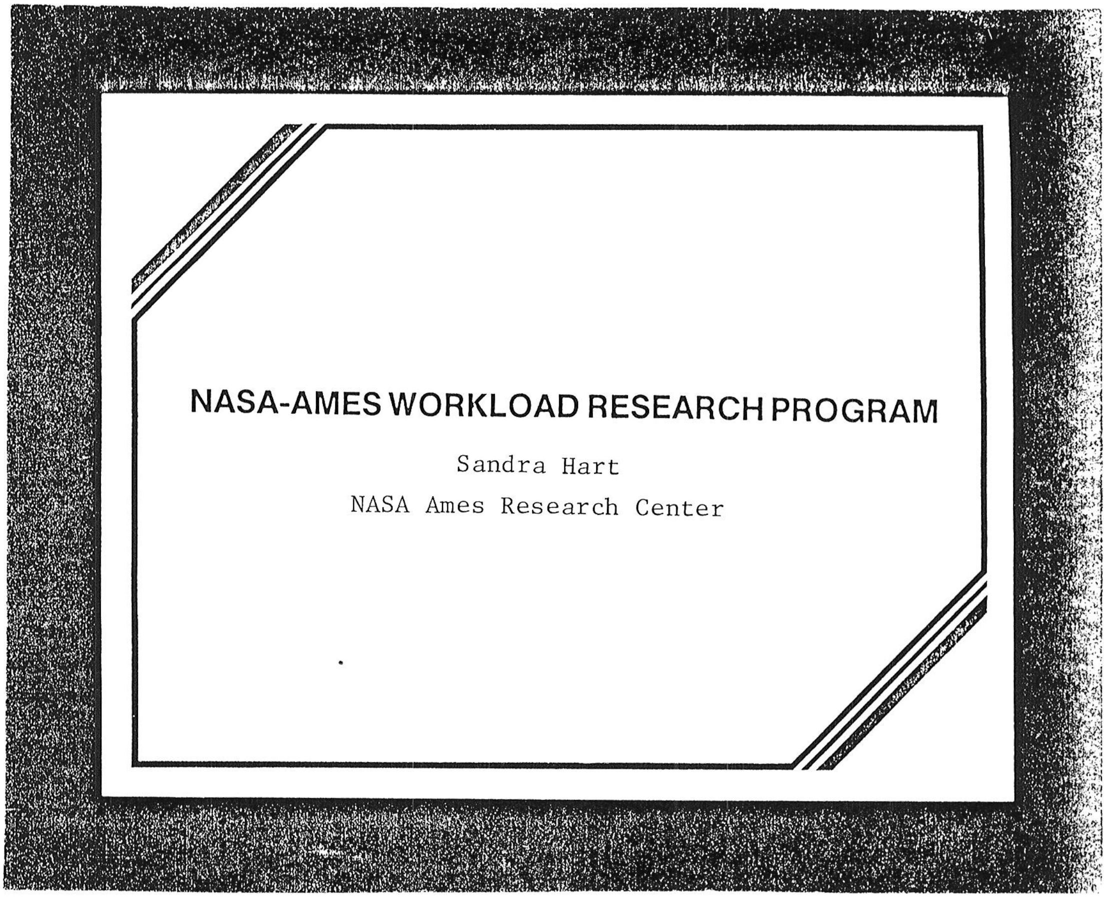


During the next hour, I will describe the purpose, philosophy, structure, and some of the accomplishments of the Human Performance Research Group of the Aerospace Human Factors Research Division. I will try to demonstrate the flow of information from generic, theoretical research to specific space-station related applications.

Although an increasing emphasis has been placed on providing computer-based automation in every phase of modern systems, the decision has been made that man will continue to play a central role in space station operations. Humans have capabilities beyond those of the most sophisticated computer systems and their flexibility and adaptibility provides a unique asset in such a remote environment. The activities that will be performed in the Space Station range from direct control of spacecraft (e.g., the orbiter, the orbital transfer vehicle, and the manned maneuvering unit) to indirect control (e.g., the orbital maneuvering vehicle and the remote manipulator arm), to housekeeping activities and the conduct of scientific experiments. Each will require specialized training, take a certain amount of very limited and precious time and will have some associated human (e.g., workload) and payload cost.

The space station provides a unique situation in which teams of astronauts, scientists, and technicians will live and work in an unfamiliar environment for prolonged periods of time. Space flight has traditionally required high levels of performance in relatively stressful environments. The stressors may include isolation from familiar work and living surrounding, physiological discomfort associated with weightlessness, and potentially high levels of workload. Major changes in the U. S. Space Program may precipitate additional problems, such as longer missions, hetereogeneous crews, more varied and complex tasks, and an expected decrease in the training provided for individual crewmembers. The increased emphasis on space commercialization will require crewmembers to exhibit new levels of productivity.

Even though previous space missions have proven to be extremely successful, the available evidence suggests that the performance and reliability of the human elements of aerospace systems is curently lower than that of other elements. Studies of human reliability show that most human-related errors involve inadequate or faulty crew coordination and inadequate or faulty man-machine interface. These problems are soluble. One of the goals of our program is to evaluate ways to predict the impact of performing a large range of tasks on the human operator and to provide guidelines for design and operation to enhance system performance and optimize human behavior and experience.

It is important to assign humans those tasks with which they can excell and to redesign, aide, automate, or eliminate those tasks which they perform poorly, unreliably, or with unacceptably high levels of workload. In addition, the presentation of information and control inputs must be designed so as to optimize human capabilities. In order to accomplish this, predictors and measures of human performance and workload are needed to evaluate the effectiveness of display, control, and automation options so as to maximize the efficiency, effectiveness and reliability of the human element in a man-machine system. This information is required early in the design and construction process, as retrofits and modifications are costly and time-consuming, if not impossible, once the actual construction process of the space station has begun.

Traditional measures of human performance (which focus on lower level, in-the-loop control) may not be applicable for high-level supervisory control tasks nor the measurement of productivity, efficiency, information seeking, decision making or control strategy for teams of operators. In addition, the impact of crewmembers' efforts to accomplish mission requirements on the human operators themselves (e.g., workload) is an important design consideration.

OUTLINE:

- o ORGANIZATION OF PROGRAM
  - PROBLEM/OBJECTIVES/APPROACH
  - RESOURCES
  - COLLABORATION
  - CONCEPTUAL FRAMEWORK
- o CRITERION TASKS
- o PREDICTIVE MODEL
- o ASSESSMENT TECHNIQUES
  - PERFORMANCE
  - PHYSIOLOGICAL
  - SUBJECTIVE
- o VALIDATION/APPLICATION OF TECHNIQUES

Research has been underway at Ames for several years to develop valid and reliable measures and predictors of workload as a function of operator state, task requirements, and system resources. Although the initial focus of this research was on aeronautics, the underlying principles and methodologies are equally applicable to space, and provide a set of tools that NASA and its contractors can use to evaluate design alternatives from the perspective of the astronauts. I will begin by describing the objectives and approach of the research program, the resources used in conducting research, and the conceptual framework around which the program evolved. Next, I will describe the standardized tasks, predictive models and assessment techniques we have developed, and their application to the space program. Finally, I will review some of the operational applications of these tasks and measures.

PROBLEM:

- NONOPTIMAL LEVELS OF WORKLOAD IMPOSED ON THE HUMAN OPERATORS OF ADVANCED SYSTEMS ARE A SIGNIFICANT FACTOR IN THE EFFICIENCY AND SAFETY OF SYSTEM OPERATIONS, OVERALL SYSTEM PERFORMANCE, TRAINING REQUIREMENTS, ADDITIONAL HARDWARE AND SOFTWARE COSTS, CREW COMPLEMENT, AND JOB SATISFACTION.

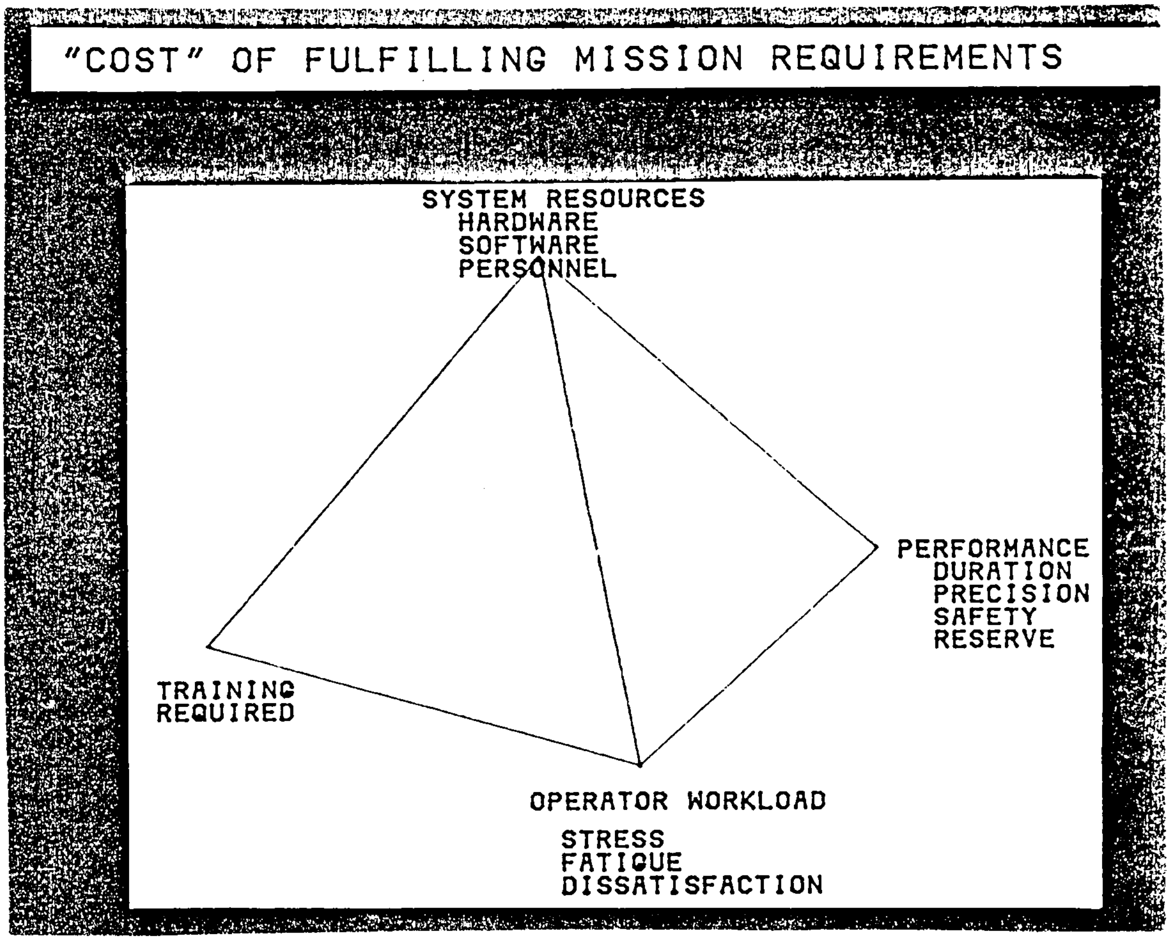
- SINCE WORKLOAD REFLECTS THE INTERSECTION BETWEEN A PARTICULAR OPERATOR PERFORMING A SPECIFIC MISSION, USING THE AVAILABLE HARDWARE, SOFTWARE AND HUMAN RESOURCES, WORKLOAD HAS MULTIPLE CAUSES AND EFFECTS.

- THUS, DIFFERENT WORKLOAD QUESTIONS AND CIRCUMSTANCES REQUIRE DIFFERENT MEASURUREMENT TECHNIQUES.

- STANDARDIZED, VALIDATED, AND SENSITIVE MEASURES ARE NOT YET AVAILABLE TO EVALUATE THE WORKLOAD OF EXISTING SYSTEMS NOR TO PREDICT THE WORKLOAD OF PROPOSED SYSTEMS DURING THE DEVELOPMENT PROCESS.

A resurgence of interest in the field of workload assessment was prompted by the President's Task Force on Crew Complement. It became clear that the question of whether or not two or three crewmembers would be required for the next generation of aircraft could not be answered satisfactorially without a clear concept of what factors affected crew workload, how workload could be measured, how much workload is too much (or too little), the relationship between measures of workload and performance, and the effectiveness of automation in reducing or redistributing workload.

Our initial premise was that nonoptimal levels of workload are a significant factor in efficient and safe system operations, training requirements, required hardware and software, crew complement, and job satisfaction. Since workload reflects the intersection between a particular operator performing a particular mission, using the available hardware, software and human resources, workload may have multiple causes and effects. Thus, different workload-related questions and circumstances require different measurement techniques. Even more important, for practical reasons, is the need for standard, valid, sensitive techniques to predict the workload of proposed systems early in the design process.



The "cost" of fulfilling mission requirements can be conceptualized in many ways. It can be quantified in terms of system resources required; the amount and sophistication of hardware and software required and the number and qualifications of personnel. The cost of the training required for crewmembers to accomplish mission objectives using existing equipment can be quantified as well, as can the cost of failure to meet mission objectives. We define the "cost" to human operators of performing their part in a man-machine system as workload. Workload is more difficult to quantify in objective terms than the other costs of system performance. It's impact may be evalutaed indirectly, however, through lowered levels of performance, additional required resources or training, and operator dissatisfaction. In order to meet mission requirements, there may be a tradeoff between additional resources, additional training or higher levels of workload. If operators are already working at their peak efficiency, then lower levels of performance might have to be accepted or additional system resources provided.

PROGRAM OBJECTIVE:

DEVELOP AND VALIDATE TECHNIQUES TO PREDICT AND ASSESS THE EFFECTS OF TASK DEMANDS, ENVIRONMENT, AND TRAINING ON OPERATOR BEHAVIOR, WORKLOAD, AND PERFORMANCE.

APPROACH:

PERFORM GENERIC RESEARCH TO DISCOVER UNDERLYING PRINCIPLES, DEVELOP AND VALIDATE ASSESSMENT TECHNIQUES, AND CREATE PREDICTIVE MODELS.

PERFORM VEHICLE-SPECIFIC APPLICATIONS OF GENERIC CONCEPTS AND METHODS TO ADDRESS OPERATIONAL PROBLEMS.

Our asumption is that workload is a hypothetical construct that represents the cost to human operators of achieving mission objectives. Thus, our definition is human-centered, rather than task-centered. An operator's experienced workload representes many other factors in addition to the objective demands placed on them. It is not an inherent property of a task but emerges from the interaction between the requirements of the task, the skills and behaviors of an operator, and the circumstances under which the task is performed.

The initial goal of the program was to develop measures and predictors of human workload that took into account all of the relevant factors. Several parallel lines of research were undertaken in which underlying principles were discovered, measurement techniques developed and validated, and predictive models created. Vehicle-specific applications of these generic concepts and methods were performed concurrently to address a variety of operational problems.

NASA/ARMY-SPONSORED WORKSHOP

OBJECTIVE: EXAMINE THE ASSOCIATIONS AMONG WORKLOAD, TRAINING, AND PERFORMANCE. IDENTIFY WHAT IS ALREADY KNOWN AND IMPORTANT ISSUES THAT REQUIRE RESEARCH.

APPROACH: CONDUCT A FIVE-DAY WORKSHOP IN WHICH EXPERTS IN TRAINING, WORKLOAD, AND ADVANCED SYSTEMS WILL BECOME FAMILIAR WITH EACH OTHERS' DISCIPLINES AND CONSIDER SPECIFIC QUESTIONS:

- o WHAT IS THE EFFECT OF EXPERTISE ON WORKLOAD?
- o IMBEDDED TRAINING
- o EXPERT SYSTEMS/COMUPTER-BASED TRAINING
- o INDIVIDUAL DIFFERENCES IN WORKLOAD AND TRAINING

PRODUCT: PUBLISHED BOOK EDITED BY DR. EMANUEL DONCHIN THAT INCLUDES INVITED ADDRESSES AND SUMMARIZES PANEL DISCUSSIONS.

SCHEDULE: 10/86 (FUND GRANT TO UNIVERSITY OF ILLINOIS)
1/86 (CONFERENCE)
9/86 (PUBLICATION)

The initial focus of the research was on assessment. The focus moved toward predition as the theoretical problems associated with assessing workload in existing systems were resolved. I will describe the results of this research in greater detail in a moment. More recently, our focus has been on training. Specifically, we wish to investigate the interrelationships among workload, training, and performance in highly automated systems, such as the LHX helicopter and the space station.

