results of a 3-year research program to identify the factors associated with variations in subjective workload within and between different types of tasks are reviewed. Subjective evaluations of 10 workload-related factors were obtained from 15 different experiments. The experimental tasks included simple cognitive and manual control tasks, complex laboratory and supervisory control tasks, and aircraft simulation. Task, behavior, and subject-related correlates of subjective workload experiences varied as a function of

difficulty manipulations within experiments and different sources of workload definition. A multidimensional rating scale was proposed in which information about the magnitude and sources of six workload-related factors are combined to derive a sensitive and reliable estimate of workload.

Haworth, L. A., Bivens, C. C., and Shively, R. J. (in press). Advanced cockpit and control configurations for single pilot nap-of-the-earth flight. *7th IEEE/AIAA-Digital Avionics System Conference*, October 1986.

The U.S. Army Aeroflightdynamics Directorate conducted a pilot workload and aircraft handling qualities investigation of single pilot operation in the combat Nap-of-the-Earth (NOE) environment. As predicted, superimposing mission management tasks on NOE flight-path management tasks resulted in degraded pilot Handling Quality Ratings (HQRs) and higher workload. Of the control configurations studied, only one configuration was rated satisfactory for single pilot NOE flight in comparison to eight control configurations rated satisfactory for dual crewmember NOE flight.

Higgins, T., Chignell, M. and Hancock, P. A. (in press). Augmented in supervisory control: An aviation case study. In M. G. Chignell, P. A. Hancock and A. Lowenthal (Eds.) *Intelligent Interfaces*. Amsterdam: North Holland.

As a result of the continual progress in aircraft capabilities and safety improvements, today's pilots are deluged with information and controls. Unfortunately, these technologically advanced controls and displays cause an operating risk unless the pilot is able to assimilate the data and take appropriate action in a timely manner should a problem develop. While the increasing trend towards automation has helped to keep the pilot's active workload at acceptable levels when the systems are functioning properly, the sheer amount of information to be absorbed by the pilot has created a control task with high demands for attentional resources.

Johnson, W. W. and Hart, S. G. (in press). Step Tracking Shrinking Targets. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

Four models describing how people might acquire targets that dynamically vary in size were examined; two that described movement speed as a simple function of target size (either initial or final) and two that described movement speed as a function of the predicted size of the targets as a fixed time in the future (one was referenced to the beginning of the reaction time phase, and the other to the end of the phase). It was found that movement time was best described as a function of a size prediction made at the end, rather than the start, of the reaction time phase. Subjective workload ratings primarily reflected the total amount of time needed to acquire the targets rather than the time pressure imposed by the diminishing size of these targets.

Kadlec, H. (in press). A production system model for the performance of complex task with applications for the study of mental workload. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

In an effort to begin to incorporate the concept of subjective mental workload within a theoretical framework, a production system model for the performance component of the complex task, called POPCORN, is presented. The production system was developed for the second level of complexity, and includes six of the possible twelve functions available to the operator. Following a brief review of recent studies on the relationships between subjective ratings of mental workload, performance, and task characteristics, the POPCORN task is described. The production system model of performance component of the task is represented by a hierarchical structure of goals and subgoals where the information flow is controlled by set of condition-action statements. The implementation of the production system which "plays" POPCORN (implemented on the IBM PC AT and called POPEYE) can be used to generate computer simulations of human operators performing the task under different task difficulty conditions. Although developed to simulate an operator at an asymptotic level of performance, the model is also discussed with respect to the acquisition and refinement of the productions (i.e., learning) to optimize performance, and the possibility of operator errors. The performance model will be embedded into a dynamic psychological model which will allow us to examine and quantify relationships between performance and psychological aspects of a complex task, and their contributions to subjective mental workload.

Kantowitz, B. H. (in press). Can conditioning concepts aid the study of human information processing? In J. Sidowski (Ed.). *Conditioning, Cognition, and Methodology: Contemporary Issues in Experimental Psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.