The focal point of this area of research is a workshop sponsored by NASA that will be held in January. The workshop participants will consider how to quantify and predict performance and workload changes as training progresses, and, conversely, to determine the role of workload in training effectiveness. The proceedings of this workshop will be published in a book for public dissemination. The specific focus of the discussions will be on the two vehicles that represent two workload and environmental extremes faced by technology - - single-pilot, nap-of-the earth helicopter flight at night during the performance of Army missions and Space Station operations. Training may well emerge as a significant problem area in space station operations. Due to new mission goals and characteristics, it is anticipated that the training time allowed for space station operators will be reduced. Some of the training now accomplished on the ground may be performed in orbit and recurrent training may be required on orbit due to the extended mission durations. More effective and efficient training programs, particularly those that focus on understanding and operating highly automated subsystems, will be needed to maintain workload and performance at acceptable levels.

RESEARCH GRANTS FUNDED BY THE PROGRAM

* VIRGNINA POLYTECHNIC INSTITUTE	WIERWILLE
* ARIZONA STATE UNIVERSITY	DAMOS
* UNIVERSITY OF CALIFORNIA, LOS ANGELES	LYMAN
* OHIO STATE UNIVERSITY	JENSEN
* SAN JOSE STATE UNIVERSITY	JORDAN
* MASSACHUSETTS INSTITUTE OF TECHNOLOGY	SHERIDAN
* U.S. AIR FORCE ACADEMY	SWINEY
* SANTA CLARA UNIVERSITY	SWEENY
* UNIVERSITY OF ILLINOIS	WICKENS, KRAMER
* PURDUE UNIVERSITY	KANTOWITZ
* UNIVERSITY OF TORONTO	MORAY
* TECHNION, ISRAEL INSTITUTE OF TECHNOLOGY	GOPHER
* BEHAVIORAL INST. TECHNOLOGY AND SCIENCE	KANTOWITZ, TOWNSEND
* UNIVERSITY OF SOUTHERN CALIFORNIA	HANCOCK
* WAYNE STATE UNIVERSITY	FRANKEL
* STANFORD UNIVERSITY	CALFEE

RESEARCH CONTRACTS FUNDED BY THE PROGRAM

* GENERAL PHYSICS CORPORATION	GOMER
* STRUCTURAL SEMANTICS	LINDE, GOGUEN
* STANFORD RESEARCH INSTITUTE	CHESNEY
* DOUGLAS AIRCRAFT COMPANY	BIFERNO
* SEARCH TECHNOLOGY	ROUSE

Our program represents an active collaboration between inhouse research, joint research with other government agencies and industry, and research funded through grants and contracts. The personnel involved in the program include psychologists, pilots, and engineers. The facilities used range from laboratory settings to part-task simulations, full-mission simulations, and inflight experiments. The research efforts differ with respect to theoretical perspective, assessment techniques used, research facilities, and focus (theoretical or applied, prediction or assessment). For each critical area, several different lines of research have been undertaken. Coodination and integration has been accomplished though publications and scientific presentations, meetings, and shared experimental tasks and measurement techniques.

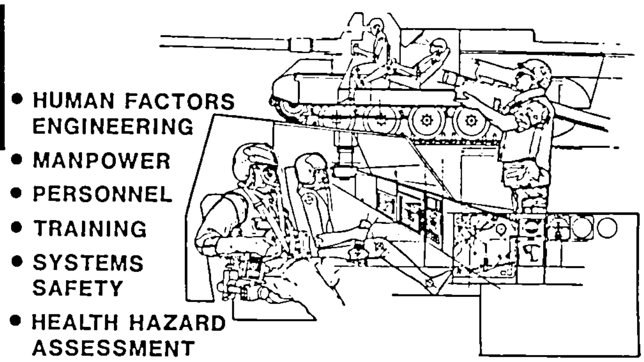
INTERACTIONS WITH OTHER AGENCIES: COLLABORATIVE RESEARCH

* ARMY (CDEC)	SCOUT II Helicopter Experiment
* ARMY (AVSCOM)	COBRA/Pilot Night Vision System Inflight Training 1 vs 2 Pilot (ADOCS Simulation in VMS) ARTI Contractor Simulations-Government scenario
* NASA-JSC	Space Suit Comparison RMS Workload Prediction/Evaluation
* FAA	TCAS Workoad Evaluation (MVSRF B-727 simulator)
* Navy (NATC)	Tilt-rotor Workload Evaluation
* Air Force (Brooks)	Pilot Recertification Test Battery
* British CAA	North-Sea Oil Operations Workoad Evaluation

We have played a support role in a number of simulation and inflight experiments conducted by outside organizations. In general, we provided workload assessment methodologies and application procedures to assist these organizations in addressing oeprationaly relevant workload-related problems.

MANPRINT

MANPOWER & PERSONNEL INTEGRATION



NASA CONTRIBUTION:

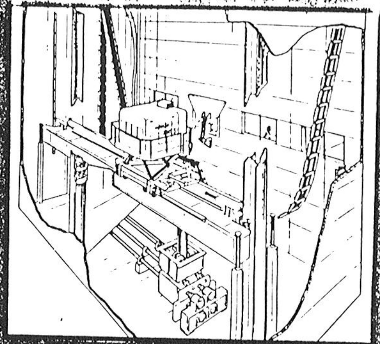
- o BRIEFING: OVERVIEW OF WORKLOAD AND PERFORMANCE ASSESSMENT RESEARCH
- o COURSE SYLLABUS: BASED ON NASA WORKLOAD REPORT
- o COMPUTER-BASED TRAINING PROGRAM: NASA "EXPERT" SYSTEM FOR SELECTING WORKLOAD ASSESSMENT METHODOLOGY

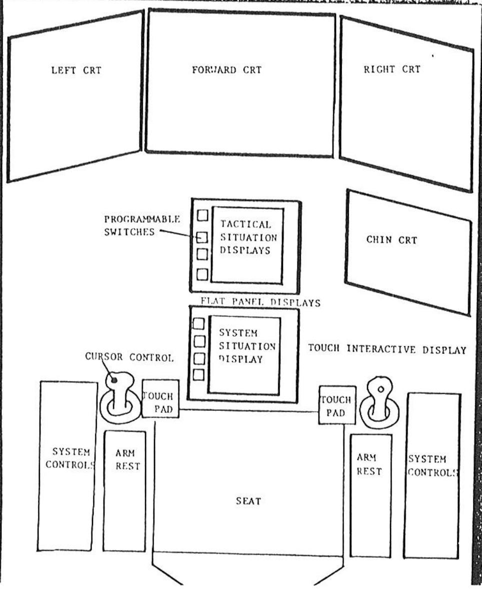
Operational validity and applicability have been insured by frequent involvement in addressing operational problems posed by members of other organizations. One example of such involvement is the role that we played in the development of the Army MANPRINT course. This program represents a major effort by the Army to integrate human factors issues, manpower and personnel, and training into the materiel acquisition process. The results of our research provided the foundation for the course presented by the Army to familiarize Army managers with human factors engineering and several of the programs developed at Ames will be used as teaching aides.

COLLABORATIVE RESEARCH:
ADVANCED DIGITAL OPTICAL CONTROL
SIMULATION (ADOCS)

OBJECTIVE:

(1) COMPARE ONE vs TWO PILOT WORKLOAD
(2) COMPARE WORKLOAD OF DIFFERENT COMBAT MISSIONS
(3) EVALUATE WORKLOAD IMPACT OF DIFFERENT LEVELS OF AUTOMATION

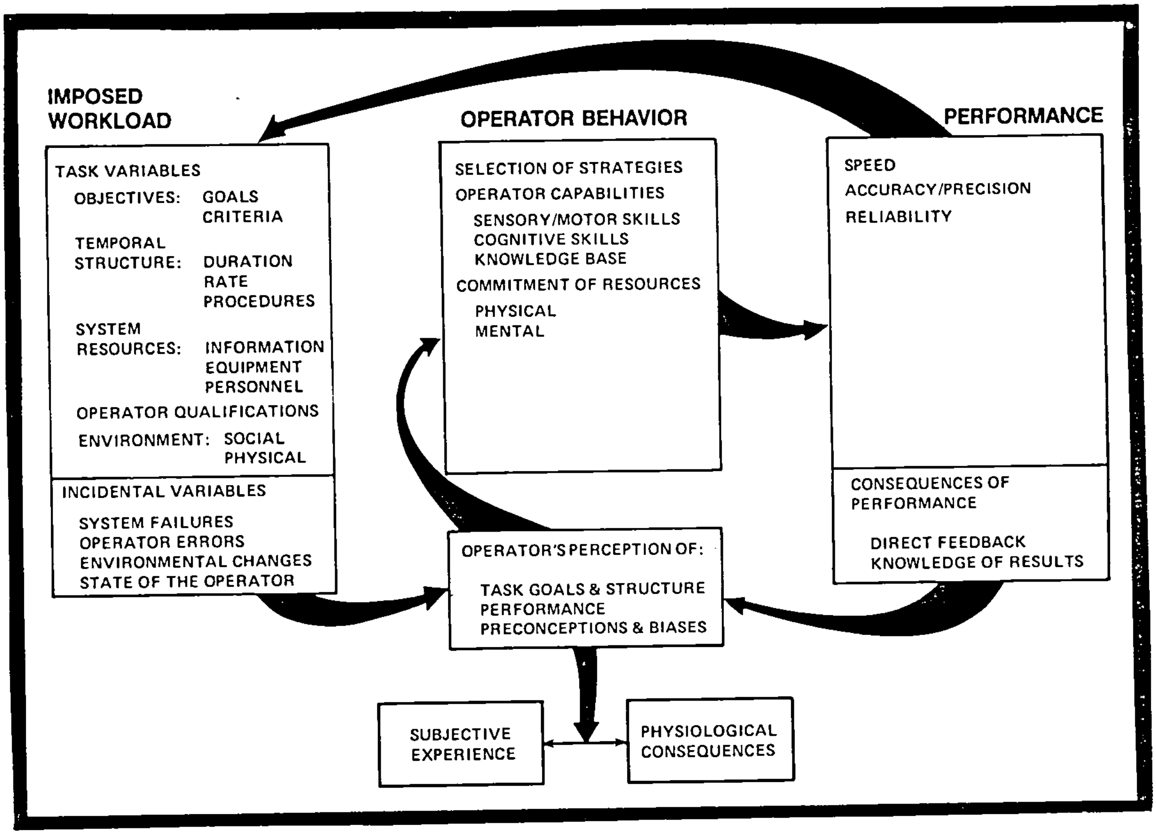




APPROACH:

(1) CONDUCT SIMULATED NOE COMBAT MISSIONS IN VMS
(2) DISPLAYS: HMD, TSD, SMD, TOUCH PANEL, BUTTON I/O
(3) CONTROLLERS: CONVENTIONAL, SIDEARM
(4) WORKLOAD MEASURES:
    INFLIGHT AND POSTFLIGHT RAT
    HEART RATE AND VARIABILITY
    HOVER/BOB-UP TIME ESTIMATES

One example of such joint research is a recent simulation which we completed with the Army Aeroflightdynamics Division. The goal of this study was to compare the workload of pilots flying one- or two-pilot configurations with different levels of automation. The tasks represented missions that an LHX-type helicopter might perform in the 1990s. The flights were performed in the Ames Vertical Motion Simulator using the Advanced Digital Optical Control Simulation (ADOCS).



As I mentioned before, the focal point of the program was a conceptual model in which task-related, behavior-related, and operator-related variables were related to each other. Imposed workload refers to the situation encountered by a specific operator or team of operators in performing a task. The intended demands of a task are created by its objectives and performance criteria, temporal structure, system resources provided and the environment in which it is performed.

Task objectives are particularly critical because they determine the target performance levels that operators attempt to achieve. The temporal structure of the task refers to the length of time available to perform the task or subtask elements, the degree to which task elements overlap in time, the procedures and organization, and the degree to which operators can select which tasks to perform and in which order. The objectives and temporal structure of a task create the task requirements. This can be distinguished from the workload associated with the system resources provided to the operators to perform such a task.

System resources refer to the information, equipment, controls, displays, and personnel that are provided to assist the operator in performing the task. System resources include automation that has become such an important element in most advanced systems. A major focus of our research program has been to investigate the workload-impact of different types of automation on operator workload. In general, the trend has been to reduce the physical workload of operators and to remove them from in-the-loop control activities, but often at the cost of an increase in mental workload. An additional concommitant of automation has been to alter the nature and impact of operator errors - - relatively "minor" typographical errors can lead to extremely grave consequences that are difficult to detect becasue the operator is not sufficiently integrated into the performance of the system.

The environment can have a significant effect on operator workload and performance. The social environment, that is crew interactions, leadership styles, group dynamics, can all play a significant in the safe and efficient functioning of a crew. This particular issue will become particularly salient in space station operations, where crew members live and work together in a very confined environment for a prolonged period of time. The physical environment refers to the workstation layout, personal space, climate, threat from man-made or natural sources.

Each time a particular task is performed by a specific operator, incidental variables may occur that can alter the workload demands of the task either subtly or substantially. In this regard, the primary focus of our research efforts has been to examine the role of system failures and operator errors on subsequent task performance and crew workload. We consider errors to be a potent source of workload rather than an indicator of workload. The disruption caused by errors is particularly acute for well-trained operators, as they must step out of over-learned, automatic patterns of behavior to diagnose and solve the error and then continue with the interrupted activities with conscious attention.

System response refers to the behavior and accomplishments of a man-machine system. Operators are motivated and guided by the imposed demands, but the strategies selected and effort exerted reflects the operators perception of what it required of them. In most tasks, a variety of strategies are possible and different tasks, obviously, required different skills and capabilities. Thus, the role of human behavior in workload can be complex. Physical effort is the easiest to conceptualize and measure, but its contribution to advanced systems in diminishing. The problems associated with physical effort exerted in zero-G environments should be relatively unique, as the astronauts cannot rely on highly overlearned (and thus automatic) patterns of motor behaviors learned in a one-G environment. This source of workload - - that is the conscious attention to physical activities that are normally performed without conscious attention should be relatively great early in a mission, but should be reduced as time on orbit increases, and new patterns of response are developed. Mental effort serves as a potent intervening variable between measurable stimuli and measurable responses but it is difficult to quantify directly. It is unlikely that this aspect of human workload should be affected significantly by a zero-G environment, except for those aspects involved with motor control and spatial orientation.

Performance represents the product of the operators' actions and the limitations, capabilities and characteristics of the system controlled. Performance feedback provides information to the operators about their success in meeting task requirements, the appropriatness of the strategies selected, and the level of effort exerted, allowing them to modify their behavior to achieve more acceptable levels. We have examined performance from two perspectives: (1) As an indicator of the degree to which operators were able to satisfy task requirements and (2) As an indicator of the cost incurred by the operator in doing so. Performance levels tend to remain fairly constant as long as the task requirements remain within the oeprator's capabilities. In this case, performance measures do not reflect the increasing levels of effort associated with meeting progressively increasing task demands. When performance requirements exceed operators' capabilities, or they lower their performance standards, decreasing levels of performance may in fact reflect the existence of higher levels of workload.

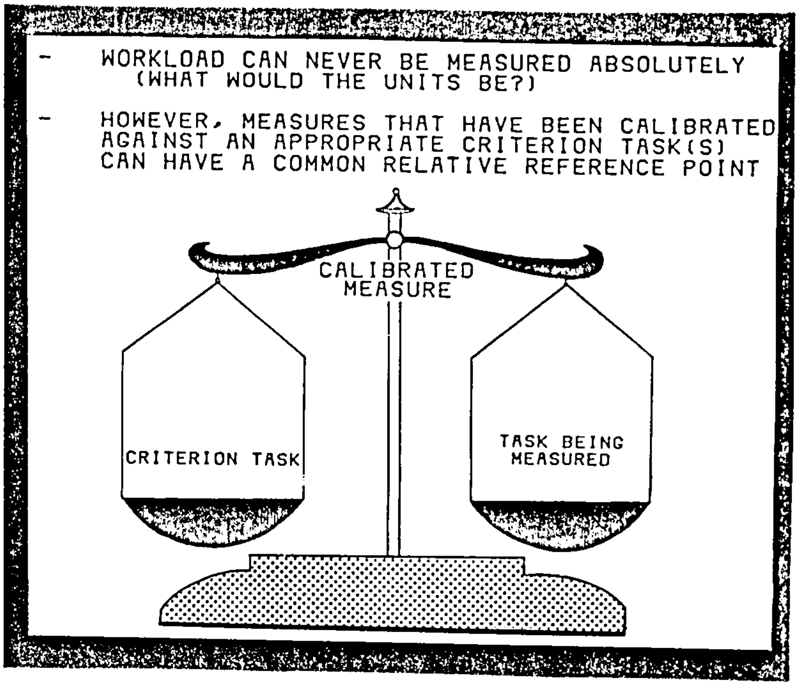
The consequences of performing a task on an operator can be physiological or subjective. Since operators may not be aware of every task variable, the processes that underly their decisions and actions, or the influence of preconceptions about the task, workload experiences may not reflect all of the relevant factors and may, in fact, reflect some that are irrelevant. Thus, we draw a distinction between the level of workload that a system designer intended to impose on an operator, the responses of a specific man-machine system to the task, and the operators' subjective experiences. The importance of subjective experiences extends beyond their association with subjective ratings, however. The phenomenological experiences of human operators affects subsequent behavior, and thus, performance. If operators consider the workload of a task to be excessive, they may adopt strategies that are appropriate for high workload situations (such as shedding tasks, hurrying, or accepting lower levels of performance) and they may experience psyiological or psychological distress.