For many, if not most, experimental psychologists the study of conditioning and that of human information processing (HIP) represent highly incompatible topics and methodologies with little, if any, overlap. The two areas of study are so divergent that they seldom even bother to criticize one another in any useful dialogue. Modern HIPers by and large discard years of research in traditional learning theory as uninteresting, staid, and just plain old-fashioned. Watson and Hull are hardly regarded as models to be emulated by this group of researchers. By the same token, many traditional researchers in conditioning and learning theory are puzzled by this new wave of information processing research. They cannot see how it is substantially different from the kinds of experiments they have been doing all along and tend to regard the changes of the last two decades or so as changes in terminology rather than anything really important and new. Indeed, at the 50th annual meeting of the Midwestern Psychological Association, one past president showed in some detail how contemporary terms could be easily translated into S-R concepts of yore. This presentation was enthusiastically received by those experimenters old enough to remember the work of Hull, while younger investigators for whom Hull was primarily one topic in a required History of Psychology course, listened with unbelieving ears making caustic sotto voice remarks to their colleagues.

Kantowitz, B. H. (in press). Defining and measuring pilot mental workload. *Proceedings of the 1987 Mental-State Estimation Workshop*. Williamsburg, VA: NASA-Langley Research Center.

The best practical tool is a good theory. Models of attention based upon a single pool of limited capacity offer an excellent starting point for measuring pilot mental workload. Thus, I define mental workload as an intervening variable similar to attention.

Objective measures are preferable for measuring pilot mental workload. Secondary tasks, especially choice-reaction time, are extremely useful in this regard. Psychophysiological tasks will be more useful in the near future as theoretical models are refined.

Kantowitz, B. H. (in press). Mental workload. In P. A. Hancock (Ed.), *Human Factors Psychology*. Amsterdam: North Holland Press.

This chapter represents my attempt at catharsis wherein I purge myself of many of the good and bad thoughts I have entertained concerning mental workload, both as a pragmatic and as a scientific concept. Furthermore, the next time I am again asked "What is mental workload?" I can simply thrust a copy of this chapter into the beseeching hands of the questioner and beat a hasty retreat.

Kantowitz, B. H. and Casper, P. A. (in press). Human workload in aviation. In E. Wiener and D. Nagel (Eds.), *Human Factors in Aviation*. New York: Academic Press.

This chapter discusses the importance of human workload in aviation systems in the areas of safety, crew size, automation, and certification. A definition of workload is addressed through a comparison of physical and mental workload, and by reviewing several measurement techniques including subjective ratings, secondary tasks, and biocybernetic indices. The authors then relate workload to the psychological concept of attention and suggest how theories of attention can help solve some of the problems in defining and measuring workload. A series of experiments conducted at NASA which use the theoretical concepts of attention to measure workload in a flight simulator is examined. This is followed by a selective review of empirical workload studies using a variety of measurement techniques. Finally, the chapter ends with a brief discussion of prospects for workload research and its pragmatic application to the aviation industry.

Kramer, A. F. (in press) Event-related brain potentials. In B. Christie and A. Gale (Eds.) *Psychophysiology and the Electronic Workplace*. New York: Wiley and Sons.

The paper provides a brief glimpse of the pitfalls and potential advantages of employing ERPs in the assessment of human performance and cognition. There are many situations in which traditional measures can provide adequate answers to our questions, thereby rendering the costly and time consuming ERP methodology unnecessary. However, there are other cases in which the issues have proven difficult to resolve with the current battery of measurement techniques. It is in these situations that ERPs can be most probability employed.

Kramer, A. F. and Strayer, D. L. (in press). P300 operating characteristics: Performance/ERP analysis of dual-task demands and automaticity. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

The present study examines the attentional requirements of automatic and controlled processing. The amplitude of the P300 component of the ERP was used as a metric of the attentional resources invested in a pair of tasks. Subjects performed two tasks (a Sternberg memory-search task and a recognition running-memory task) both separately and in dual-task conditions. Two stimuli-response mapping conditions were employed: consistent mapping (CM) and varied mapping (VM). Processing priority was manipulated between the two tasks by instructions. Subjects received extensive training prior to the experiment. In CM conditions, large P300s were elicited by all events and P300 amplitude was uninfluenced by processing priority. Dual and single task P300 amplitudes were equivalent. In VM conditions, P300 amplitude varied as a function of processing priority, and reciprocity was found between the two tasks under VM conditions. These results support the hypothesis that attentional resources are allocated to automatic processing. When a CM target is presented, attention is automatically allocated to the task. This interpretation is supported by the finding that even when subjects were instructed to ignore the Sternberg stimuli, the presence of a CM target intruded on performance of the concurrent task. This was not the case for CM non-targets, nor for VM conditions. A second purpose of the present experiment was to examine differences in automatic and controlled processing from a chronometric perspective. In CM condition, P300 latency was relatively constant and did not vary as a function of priority. In addition, reaction time preceded P300 latency in all CM conditions and the RT/P300 ratio did not vary as a function of priority. In VM conditions, P300 latency increased with memory load and varied as a function of priority. Furthermore, P300 latency preceded RT, and the RT/P300 ratio varied as a function of priority. As attention was withdrawn from the task, the RT/P300 ratio increased. The results suggest that an efficiency information extraction process emerges following consistent practice. This is similar to the tuning of a perceptual filter and may correspond to the "pop out" effect, where CM targets appear to jump out of the display.

Linde, C. and Goguen, J. and Devenish, L. (in press). *Aircrew Communicative Competence: Theoretical and Pragmatic*. (Final Report for NASA Contract NAS2-12379).

This is the final report of a project studying methods for communications training applicable to both civilians and military aviation personnel, including multiperson crews and teams of single pilot fixed-wing or rotary-wing aircraft. This report reviews a number of theories which have been proposed as relevant for producing training materials on improved communications to be used in aviation contexts, gives criteria for evaluating the applicability of training programs in the aviation context, and applies these criteria to United Airlines' Resources Management Training, as well as to a number of commercially available general purpose training programs. The theories considered in detail are assertiveness training and grid management training. The report examines their theoretical background and the attempts which have been made to validate their effectiveness.

Linde, C., Goguen, J., Finnie, E., MacKaye, S., and Wescoat, M. (in press). Rank and status in the cockpit: Some linguistic consequences of crossed hierarchies. *Proceedings of the Fifteenth New Ways of Analyzing Variation Conference*. Stanford University, October 1986.

This study examines two different types of social stratification which may be present in the same social situation: rank hierarchy and task hierarchy, and demonstrates that each can have a separate effect on two linguistic variables: mitigation, including speech act indirection, and use of term of address, including both names and terms of rank. Using as data videotapes of 14 simulated commercial flights, this study investigates the linguistic consequences of the captain or the first officer being the pilot flying. We find that these two situations, parallel and crossed hierarchies, exhibit different patterns of use of mitigation and term of address. The study thus shows that even a very well defined social hierarchy is not sufficient to explain linguistic variation, and that situational stratification, in this case, task hierarchy, must also be considered.

Liu, Y. and Wickens, C. D. (in press). The effect of processing code, response modality and task difficulty on dual task performance and subjective workload in a manual system. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

We report here the first experiment of a series studying the effect of task structure and difficulty demand on time-sharing performance and workload in both automated and corresponding manual systems. The experimental task involves manual control modes of response (voice or manual). The results provide strong evidence that tasks and processes competing for common processing resources are time shared less effectively and have higher workload than tasks competing for separate resources. Subjective measures and the structure of multiple resources are used in conjunction to predict dual task performance. The evidence comes from both single task and from dual task performance.

Moray, N., Eisen, P., Money, L., and Turksen, I. B. (in press). Fuzzy analysis of skill and rule-based mental workload. In P. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. The Netherlands: Elsevier.

With the introduction of Rasmussen's taxonomy of skill, rule, and knowledge based behavior the question arises of their relative importance as sources of workload. If workload is rated using fuzzy measurement, it can be shown that the ratings approximate an interval scale. Regression models show that the difficulty a task with both skill and rule based components can be predicted from the ratings of the difficulty of the skill and rule based components measured separately. The major source of difficulty is the skill based component with the rule based component modulating the overall task difficulty.

Moray, N., King, B., Turksen, B., and Waterton, K. (in press). A closed-loop causal model of workload based on a comparison of fuzzy and crisp measurement techniques. To appear in *Human Factors*, 29 (1).

Fuzzy and crisp measurements of workload are compared for a tracking task that varied in bandwidth and order of control. Fuzzy measures are as powerful as crisp measures, and can under certain conditions give extra insights into workload causality. Both methods

suggest that workload arises in a system in which effort, performance, difficulty, and task variables are linked in a closed loop. marked individual differences were found. Future work on the fuzzy measurement of workload is justified.

Nagel, D. C. and Hart, S. G. (in press). Helicopter human factors research. *Proceedings of the 1987 NASA/Army Rotorcraft Technology Conference*. Moffett Field, CA: NASA Ames Research Center.