One example of a misperception of task requirements was presented to us by JSC as a problem requiring an experimental solution. The mission commander on an early Shuttle flight reported experiencing "time compression" during approach and landing - - that is the feeling that time was passing too quickly. One suggestion was that experiencing zero-G had somehow disrupted his ability to perceive the passage of time accurately. The more likely explanation, based on a series of experiments, was that failures of time perception is a common concommitant of stress and high levels of workload.

Physiological responses may reflect momentary responses to task demands (such an elevated heart rate or pupil dilation) or relatively long term effects following prolonged exposures. It might be expected that this aspect of operator's responses to workload might be relatively more extreme in orbit, as task-related stressors might interact with environmental stressors associated with zero-G.

CRITERION TASKS DEVELOPED AT AMES:

- o FITTSBERG
- o POPCORN
- o MULTI-COCKPIT SIMULATION
- o STANDARD FLIGHT SCENARIO MODEL



The fact that workload validation procedures are often circular presents a significant problem in the development and validation of candidate workload measures. since there is no objective standard against which a measure can be compared, the decision of whether or not it is sensitive is often made ad hoc. That is, if the measure varied in accordance with the supposed levels of workload imposed by the task, the assumption is that it is sensitive, and if it does not, it may either indicate that the measure was not sensitive or that the experimenter did not, impose the intended levels of workload.

For this reason, we have developed a set of "criterion tasks", for which standardized levels of workload can be created according to well-known psychological principles. These tasks represent stylized versions of the activities that operators normally perform in advanced systems. Candidate measures or models can then be compared against known workload levels imposed by these tasks. I will describe two such tasks.

CRITERION TASKS: FITTSBERG

OBJECTIVE:

DESIGN A SIMPLE, RELIABLE, AND FLEXIBLE LABORATORY TASK IN WHICH TASK ELEMENTS ARE FUNCTIONALLY RELATED BUT:

(1) RESPONSE SELECTION AND RESPONSE EXECUTION DIFFICULTY CAN BE MANIPULATED INDEPENDENTLY

(2) PERFORMANCE ON SUBTASK ELEMENTS CAN BE MEASURED INDEPENDENTLY

APPLICATIONS:

(1) IDENTIFICATION OF SUBTASK ELEMENTS TO AUTOMATE

(2) DISPLAY MODALITY (AUDITORY/VISUAL)

(3) DISPLAY FORMAT (SPATIAL/VERBAL/NUMERIC)

(4) PREDICTION OF COMPLEX TASK PERFORMANCE

(5) SUBJECTIVE ASSESSMENT OF NON-HOMOGENEOUS INTERVALS

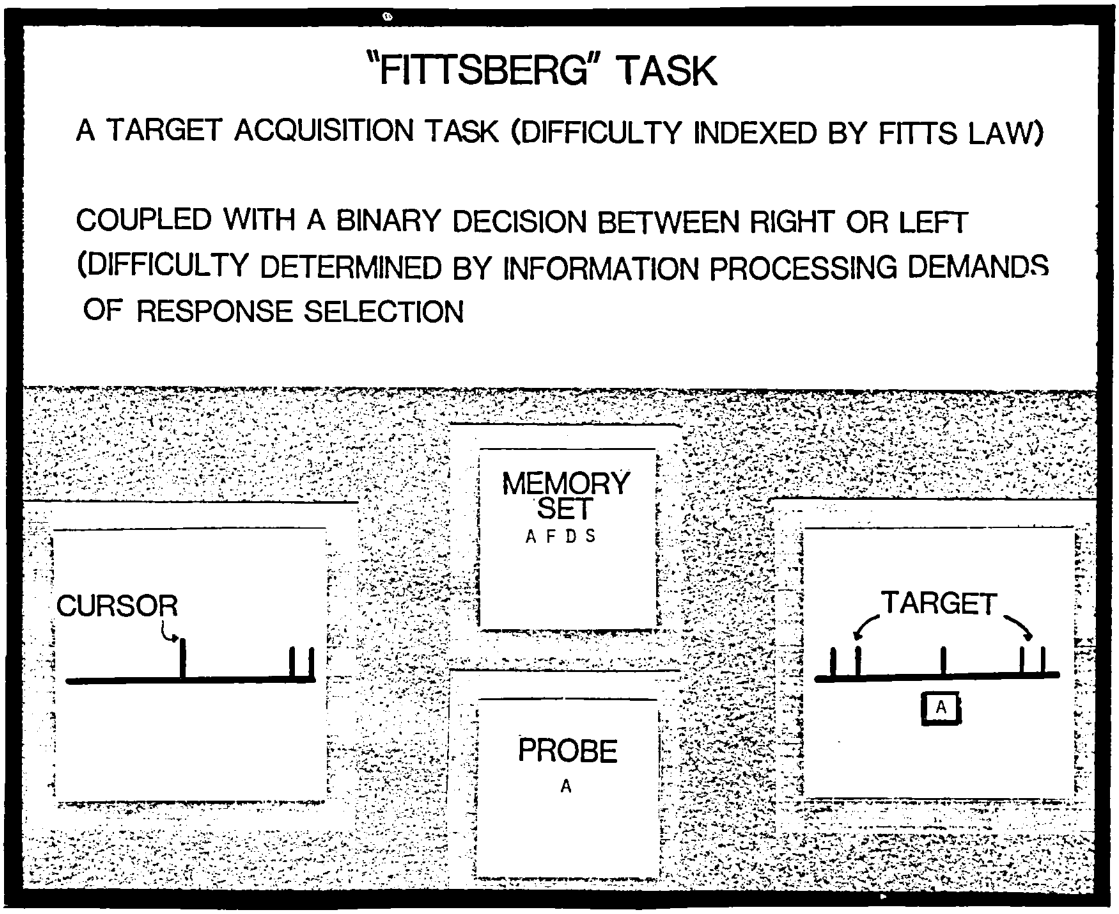
(6) IMMEDIATE vs RETROSPECTIVE WORKLOAD EVALUATION

(7) ASSOCIATION AMONG MEASURES OF WORKLOAD AND PERFORMANCE

(8) BASIS OF SPACE SUIT EVALUATION TEST BATTERY

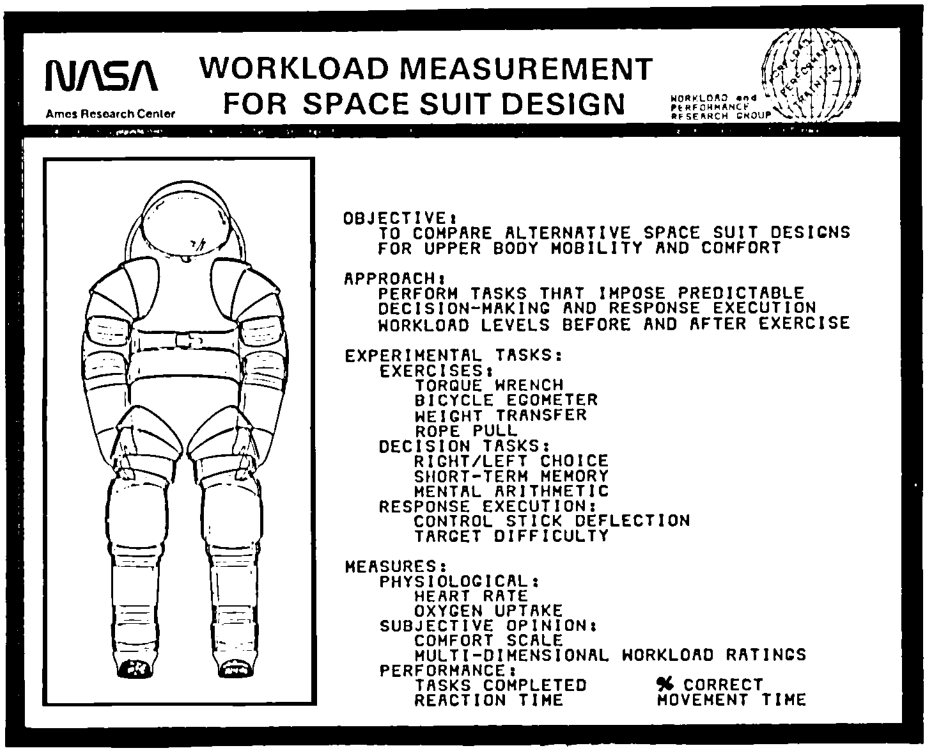
(9) PRIMARY TASK FOR CURSOR CONTROL EVALUATIN IN SHUTTLE

The "Fittsberg task" is a simple, flexible laboratory task where subtask workload levels can be independently manipulated and measured over a wide range. It provides an alternative to the traditional dual task paradigm in which two unrelated tasks are performed during the same time interval. It represents the types of tasks that are performed in many automated systems: a requirement for action is recognized and the appropriate plan of action selected. The plan of action is executed by an automated system in response to a discrete command.



Fittsberg task components are functionally related - response selection provides information for and initiates response execution. The response selection task is a target acquisition based on Fitts' Law. Two identical targets are displayed equidistant from a centered probe. The decision about which target to acquire is based on a Sternberg memory search task; Subjects acquire the target on the right if the information presented in the center of the display is the same as a remembered value or the target on the left if it is not. A wide variety of response selection tasks have been used in addition to the Sternberg Task - - mental arithmenic, pattern match, time estimation, etc. Workload levels of one or both task components can be held constant or systematically varied within a block of trials. The stimulus modality of the two components can be the same (visual/visual) or different (auditory/visual).

Response selection performance is measured by RT and percent correct. Response execution performance is measured by MT. RT, but not MT, increases as the difficulty of the response selection task is increased. MT, but not RT, increases as target acquisition difficulty is increased. Workload ratings for the Fittsberg task integrate the influences of the component subtask components. Workload ratings and performance levels for the combined task are often substantially less that would be predicted by simply adding single-task workload ratings or response times .

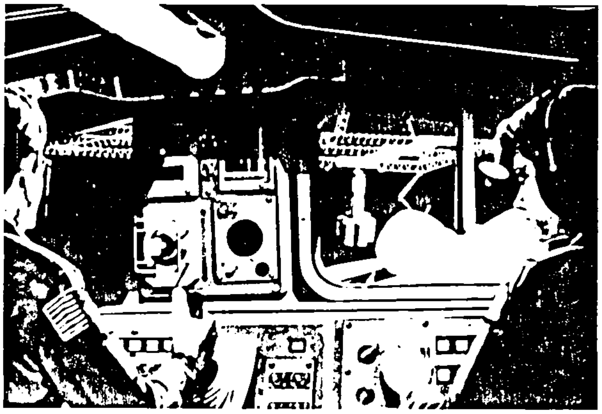


This task has proven to be a useful focal point for several space-related applications. In response to a request by Johnson Space Center, we provided the hardware and software to use the Fittsberg task in a series of experiments in which two alternative space suit configurations were compared with respect to upper body mobility and comfort. Several Fittsberg tasks are performed using either fine or gross arm movements before and after a battery of physical exercises are completed. Physiological, subjective and performance measures are obtained to aide in the comparison between the two suit configurations.

Again the advantage of using this task is the fact that it has been calibrated in advance of the experiment with respect to expected workload and performance levels.

NASA-AMES WORKLOAD AND PERFORMANCE RESEARCH

STUDY OF CURSOR CONTROL DEVICES IN ZERO-G



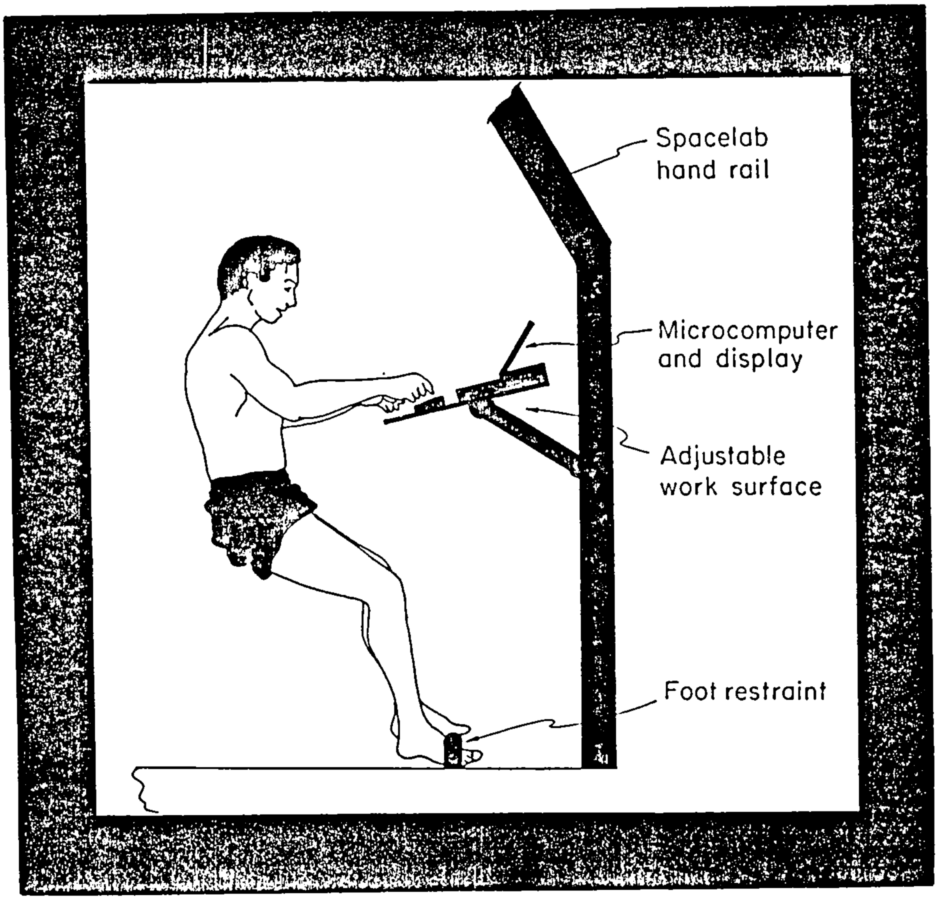
OBJECTIVE

- EVALUATE 3 CURSOR CONTROL DEVICES EARLY AND LATE IN ZERO-G EXPOSURE DURING FY86 SHUTTLE MISSION

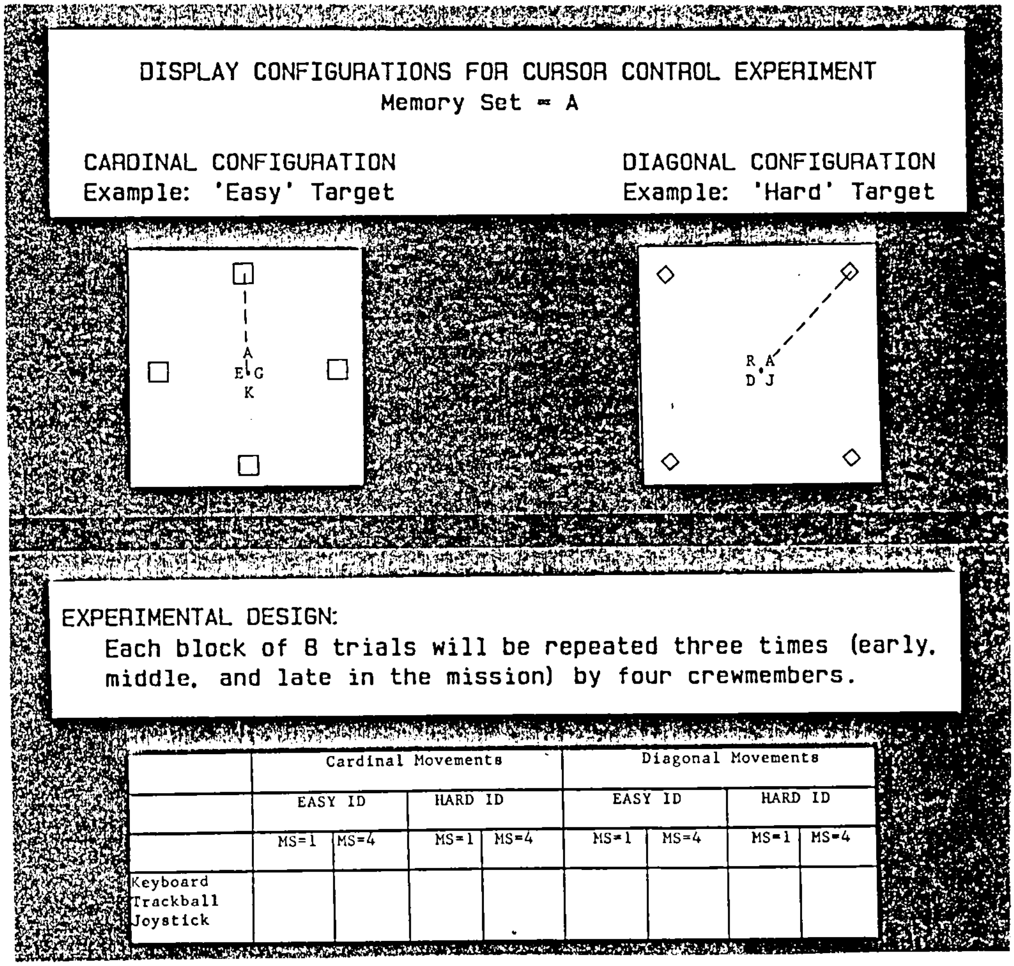
APPROACH

- ARC/UNIVERSITY COLLABORATION
- ARC-DEVELOPED "FITTSBERG" TASK AS CRITERION TASK
- COMPARISON OF VERTICAL, HORIZONTAL, AND ANGULAR MOVEMENTS TO ACQUIRE TARGETS WITH:
  - TRACK BALL
  - JOYSTICK
  - ARROW KEYS

The Fittsberg task was selected for an experiment that will be flown in the Shuttle in the fall of 1986. The purpose of the experiment, which will be conducted jointly with MIT and JSC, is to evaluate three alternative cursor control devices in zero-G.