Helicopter flying tasks are among the most demanding of all human-machine interaction. The inherent manual control complexities of rotorcraft are made even more challenging because of the small margins of error created by the proximity of terrain. Accident data recount numerous examples of unintended conflict between helicopters and terrain and attest to the perceptual and control difficulties associated with low-altitude flight tasks. NASA Ames, in cooperation with the U.S. Army Aeroflightdynamics Directorate, has initiated an ambitious research program aimed at reducing the difficulties and increasing the margins of safety for helicopter operations, both civilian and military ones. The program is broad, fundamental, and focused both on the development of scientific understandings and technological countermeasures.

This paper focuses on research being conducted in several areas. First, studies of workload including its assessment, prediction, and the validation of being approaches to measurement are described. Next, we discuss research done to understand the decomposition of flying tasks and the relationship of workload and training. Since the visual sense is so significantly involved in helicopter flying, and particularly NOE flight, we next describe studies that are being done to understand what visual cues are important and the ways that various artificial sensors and artificial visual aids affects the perception and use of such cues. Finally, the broad topics of displays and the development of effective pilot/automation interfaces are discussed. A companion paper (Hart, Hartzell, Voorhees, Bucher, and Shively, 1987) describes a second program at Ames that attempts to integrate the information, understanding, and technology described here into specific requirements for advanced rotorcraft development programs.

Pashler, H. and Johnston, J. C. (in press). Dual-task interference and response grouping in temporally overlapping classification tasks: Validating single-channel predictions. Submitted to *Journal of Experimental Psychology: Human Perception and Performance*.

When the stimuli from two tasks arrive in rapid succession (the overlapping tasks paradigm), response delays are typically observed. Two general types of models have been proposed to account for these delays. *Postponement models* suppose that processing stages in the second task are delayed due to a single-channel bottleneck. *capacity-sharing models* suppose that processing on both tasks occurs at reduced rates because of sharing of common resources. Postponement models make strong and distinctive predictions for the behavior of variables slowing particular second-task stages, when assessed in single- and dual task conditions. In Experiment 1, subjects were required to make manual classification responses to a tone (S1) and a letter (S2), presented at stimulus onset asynchronies of 50, 100 and 400 msec, making R1 responses to S1 as promptly as possible. The second response, R2, but not R1, was delayed in the dual task condition, and the effects of two S2 variables (degradation and repetition) on R2 response times in dual- and single-task conditions closely matched the predictions of a postponement model with a

processing bottleneck at the decision response-selection stage. In Experiment 2, subjects were encouraged to emit both responses in tandem. Use of this response grouping procedure had little effect on the magnitude of R2 response times, or on the pattern of stimulus factor effects on R2, supporting the hypothesis that the same underlying postponement process was operating. R1 response times were, however, dramatically delayed, and were now affected by S2 difficulty variables. The results provide strong support for postponement models of dual-task interference in the overlapping tasks paradigm, even when response times are delayed on both tasks.

Pepitone, D. D. and Shively, R. J. (in press). Pilot workload prediction. *1987 SAE Conference Proceedings*. Long Beach, CA: Society for Automotive Engineers.

Workload prediction requires that we know how a pilot experiences inflight workload. The method by which a pilot estimates inflight workload is most likely very complex, with the workload experiences based not only on task combinations (e.g., communication tasks paired with psychomotor or navigation tasks), but also various combination of tasks paired with various flight situations (e.g., routine versus emergency situations). In a general sense there may be two ways in which variables combine to influence subjective workload ratings. One may be the total amount of work to be done, and the second way may be the rate at which the work is done. Three experimental flight simulations were utilized to test these two hypotheses. The results indicated that subjective pilot workload ratings were sensitive to the rate at which work is accomplished in a general aviation flight simulation. A linear regression model was utilized to examine the data. The results indicated that as the rate of work increased so did the workload ratings. It was concluded that subjective workload ratings are sensitive to the rate at which work is done rather than the total amount of work accomplished.