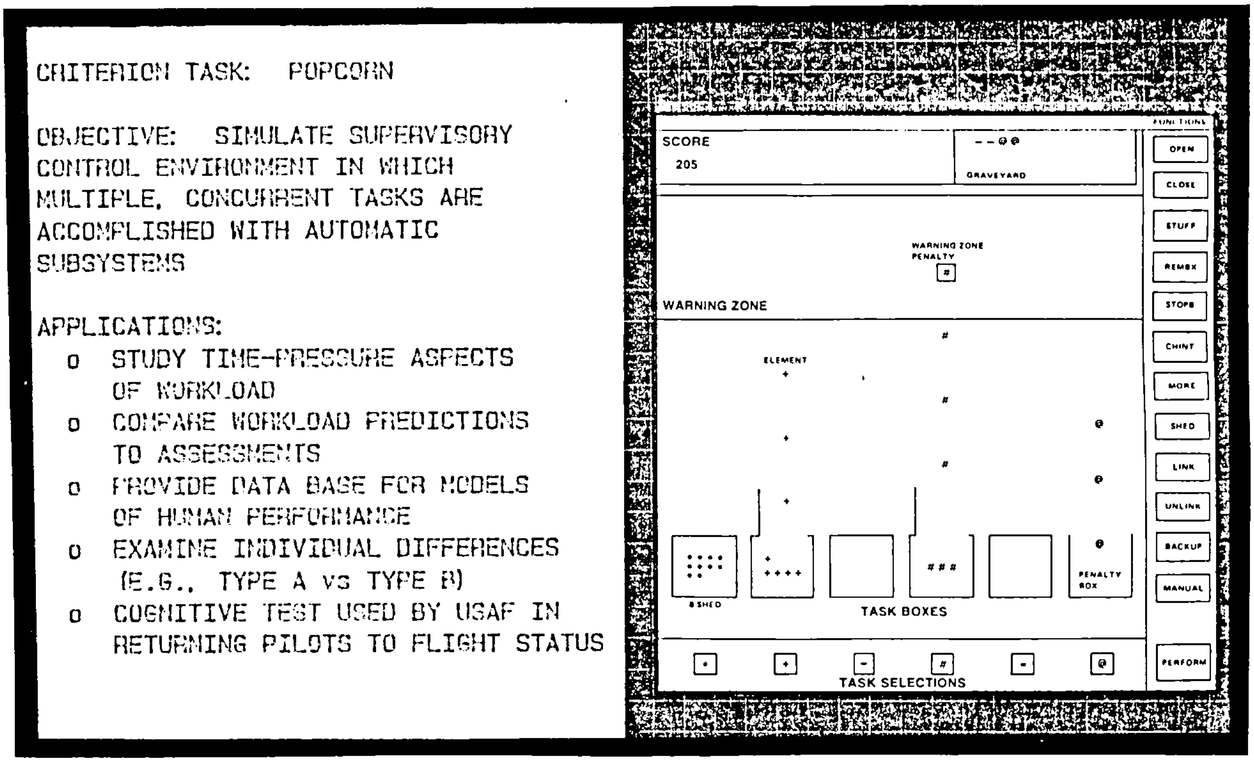


The experimental task will be presented on a Compass-Grid microprocessor mounted on an adjustable work surface attached to a Spacelab hand rail. Both foot and arm restraints will be provided. The three space-rated input devices devices - - track ball, arrow keys, and joystick will be positioned with Velcro strips.



	Cardinal Movements				Diagonal Movements			
	EASY ID		HARD ID		EASY ID		HARD ID	
	MS=1	MS=4	MS=1	MS=4	MS=1	MS=4	MS=1	MS=4
Keyboard Trackball Joystick								

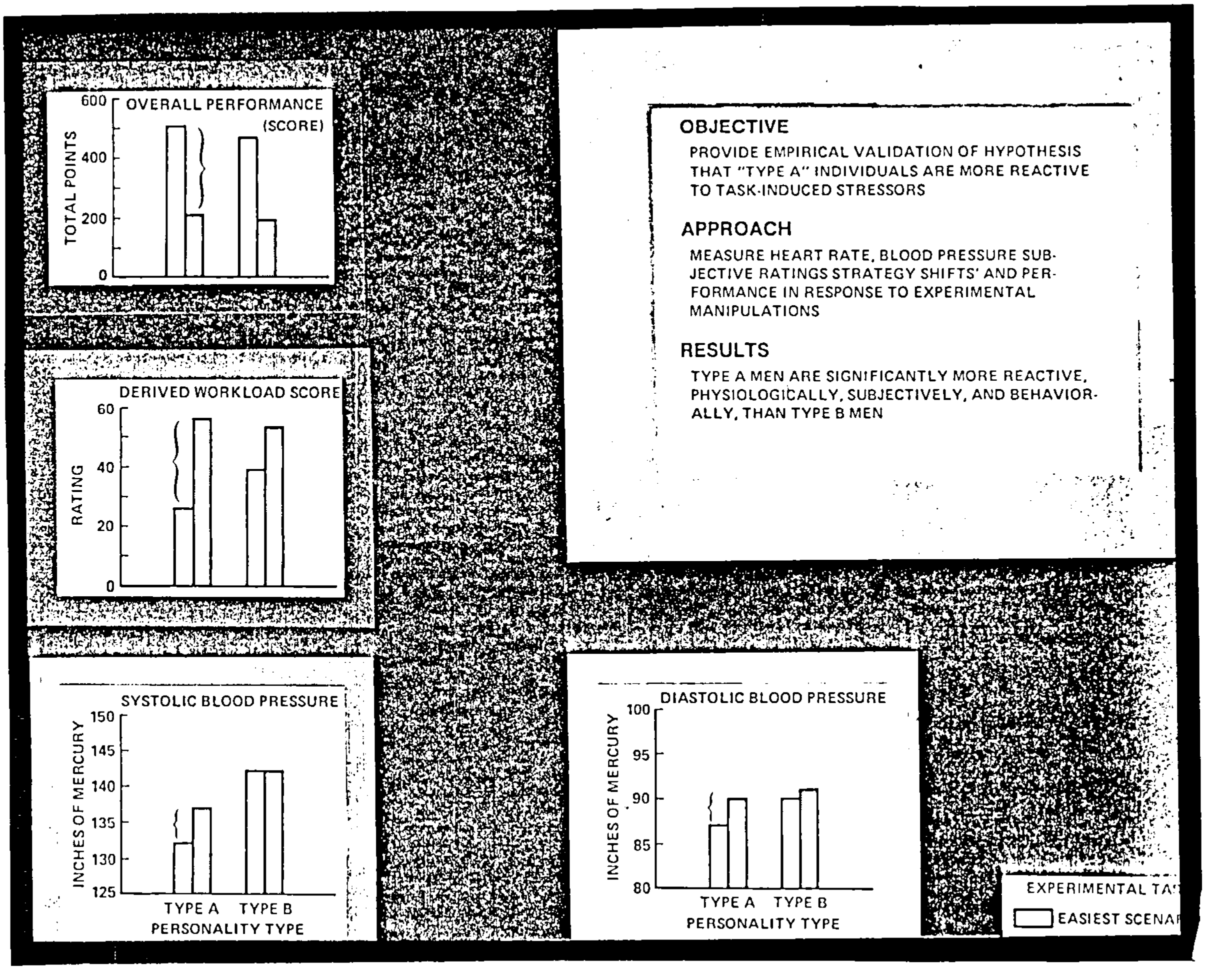
Twenty-four blocks of Fitsberg trials will be performed during three, 30-min intervals early, middle, and late in the 7-day mission by four mission specialists. The difficulty of the response selection task will be manipulated by varying the number of items to be remembered (the Sternberg paradigm). The difficulty of the response execution portion of the task will be varied by manipulating the direction of movement - - either in a cardinal direction (up/down/right/left) or at an angle - - and by varying the index of difficulty of the target (target size and distance).



A second example of a criterion task developed at Ames is POPCORN, a dynamic, multi-task, supervisory control simulation. It represents operational environments in which decision-makers are responsible for actuating semi-automatic systems according to both pre-programmed and flexible schedules. Its name, POPCORN, reflects the appearance of groups of task elements waiting to be performed (they move around in a confined area and "pop" out when selected for performance).

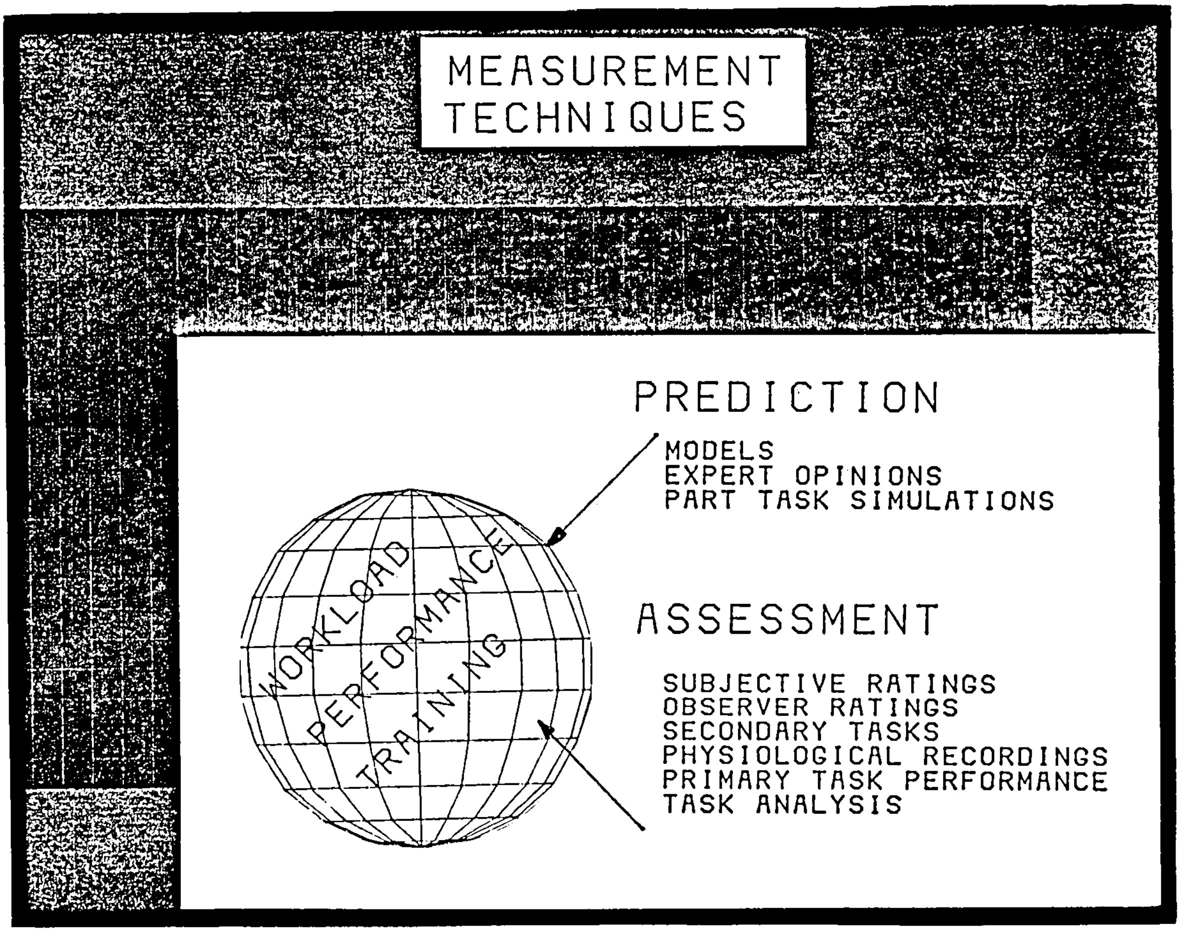
Operators decide which tasks to do and which procedures to follow based on their assessment of the current and projected situation, the urgency of specific tasks, and the reward or penalty for procrastination or failure to complete them. Simulated control functions provide alternative solutions to different circumstances. Control may be accomplished by magnetic pen and pad entry, mouse input, or a VOTAN voice recognition system.

The most compelling feature of the POPCORN task is the wide variety of time pressure sources that can be generated, the time management strategies that are available, and the penalties imposed for procrastination.



A recent experiment conducted jointly with SRI is one example of the applications in which POPCORN has been used. The objective was to provide empirical validation of the hypothesis that "Type A" individuals are more physiologically, behaviorally, and psychologically reactive to task-induced stressors than "Type B" individuals. It has been suggested that it is this differential level of reactivity that leads to the eventual development of cardiovascular disease associated with the "Type A" personality.

We found very strong empirical evidence that "Type A" men with normal resting blood pressure levels, are significantly more reactive to different levels of task-induced stress than otherwise similar "Type B" males. The results of this study have prompted researchers at Brooks AFB to adopt POPCORN as one of the battery of tests to be given when returning grounded pilots to flight status.



For the remainder of this talk I will describe typical predictive models and measures of workload that have been developed by this program and the methods used in validation.

ASSUMPTION1: FOR WELL LEARNED TASKS, FUNCTIONALLY INTEGRAL ACTIVITIES PROVIDE THE NOMINAL LEVEL

- NOMINAL DURATION
- NOMINAL PERFORMANCE LEVELS
- WORKOAD EXPERIENCED

ASSUMPTION 2: ADDITIONAL TASKS, CHANGES IN THE ENVIRONMENT, EQUIPMENT, OR PROCEDURES

- IMPAIR WHOLE OR SUBTASK PERFORMANCE
- REQUIRE ADDITIONAL TIME
- INCREASE WORKLOAD

ASSUMPTION 3: THE INFLUENCE OF LIKELY OCCURRENCES DURING DIFFERENT NOMINAL ACTIVITIES CAN BE COMPUTED AND USED TO PREDICT NEW LOAD LEVELS

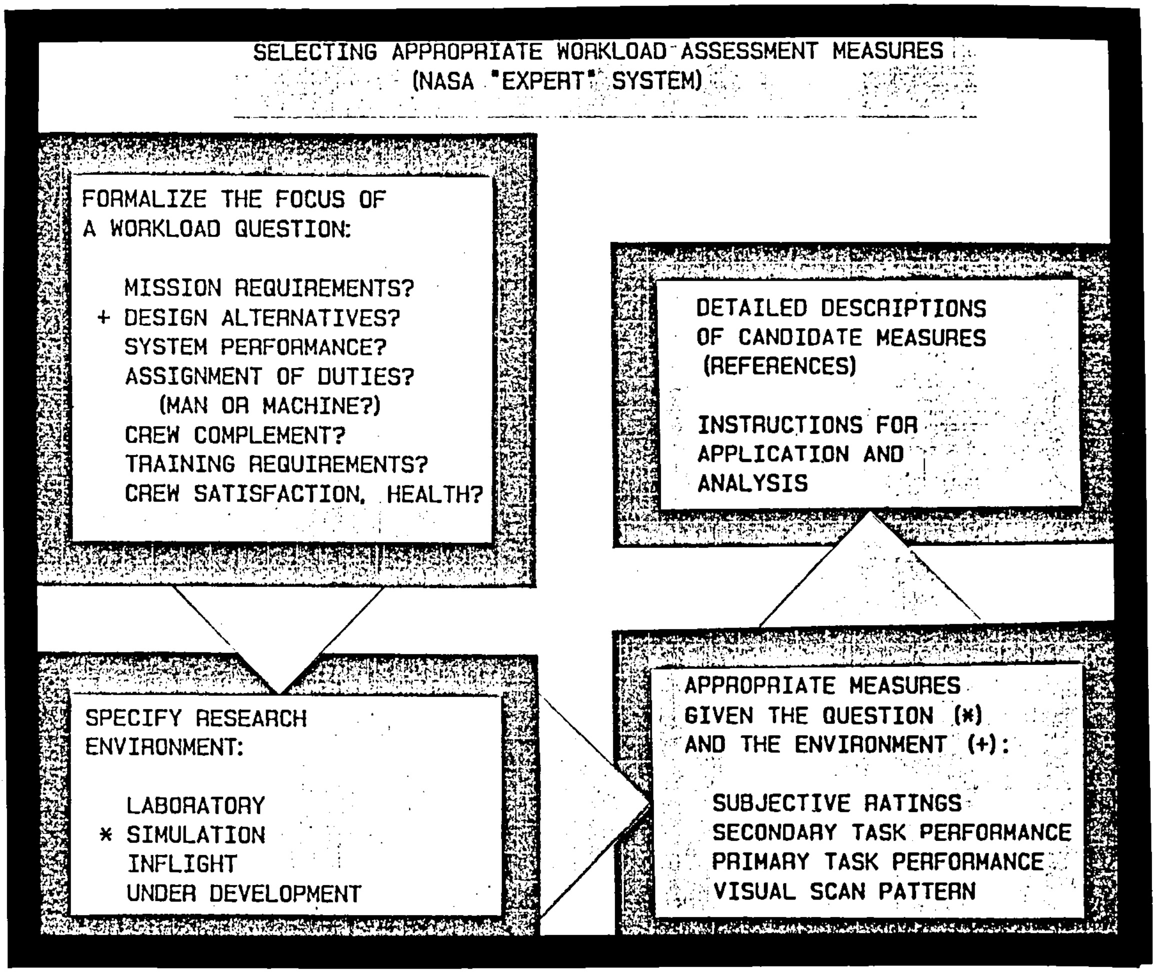
ASSUMPTION 4: THE RULES FOR COMBINING "EVENTS" WITH NOMINAL ACTIVITIES TO CREATE DIFFERENT TASKS MAY REFLECT:

- TASK INTEGRATION
- ADDITION
- COMPETITION

During the past three years, we have developed a predictive model of pilot workload. The goal was to provide a standardize method of creating simulation scenarios to use in research. The initial focus of the model was on general aviation instrument flight (for convenience), although the model philosophy is being extended to helicopter operations and the space station. The goal was to provide a standardized format for creating simulations scenarios for workload and performance validation research, flight handling quality research, display and control evaluations and so on.