Tsang, P. S. and Vidulich, M. A. (in press). Time-sharing visual and auditory tracking tasks. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

Multiple resource theory suggests that distributing demands over separate resources will reduce resource competition and improve time-sharing efficiency. A recent hypothesis, however, suggests that the benefits of utilizing separate resources for the time-shared tasks may be mitigated if the two tasks are integrated. The present experiment examined the benefits of distributing the input demands of two tracking tasks as a function of task integrality. Visual and auditory compensatory tracking tasks were used. Time-sharing two tracking tasks with the same order of control is said to be more integrated than with different orders of control. Results show that presenting the two tracking signals in two input modalities did not improve time-sharing efficiency. This was attributed to the difficulty insensitivity phenomenon. Whether utilizing the same control dynamics between the time-shared tasks could generate an integrality effect was unclear from the present data. A continuous auditory task that could offer comparable spatial information as the visual counterpart was proposed to be valuable for studying attentional processes, information display alternatives, and workload assessment.

Vidulich, M. A. (in press). The cognitive psychology of subjective mental workload. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload*. The Netherlands: Elsevier.

ORIGINAL PAGE IS
OF POOR QUALITY

The trend toward automated systems has created a need for evaluating mental workload in environments with little measurable performance. Subjective workload assessment is reviewed in terms of its suitability for such evaluations. The results reviewed suggest that subjective assessment, as currently practiced, can provide a valid assessment of the overall workload inflicted on an operator's working memory, but is relatively insensitive to demands outside that component of the human information processing system. Also, performing multiple tasks concurrently seems to render subjective workload assessments somewhat insensitive to changes in just one of the tasks.

Vidulich, M. A. and Pandit, P. (in press). Consistent mapping and spatial consistency in target detection and response execution. *Proceedings of the Fourth Mid-Central Ergonomics/Human Factors Conference*. Urbana, IL: Mid-Central Psychological Association.

Among the most robust findings in the experimental psychology literature has been the facilitating effect of consistently mapped (CM) training on human performance. Given a set of stimuli that are invariably presented as targets against the background of a consistent set of distractors, training will permit subjects to make seemingly effortless target detections despite high speed presentations, and regardless of memory set size. The present experiment expanded on previous studies in two ways. First, in addition to the usual CM manipulation, "spatial consistency" was manipulated. Half of the CM targets appeared in only specific locations, while the remaining potential targets could appear in any location. Second, the response execution demands were increased by requiring subjects to perform a Fitts' task to indicate the target's location. Furthermore, the spatially consistent targets were paired with a specific Fitts' Index of Difficulty (ID); other CM targets were not. Therefore, in the present study, spatial consistency refers to both a consistent location for the stimulus presentation and a consistent Fitts' ID. The study was designed to test the effects of both the traditional CM training and the spatially consistent training on target detection and response execution performance. Preliminary analyses indicate that spatial consistency had a powerful effect on target detection performance. Detection of the spatially consistent targets were faster and more accurate than the other CM targets. Detection performance on the other CM targets was close to that of the spatially consistent targets when they were presented in the same display locations used by the spatially consistent targets. However, the detection of the other CM targets when presented in locations not associated with either of the spatially consistent targets, was only slightly better than the detection of varied mapped targets. These results may indicate that subjects adopted a strategy of focusing attention on the display locations associated with the spatially consistent targets. If so, then the results indicate a strong interaction between the attention allocation strategy and automatic processing. Overall, the data should provide valuable insights on integrating automaticity with both spatial attention allocation and motor learning.

Vidulich, M. A. and Tsang, P. S. (in press). Absolute magnitude estimation and relative judgment approaches to subjective workload assessment. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

Two rating scale techniques employing an absolute magnitude estimation method, were compared to a relative judgment method for assessing subjective workload. One of the absolute estimation techniques used was a unidimensional overall workload scale, and the other was the multidimensional NASA-Task Load Index technique. Thomas Saaty's

Analytic Hierarchy Process was the unidimensional relative judgment method used. These techniques were used to assess the subjective workload of various single- and dual-tracking conditions. The validity of the techniques was defined as their ability to detect the same phenomena observed in the tracking performance. Reliability was assessed by calculating test-retest correlations. Within the context of the experiment, the Saaty Analytic Hierarchy Process was found to be superior in validity and reliability. These findings suggest that the relative judgment method would be an effective addition to the currently available subjective workload assessment techniques.

Vincente, K. J., Thornton, D. C. and Moray, N. (in press) Spectral analysis of sinusarrhythmia: A measure of effort. Submitted to *Human Factors*.

To resolve the uncertainty and disagreement that currently exists in the field of mental workload, a unified research approach is required. It is argued that a promising path would be to identify the various dimensions of mental workload, and then to develop a metric for each of these. The present study focussed on the dimension of mental effort. In particular, the validity of spectral analysis of sinusarrhythmia as a measure of mental effort was investigated using a psychomotor task. The strong correlation observed between the physiological measure and subjective ratings of effort suggest that spectral analysis of sinusarrhythmia is an accurate measure of operator effort. Results also indicated that the intensity of effort invested by subjects could not be inferred from objective task difficulty or performance. Thus, it is important that a measure of effort be included in experiments investigating mental workload. Future research will be directed at developing a continuous measure of operator effort by implementing the physiological measure on-line.

Wickens, C. D. (in press). Attention in aviation. *Proceedings of the Fourth Conference on Aviation Psychology*. Columbus: Ohio State University.

This paper describes the relevance of four principles or mechanisms of human attention to the design of aviation systems and the performance of pilots in multi-task environments. The principles relate to resources, confusion, integration, and tunneling. Relevance to such issues as workload prediction and measurement, control-display integration, and the use of voice and head-up displays are described.

Wickens, C. D., Fracker, L., and Webb, J. (in press). Cross-modal interference and task integration: Resources or preemption switching? *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

Data are reviewed from experiments that have contrasted intra-modal (visual-visual) information presentation with cross-modal (visual-auditory) presentation. Five different processing mechanisms that are operating in dual stimulus tasks are described, and it is concluded that in studies where visual scanning is not required, cross-modal effects are of two classes. When the visual task is continuous (tracking), a discrete auditory stimulus will *preempt* tracking performance relative to a discrete visual stimulus, leading to an effective shift in allocation bias. When both tasks are discrete, the data regarding the relative advantages of cross- vs. intra-modal interference are ambivalent.

**ORIGINAL PAGE IS
OF POOR QUALITY**

Yeh, Y. -Y. and Wickens, C. D. (in press) Dissociation between performance and subjective workload. Accepted for publication in *Human Factors*.

A theory is presented to identify sources that produce dissociations between performance and subjective measures of workload. The theory states that *performance* is determined by (1) amount of resources invested, (2) resource efficiency, and (3) degree of competition for common resources in a multidimensional space described in the multiple-resources model. *Subjective perception of workload*, multidimensional in its nature, increases (1) with greater amounts of resource investment and (2) with greater demands on working memory. Performance and subjective workload measures dissociate (1) when greater resources are invested to improve performance of a resource-limited task, (2) when demands on working memory are increased by time-sharing between concurrent tasks or between display elements, and (3) when performance is sensitive to resource competition and subjective measures are more sensitive to total investment. These dissociation findings and their implications are discussed, and directions for future research are suggested.



Report Documentation Page

1. Report No. NASA TM-100016		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Research Papers and Publications (1981-1987): Workload Research Program				5. Report Date August 1987	
				6. Performing Organization Code	
7. Author(s) Compiler Sandra G. Hart				8. Performing Organization Report No. A-87196	
				10. Work Unit No. 505-67-51	
9. Performing Organization Name and Address Ames Research Center Moffett Field, CA 94035				11. Contract or Grant No.	
				13. Type of Report and Period Covered Technical Memorandum	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546-0001				14. Sponsoring Agency Code	
15. Supplementary Notes Point of Contact: Sandra G. Hart, Ames Research Center, M/S 239-3 Moffett Field, CA 94035 (415) 694-6072 or FTS 464-6072					
16. Abstract This document contains an annotated bibliography of the research reports written by participants in NASA's Workload Research Program since 1981. It represents the results of theoretical and applied research conducted at Ames Research Center and at universities and industrial laboratories funded by the program. The major program elements included: (1) Developing a fundamental understanding of the concept of workload, (2) providing valid, reliable, and practical measures of workload, and (3) creating a computer model to predict workload. The overall goal is to provide workload-related design principles, measures, guidelines, and computational models. The research results are transferred to user groups by establishing close ties with manufacturers, civil and military operators of aerospace systems, and regulatory agencies; publishing scientific articles; participating in and sponsoring workshops and symposia; providing information, guidelines, and computer models; and contributing to the formulation of standards. In addition, the methods and theories that have been developed have been applied to specific operational and design problems at the request of a number of industry and government agencies.					
17. Key Words (Suggested by Author(s)) Workload assessment Workload prediction Human performance			18. Distribution Statement Unclassified-Unlimited Subject Category - 53		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of pages 122	22. Price A06