Workload prediction must, by necessity, focus on imposed task demands as a starting point. We assume, that for well-learned tasks, functionally integrated activities that are normally performed as a unit should provide the basic ingredients of the model. Rather than performing a fine-grained analysis of the components of highly overlearned tasks (which tends to overestimate the workload of experienced operators), we chose to focus on a level of analysis that most closely represents that used by expert performers when describing, performing and evaluating their actions.

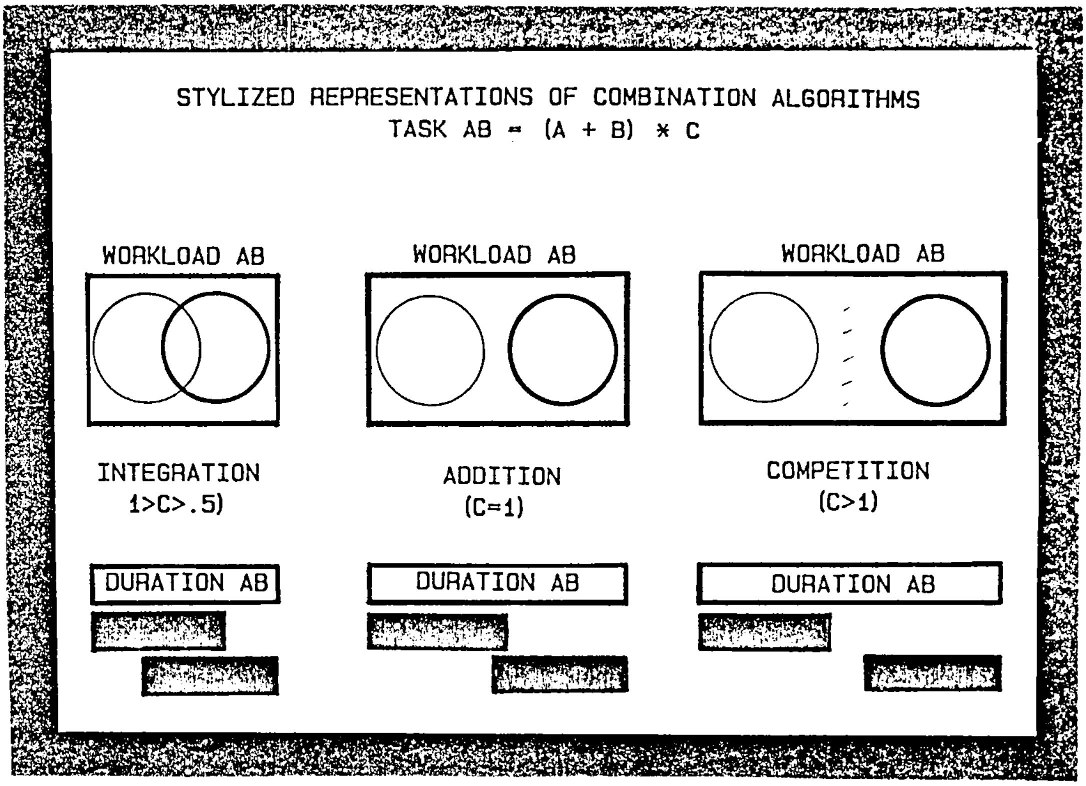
The workload of these functional units - - such as specific phases of flight, sequences of control activities, etc - - is quantified and serves as the starting point for the model. Additional tasks, changes in the environment, equipment, procedures, or time available can be superimposed on these basic elements to modify the workload of the target scenario. The influence of these events can be computed as well, and the rules by which they combine with different nominal segments determined analytically, empirically and through expert opinions.



We are in the process of developing a simple "Expert" system for the selection and application of workload measures on an IBM-PC. The goal is to provide an interactive system whereby an individual who is not familiar with workload assessment, but needs to obtain information about the workload of a particular task or alternative pieces of equipment, can select and apply an appropriate measure. This system will serve to summarize and allow practical application of the results of our research.

This system will assist the user in formulating the question to be addressed and to specify the research environment. Appropriate measures will be suggested and evaluated. Detailed descriptions about how to apply the measure will be provided along with examples and references. The system will be a stand-alone, user-friendly, and provide easily accessible information. The first application will be as a hands-on component of the Army MANPRINT course.

As long as the human remains an integral element of complex, advanced systems, the need for standardized measures and predictors or human workload and performance will be required. The need for such tools is obvious both during the design and construction of the space station. Although the environment and activities to be accomplished in the space station are unique, the fundamental principles of human behavior and experience remain the same, and we are confident that the concepts and techniques that we have developed will provide a useful and informative tool for the development and operation of the space station.

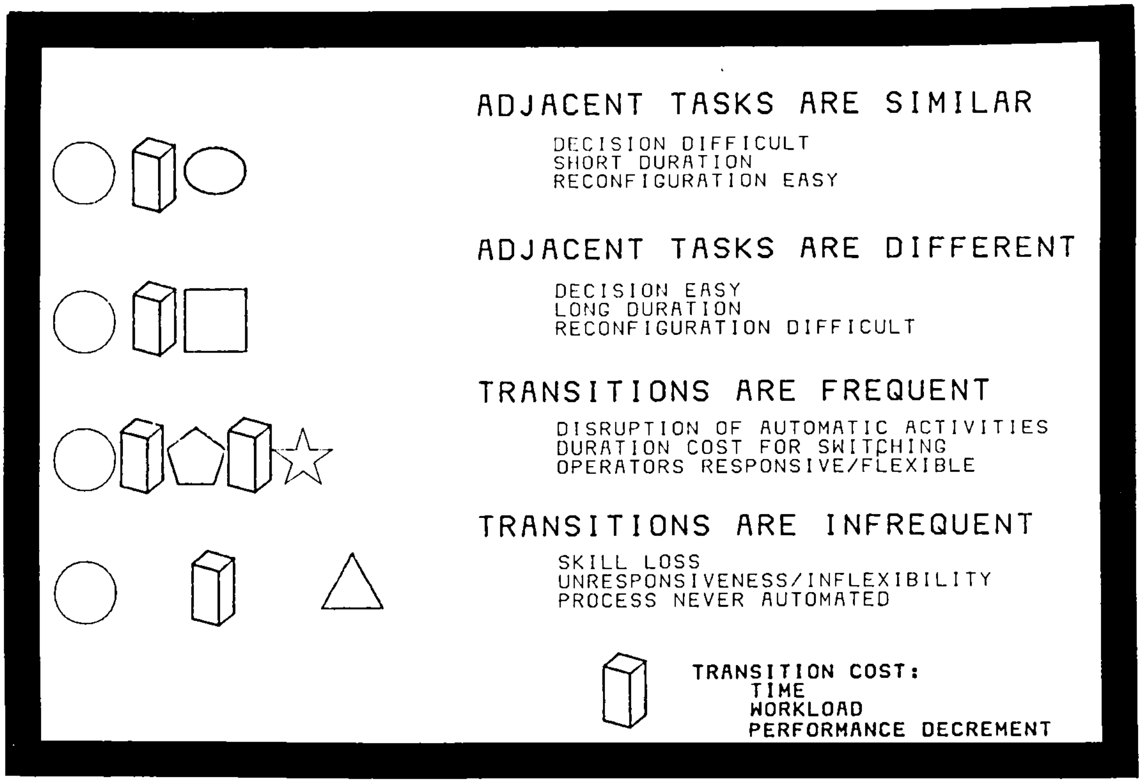


Through extensive research, we have identified a continuum of task combination rules that range from:

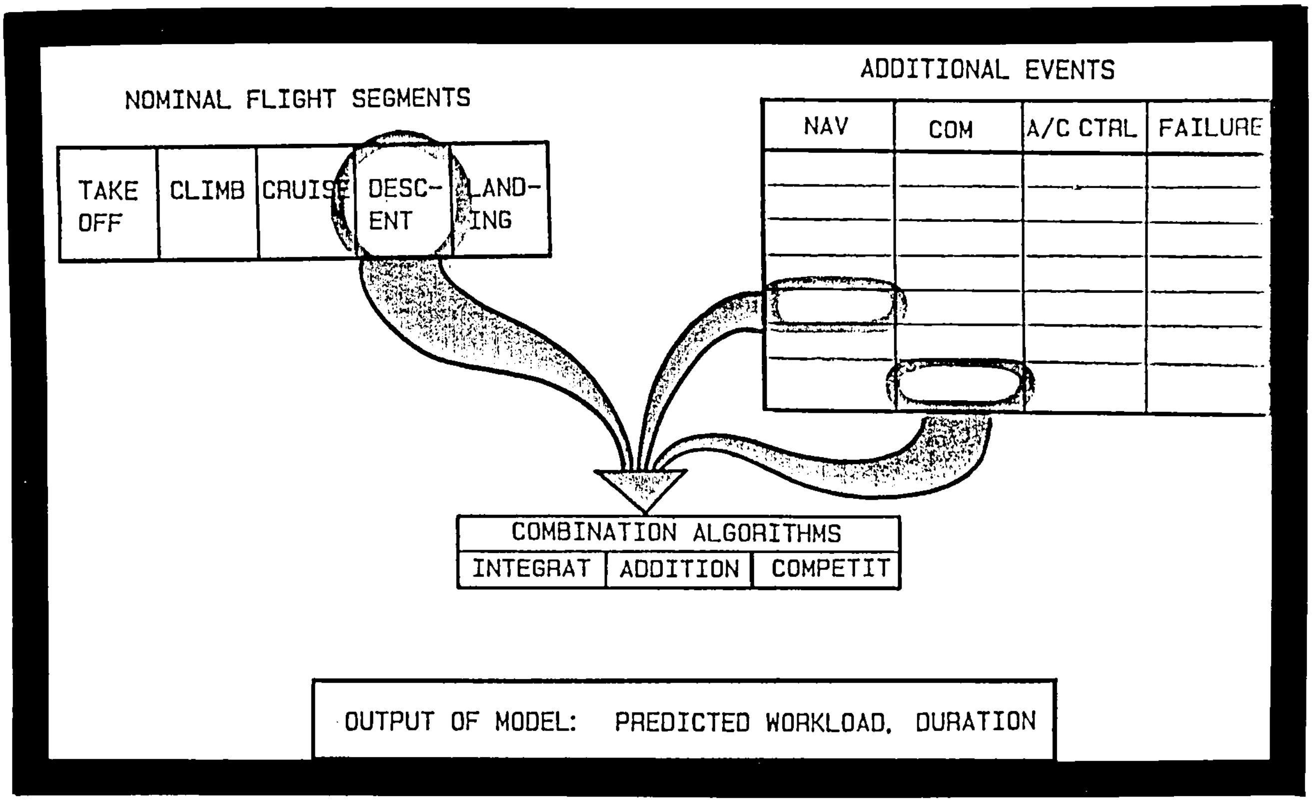
(1) INTEGRATION: The workload or time required to perform concurrent tasks approximates that of the more demanding of the components

(2) ADDITION: The workload or time required for a complex task is equal to the sum of the components

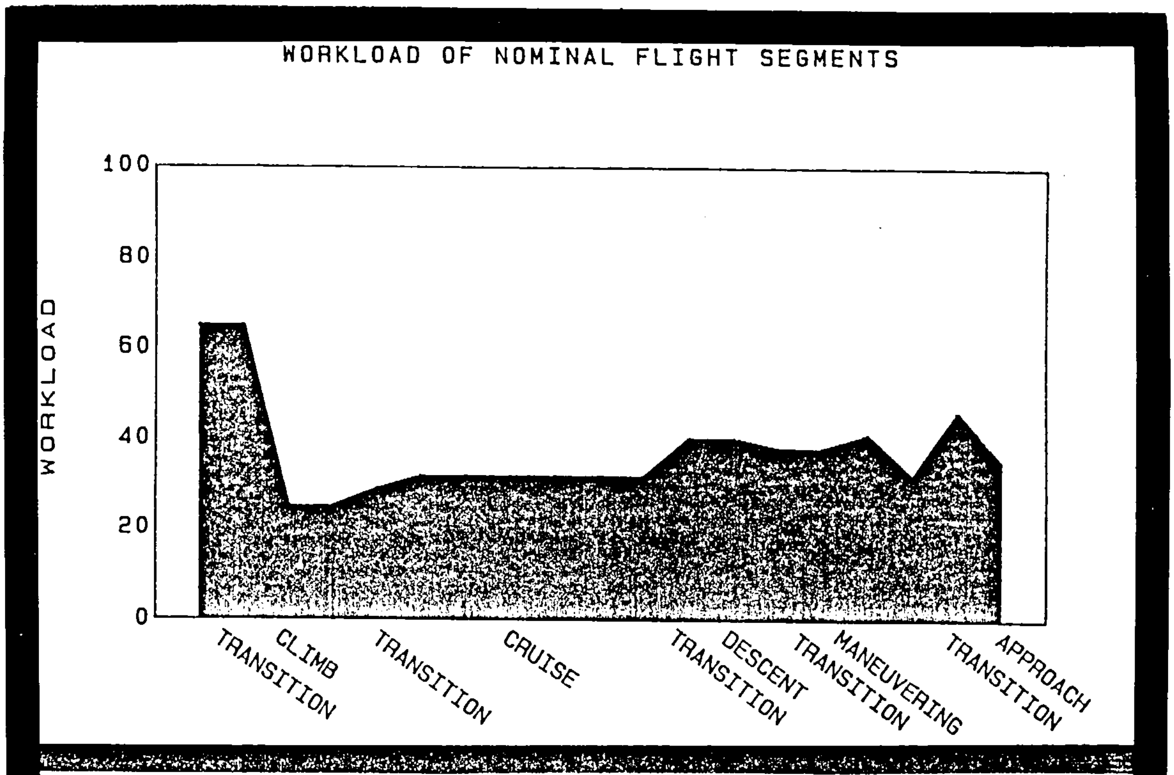
(3) COMPETITION: Task components compete for operator's attention and "resources" and cannot be performed within the same time interval There is an additional cost for switching among them and the cost of performing both tasks is greater than the sum of the parts.

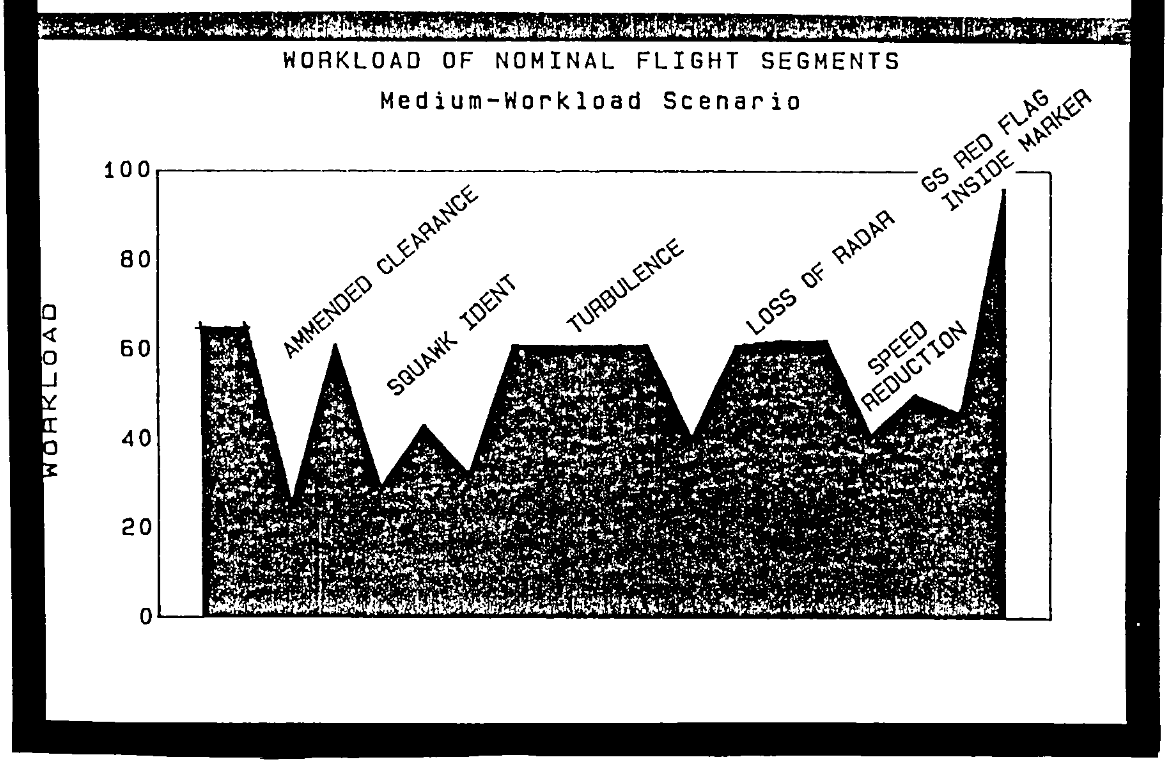


In addition to the basic workload associated with task segments and additional events, there may be brief periods of relatively high workload associated with the transition from one task segment to another. If the successive tasks are similar or frequently occur together, the transitions may occur quickly and with low workload. If they are not, the transitions may be time-consuming and demanding. In addition the sheer number of transitions that occur during a duty period may lead to high workload levels

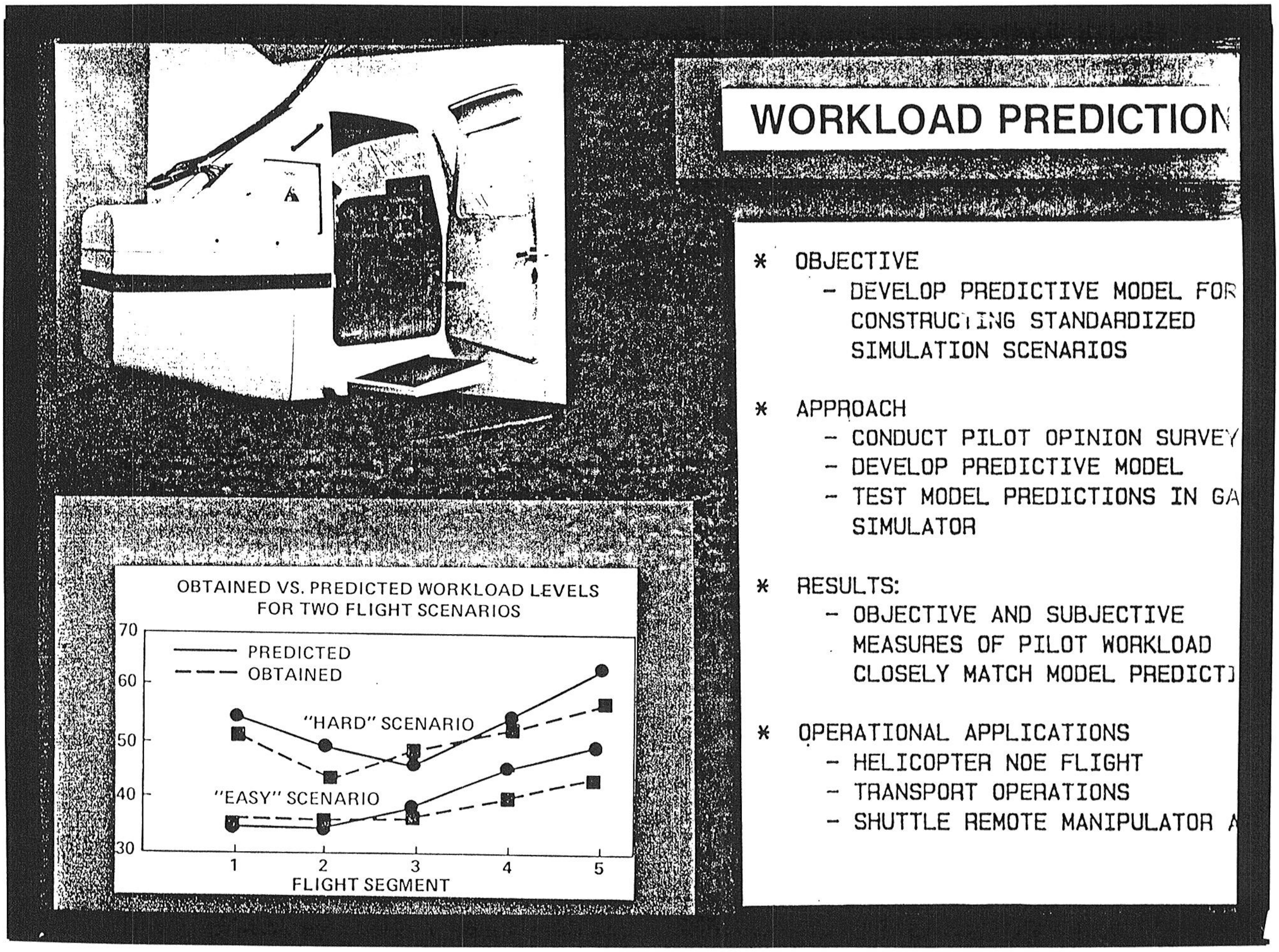


For each of the operational tasks to which this model is extended, a vehicle-specific data base is required, although the philosophy and structure of the model may be transferred. These nominal elements and additional events are entered into the computer data base and combined according to the appropriate algorithms dynamically by a researcher who wishes to create a simulation scenario of a specific duration, type, and workload level. The user may add and delete tasks until the predicted workload profile approximates the desired levels of imposed workload. The output of the model is a graphic representation of the predicted workload levels across time and a printed script to follow in conducting the simulation or operational test.



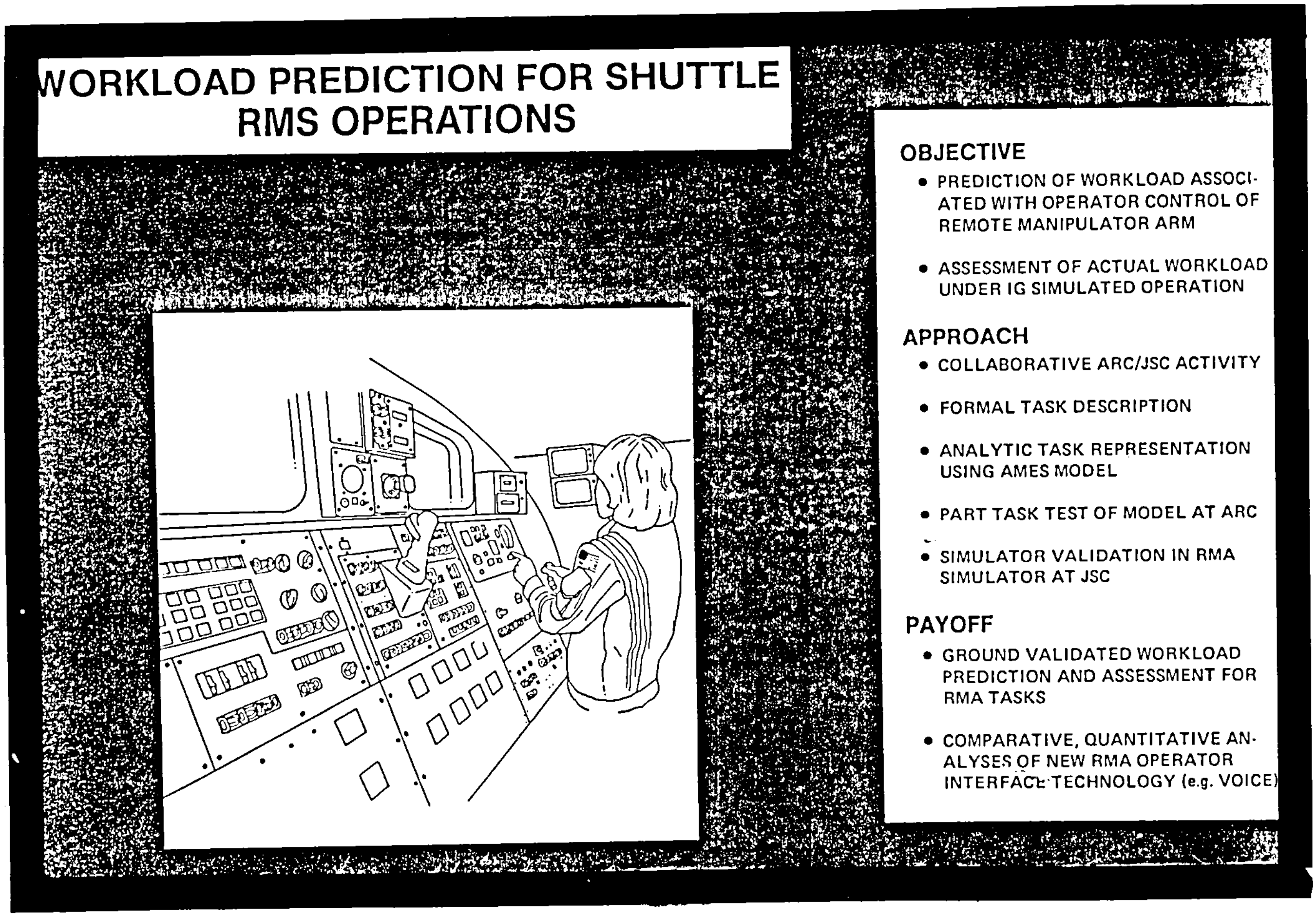


The following graphs represent one such nominal and modified scenario developed for instrument flight for a general aviaiton aircraft.



The predictions of the model have been validated in a series of simulation experiments. A battery of converging workload assessment measures are imposed to test the predictions of the model.

The first operational application of the model will be for advanced helicopter missions. Subsequent applications will focus on the space station as part of a Focused Technology Work Integration effort we will perform jointly with JSC.

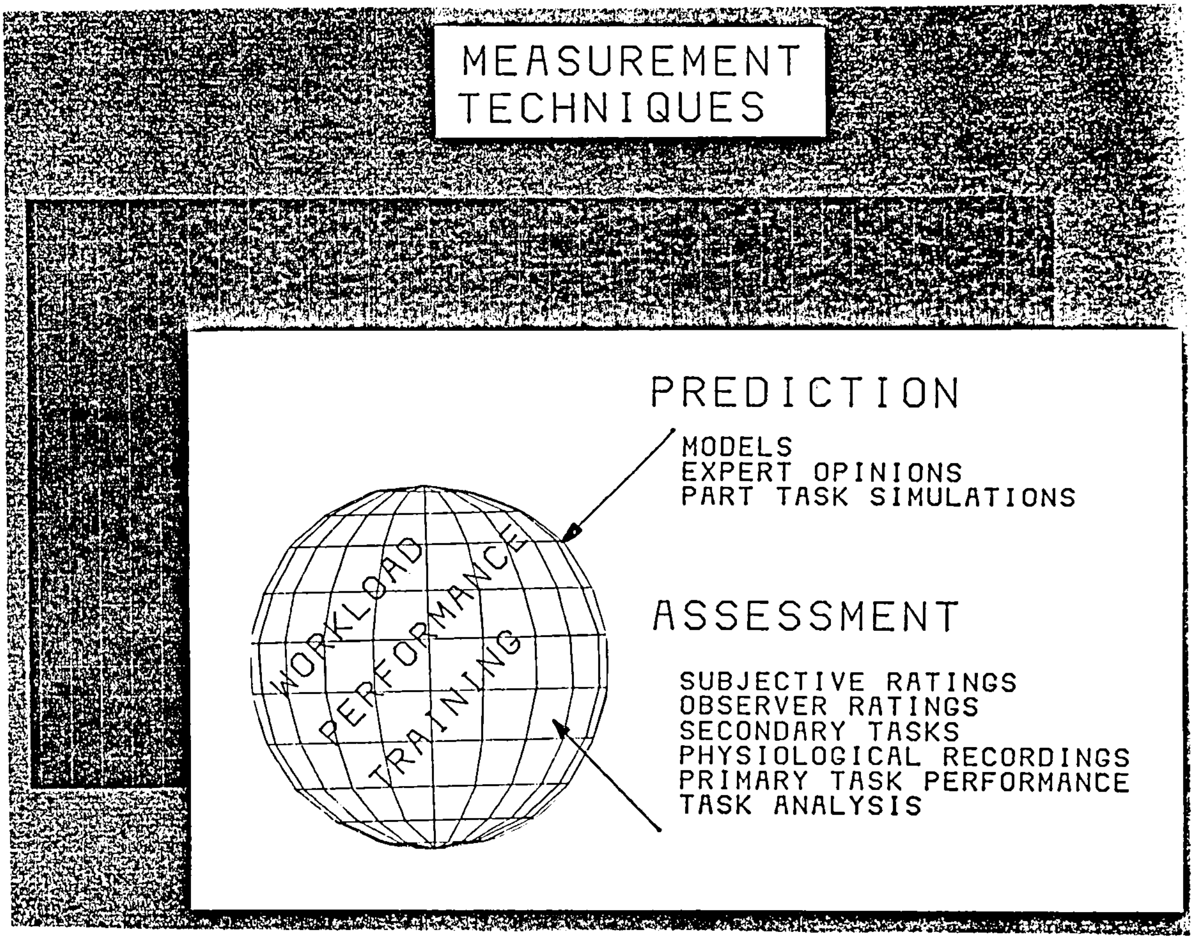


The objective of this task is to develop and test a workload model for evaluation and prediction of a Space Station human operated system. The system selected as the first test of the model is the Remote Manipulator Arm. The initial focus will be on the existing RMS used in the shuttle, although space-station specific modifications will be incorporated as they are specified.

A functional task analysis will be provided by JSC. It will be used as the initial data base for the prediction model. Using analytic, part-task simulation, and expert opinion approaches, the appropriate workload levels and combination rules will be determined.

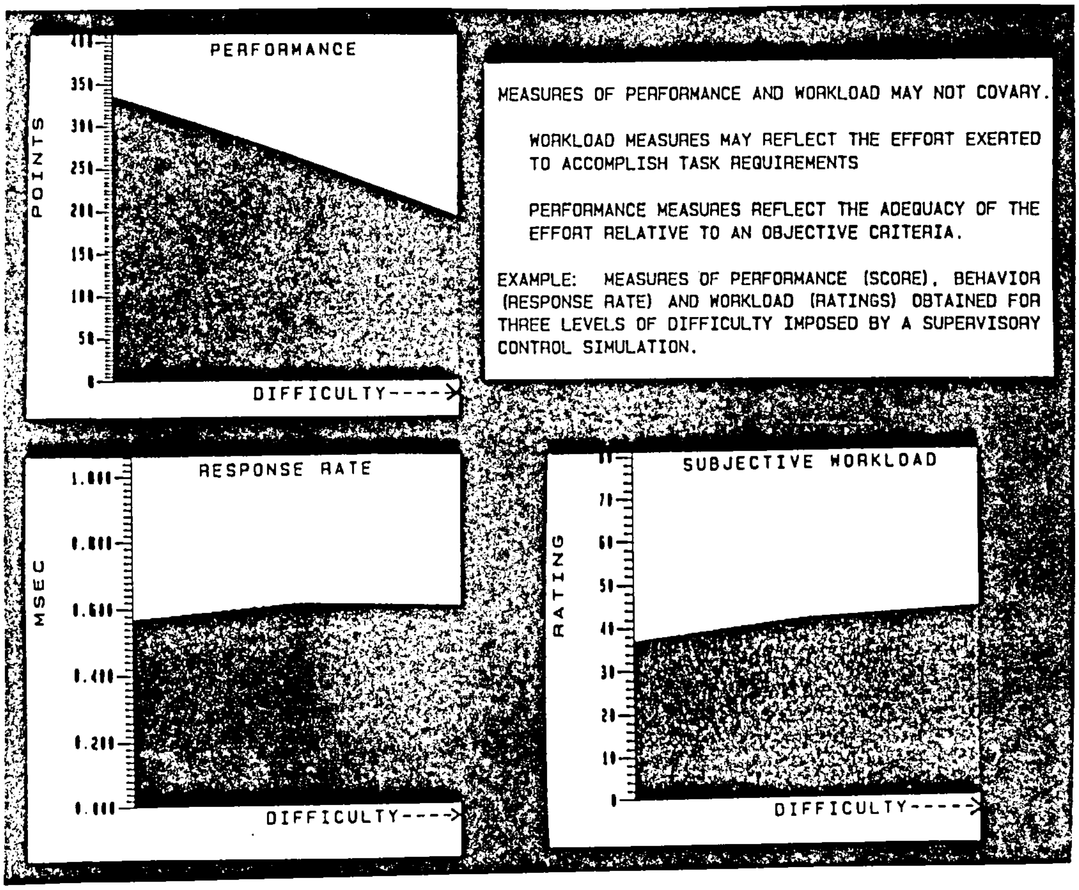
An initial test of the model will be performed at Ames, in the proximity opearations mockup. A simulator evaluation will be performed at Johnson Space Center in the RMS simulator during the second year of the project. This model will be used to predict the workload of alternative configurations and advanced RMS technology from the perspective of the human operator. Future applications might be to provide workload estimates as a feature in the existing OPSIM model developed at Ames.

The expected product of this effort is a ground-validated workload and performance model that is suitable for use by contractors and Levels B and C personnel for the prediction and evaluation of workload and performance-effectiveness of human-operated Space Station systems.



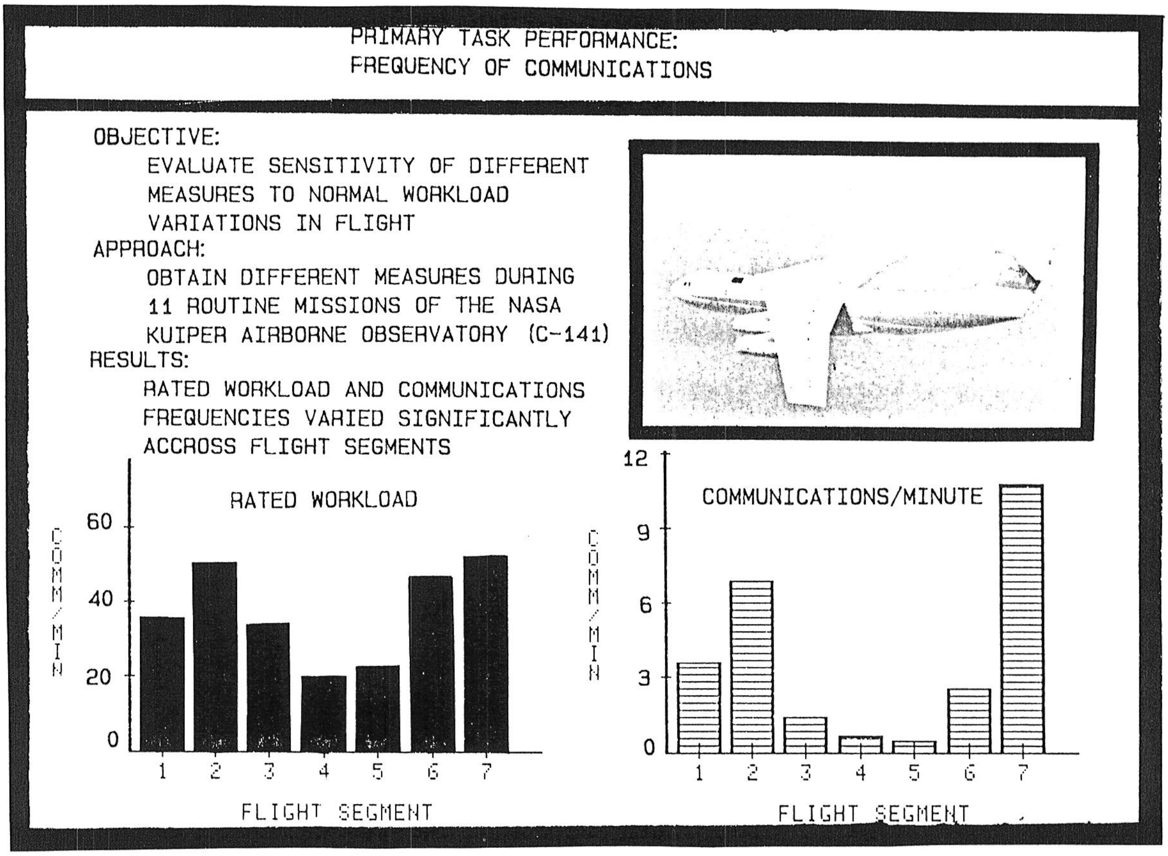
The primary focus of this program has been the development and validation of of a battery of workload and performance assessment tools that reflect sound theoretical models of human operator performance and information processing. We examined existing techniques and developed additional ones to meet the needs of a wide variety of operational environments. Our goal was to provide sensitive and reliable tools and to disseminate information about them to make the results of our research widely available and practically useful.

For each of three categories of measures - - performance, physiological, and subjective - - I will describe a typical technique and describe how it was developed and validated.



Early in the program, it became clear that, although human and system performance provided the most common motivation for workload analyses, performance measures themselves do not always reflect variations in operator workload. Within the range of their capabilities, skilled, motivated operators exert increasing levels of effort to accomplish increasing task demands. Performance degredation often occurs only after their capabilities are exceeded, or when they choose to maintain a consistent level of effort in the face of increased task demands. Subjective secondary, and physiological indicators of workload are more reflective of the cost of performance to the operator in such cases, and are able to quantify how much reserve capacity an operator still has when performing the task of interest. In addition, workload measures are able to predict future performance - - should task demands be increased yet farther - - while measures of performance are not.

One example of a dissociation between measures of workload and performance is represented by a recent study completed with the POPCORN simulation. As time pressure was increased, performance (as measured by the subject's score) dropped, as predicted. Workload levels remained constant however. They reflected the fact that operators maintained a consistent response rate in the face of increased tasks demands, and thus the cost of task performance - - at least as far as the operators were concerned - - remained constant.



Selected measures of performance may covary with operator workload. In a study that we conducted in the Kuiper Airborne Observatory, we found that the rate of communications activities provided a convenient and sensitive measure of the overall levels of workload imposed on the flight crewmembers.

In addition, we have found that specific types of communications are associated with different levels of workload. A post hoc communiations analysis can provide a sensitive workload evaluation in a many of environments, using data that is readily available in most operational environments.

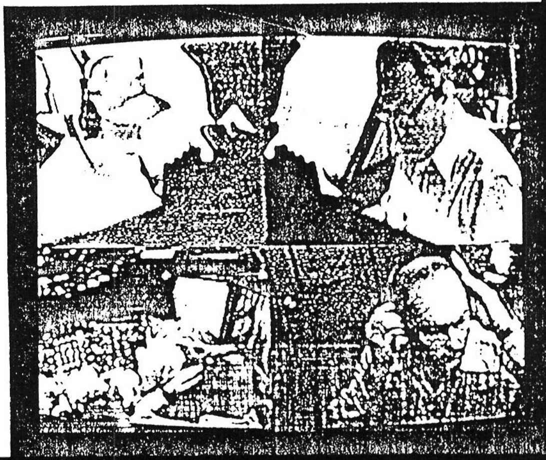
COMMUNICATIONS ANALYSIS: MEASURES OF CREW COORDINATION AND DECISION MAKING

OBJECTIVE:

ANALYZE FLIGHT DECK AND ATC COMMUNICATIONS TO ASSESS AIRCREW DYNAMICS COMMUNICATIONS COMPETENCY, AND AIRCRAFT MANAGEMENT

APPROACH:

- o CONDUCT SIMULATIONS IN B-707 SIMULATOR
- o OBTAIN POST-FLIGHT EVALUATIONS BY:
  - (1) CREWMEMBERS
  - (2) EXPERTS IN LINGUISTIC AND SEMANTIC ANALYSIS
  - (3) EXPERTS IN FLIGHT SAFETY



RESULTS:

- o CREWS DIFFERED IN COMMUNICATIONS COMPETENCY AND LEADERSHIP ROLES
- o CREW COORDINATION AFFECTED DECISION MAKING AND AIRCRAFT MANAGEMENT

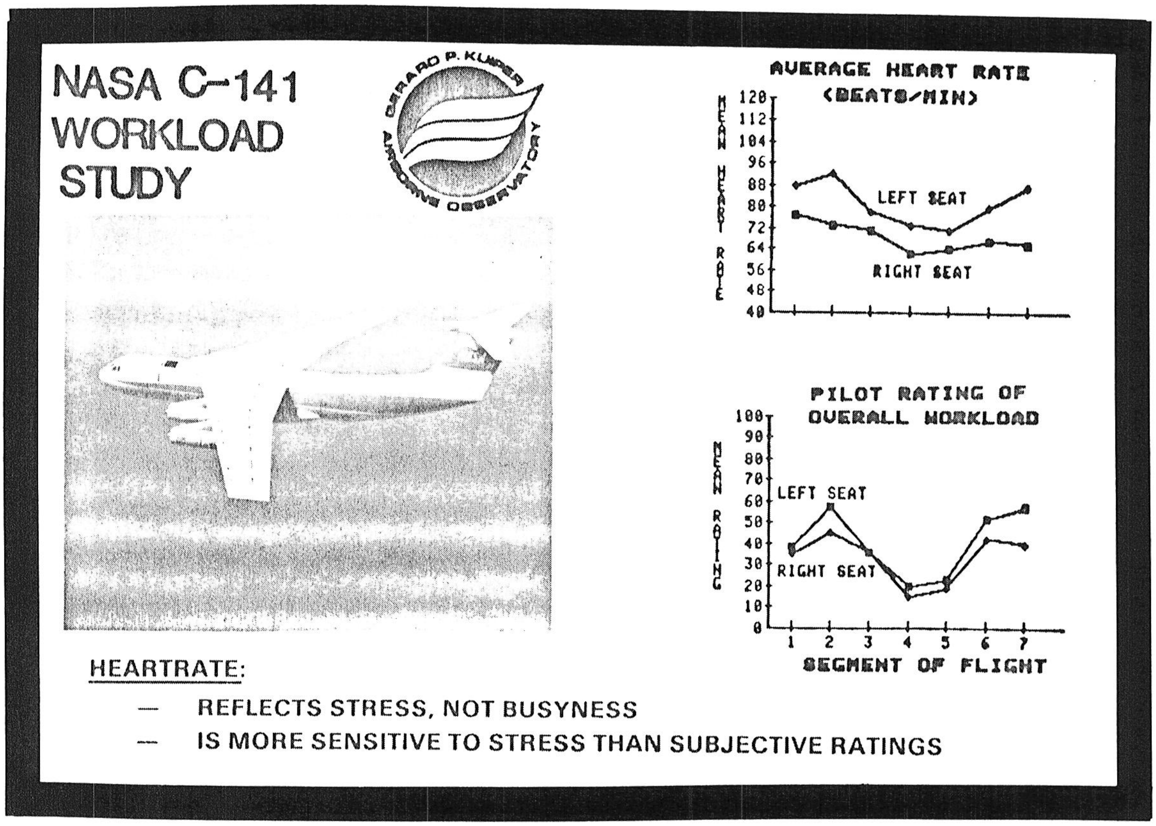
Another facet of communications that we have investigated is the role of flight deck communications in aircrew organization and coordination. In a recent simulation of transport operations, we found that crews differed in communications competency. Communications analyses provided a sensitive measure of leadership and crew coordination - - factors that play important roles in the safety and efficiency of aircrew performance. Crew coordination affected decision making behavior and aircraft management.

The primary goal of this part of the program is to develop a training program to improve crew communications competency, corrdination and leadership.

PHYSIOLOGICAL MEASURES: EXAMPLES

- o MEASURES OF MENTAL AND PERCEPTUAL PROCESSING
  - * EVOKED CORTICAL POTENTIALS
  - * EYE POINT OF REGARD
- o MEASURES OF EMOTIONAL AND PHYSICAL ACTIVATION
  - * HEART RATE AND VARIABILITY
  - * BLOOD PRESSURE
  - * GALVANIC SKIN RESPONSE
  - * RESPIRATION RATE
  - * MUSCLE TENSION
  - * VOCAL STRESS
  - * PUPIL SIZE

We have investigated a number of physiological measures of workload. Several measures provide relatively specific indicators of mental and perceptual processing - - such as auditory evoked cortical potentials. In addition, we have examined a number of measures that reflect more general levels of activation, such as heart rate, and pupil size. The advantage of physiological measures is that they are unobtrusive, do not interfer with primary task performance, and they provide common, objective measures across a variety of tasks.



The research we have conducted in evaluating heart rate and heart rate variablity is one example of this area of research. Heart rate provides a convenient and nonintrusive indicator of the overall level of activation of an operator. It is less likely to reflect more subtle changes in workload associated with different levels of mental activities, however. In the study that I mentioned earlier, we obtained measures of pilot heart rate during 11, eight-hour routine missions of the Kuiper Airborne Observatory using the portable Vitalog physiological recording unit.

The heart rate profiles of the pilot-flying, reflected the expected peaks during take-off and landing. The profiles of the pilots-not-flying reflected no significant changes, however. These results, in agreement with earlier studies, suggest that heart rate reflects responsibility and stress, rather than mental workload.

These data are particularly interesting because the test pilots who participated in the study were qualified in both positions, and the same pilots are represented in the data for both. The pilots experienced and reported apparently similar levels of subjective workload throughout the flight, but the heart rates suggested that there were differences in the physiological consequences of performing the duties required by the two positions.

In other studies, we have found that heart rate is quite insensitive to the variations in levels of workload imposed by a wide variety of laboratory tasks unless rather heavy physical effort is involved.

These data again point out the need for multiple, converging measures of workload to obtain the most complete picture possible of the impact of performing a task on the operator.

We are focusing most of our research efforts in the area of heart rate variability. In particular, we have evaluated the power in the .1 Hz range of the frequency spectrum of the beat-to-beat intervals as a very promising measure. There is considerable evidence that this measures provides a sensitive indicator of different levels of mental workload. The typical finding is that heart rate variability (and the power in the .1 Hz region) decrease as mental workload is increased. A "black box" has been developed to obtain and process this measure automatically online.

SUBJECTIVE RATINGS: ISSUES

* PROVIDE SIGNIFICANT SOURCE OF INFORMATION
* MAY TAP THE ESSENCE OF MENTAL WORKLOAD
* REFLECT SUBSET OF INFORMATION AVAILABLE DURING TASK PERFORMANCE
  - RESULTS OF INFORMATION PROCESSING
  - MEMORIES
  - OVERT BEHAVIOR
  - FEELINGS
* INDIVIDUAL DIFFERENCES IN DEFINITION AND AND EXPERIENCE
* NO MENTAL REFERENCE SCALE FOR "WORKLOAD"
* BEST TO COMPARE SHARED QUALITIES AND SIMILAR ACTIVITIES
* CALIBRATION OF RATERS
* TIMING
  - ON-LINE vs RETROSPECTIVE
  - PRIMACY/RECENCY OR ODDBALL EFFECTS
* PSYCHOMETRIC CONSIDERATIONS
  - EQUALITY OF INTERVALS
  - NO "ZERO" POINT OR "MAXIMUM"

Considerable effort has been devoted to understanding and measuring the subjective workload experiences of operators, as this is the most convenient and practically useful measure. In addition, it is the measure against which most other measures are calibrated. We have found that subjective ratings provide a significant source of information, come closest to tapping the essence of mental workload, and provide the most direct indicator about the subjective impact of a task on operators.

People often generate evaluations about the workload of ongoing experience, however they rarely quantify or remember such experiences. Thus, experiencing workload is unique to experimental situations, although the requirement to verbalize, remember or quantify such experiences may not be a commonplace activity. The goal of our research has been to determine what factors influence such subjective experiences (and which ones do not) and to develop a valid, sensitive, and reliable measure of them.

THE TYPES OF EXPERIMENTAL TASKS INCLUDED IN THE WORKLOAD RATING SCALE DEVELOPMENT EFFORT

- o SIMPLE, COGNITIVELY-LOADING TASKS
  CHOICE REACTION TIME, MEMORY SEARCH, MENTAL ARITHMETIC, MENTAL ROTATION, PATTERN MATCH
- o SIMPLE, MANUALLY-LOADING TASKS
  ONE AND TWO AXIS TRACKING
- o CONCURRENT, INDEPENDENT DUAL-TASKS
  TRACKING + MEMORY SEARCH, MENTAL ROTATION
- o SERIAL, INTEGRATED "FITTSBERG" TASKS
  TARGET ACQUISITION + MEMORY SEARCH, MENTAL ARITHMETIC, RHYMING, PATTERN MATCH, PREDICTION, TIME ESTIMATION
- o COMPLEX SUPERVISORY CONTROL SIMULATIONS ("POPCORN")
- o PART-TASK AND FULL-MISSION AIRCRAFT SIMULATIONS

During the past three years, we have conducted a series of 25 experiments in which a multi-dimensional battery of bipolar rating scales were presented to subjects following a variety of tasks. For 15 of these experiments, the ratings, and individual definitions of workload were combined into a data base and a number of global analyses were performed.

The objective was to determine:

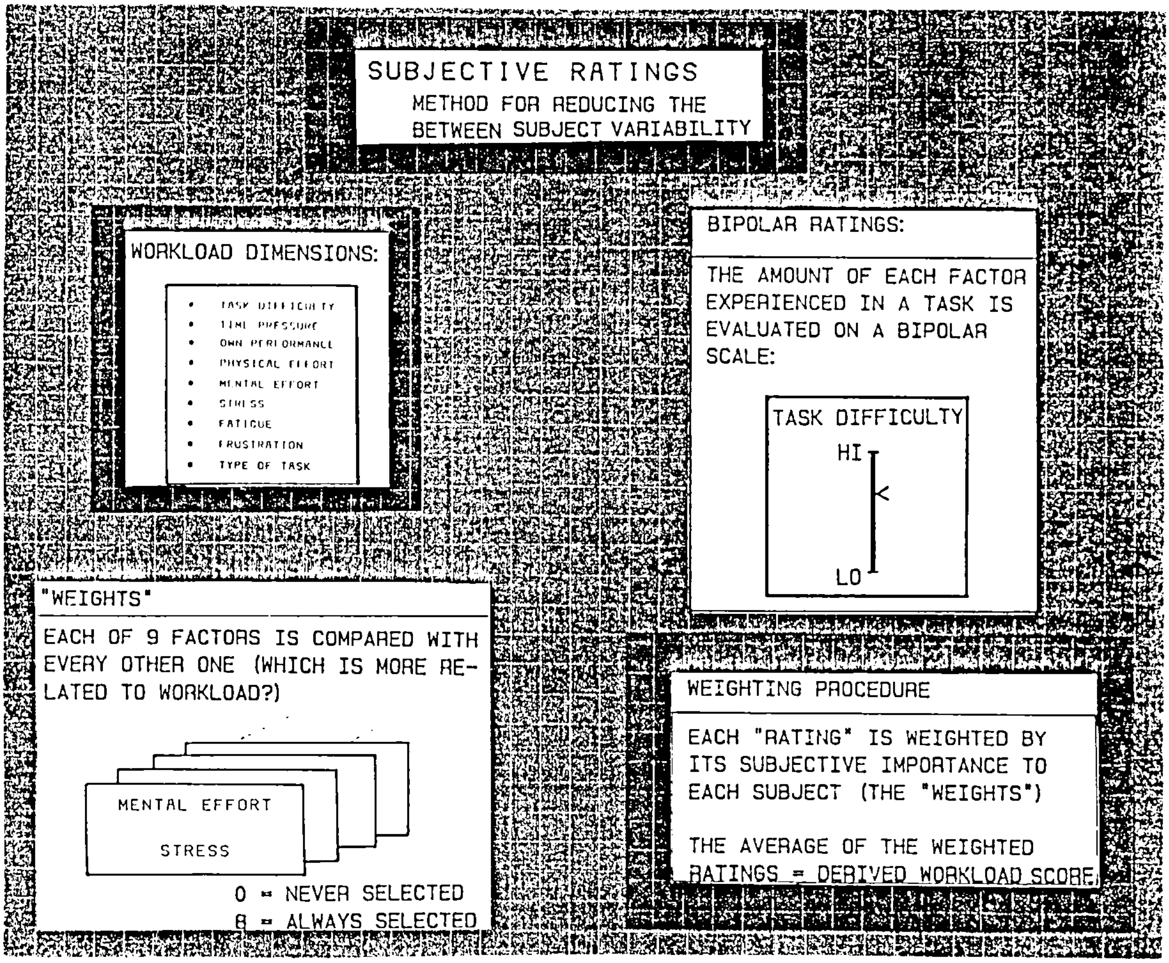
(1) What factors are sensitive to workload differences between different types of tasks
(2) What factors are sensitive to workload differences within tasks
(3) What factors are included in the workload definitions of most individuals
(4) What is the appropriate scale format

The primary problems that we encountered in this effort were:

(1) There is no objective standard against which workload ratings can be compared
(2) The workload of a task is not uniquely defined by its objective demands but represents the behaviors and psychological responses of individual subjects as well
(3) Different individuals may adopt different references activities and have diffferent personal definitions of workload

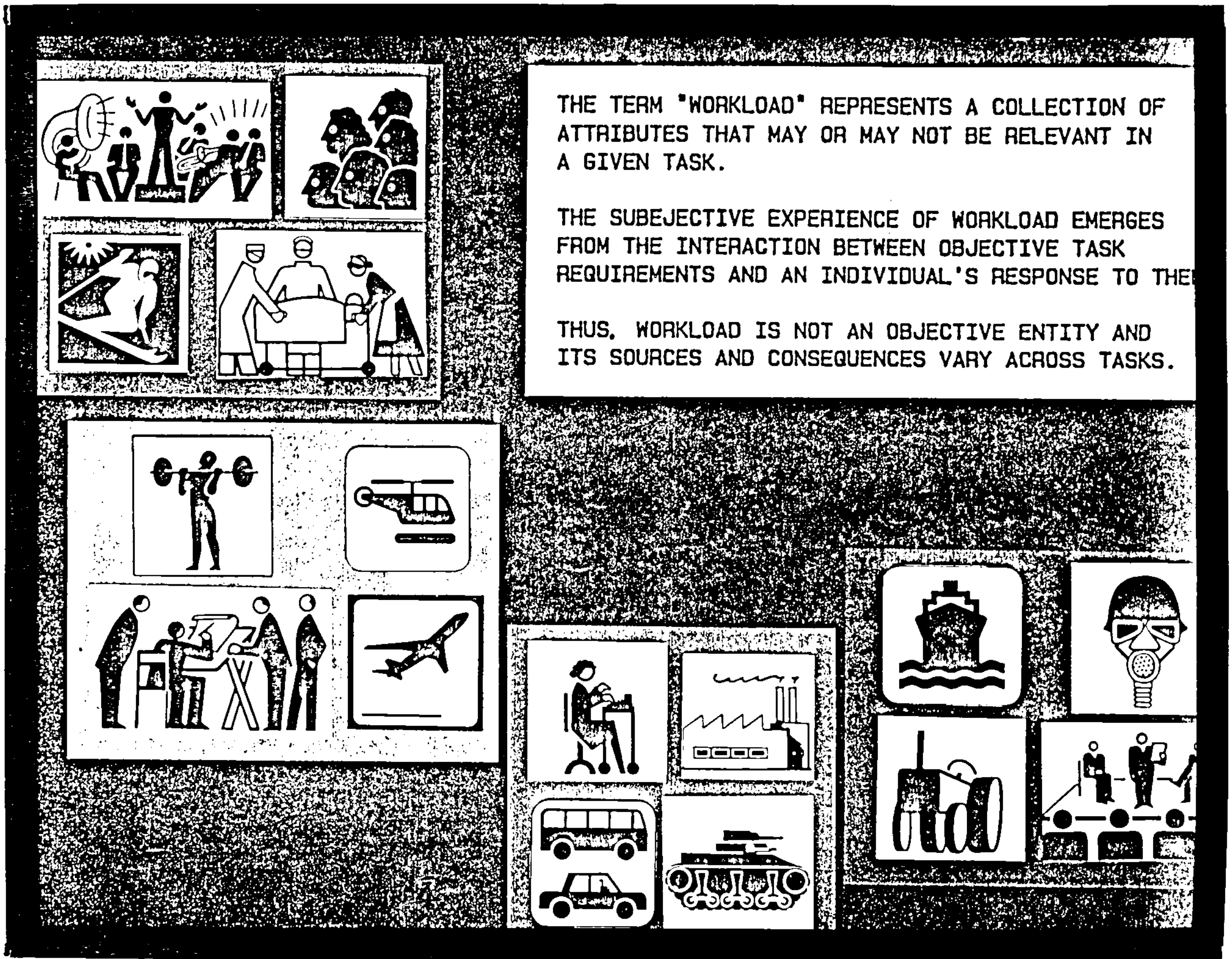
We organized the experimental tasks into six categories. These tasks ranged from simple, cognitively loading tasks to complex aircraft simulations. Several thousand data points were included in each category.

We found that different individuals consider different variables in formulating workload ratings. Thus, one person's overall workload rating might reflect the level of time pressure experienced while another's might reflect the level of cognitive effort exerted or their apprarent failure to accomplish the task requirements. People are generally unaware of the fuzziness of their definitions, however, they are able to express their biases when asked to do so.



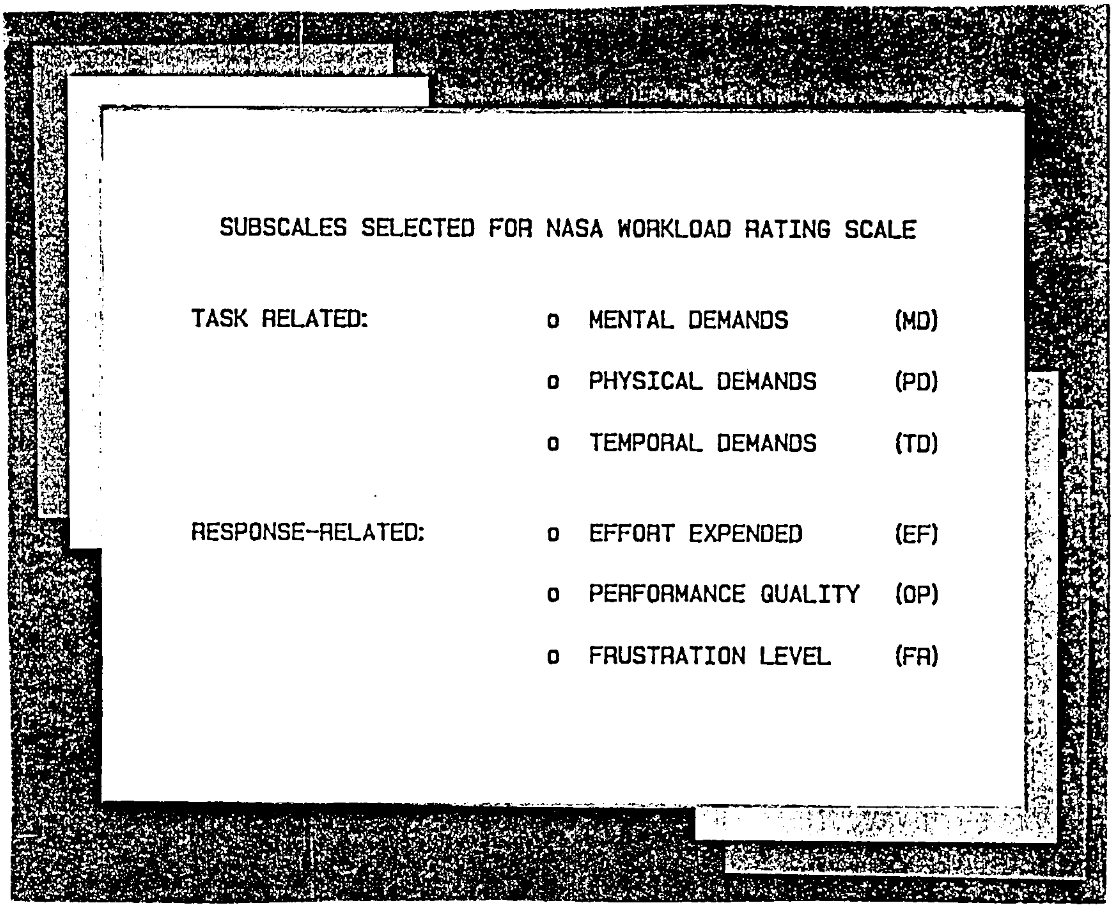
We found, that by weighting the bipolar ratings obtained on the component scales by the subjective importance of each factor to each subject, and by averaging these weighted ratings, we were able to obtain a significant reduction in between-subject variability in a summary estimate of overall workload.

These summary scores reflected the same workload levels indicated by overall workload ratings, but with a 25-50% reduction in variability. However, the sensititvity of the summary measure to experimental manipulations was not significantly enhanced.



Since workload represents a collection of attributes, the sources of workload may vary from one activity to the next as a result of the requirements, equipment, and environment. Thus, the workload of one task or task segment might be created by very heavy physical demands, while that of another by the level of time pressure or danger.

Although individuals may define workload differently, they are, none-the-less responsive to the specific sources of loading imposed by a task. Since the subjective experience of workload emerges from the interaction between objective task requirements and an individual's response to them, we found that it was critically important to determine the subjective importance of specific factors in creating the workload of a specific activity (as well as the magnitudes of those factors) to develop a sensitive and accurate multi-dimensional rating of overall workload.



We found that at least six factors are necessary to discriminate between workload levels within and between tasks. They are:

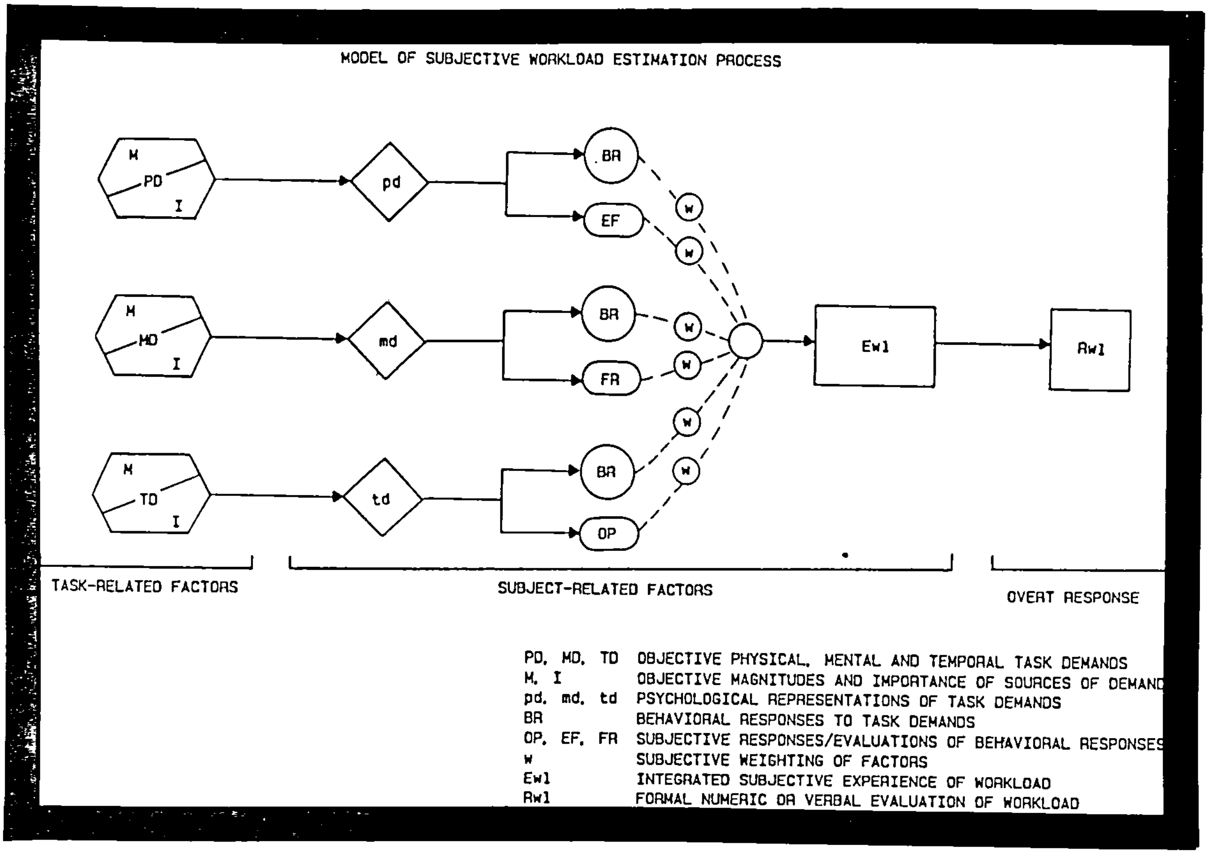
Task related:

Temporal Demands, Physical Demands, and Mental Demands

Subject-related:

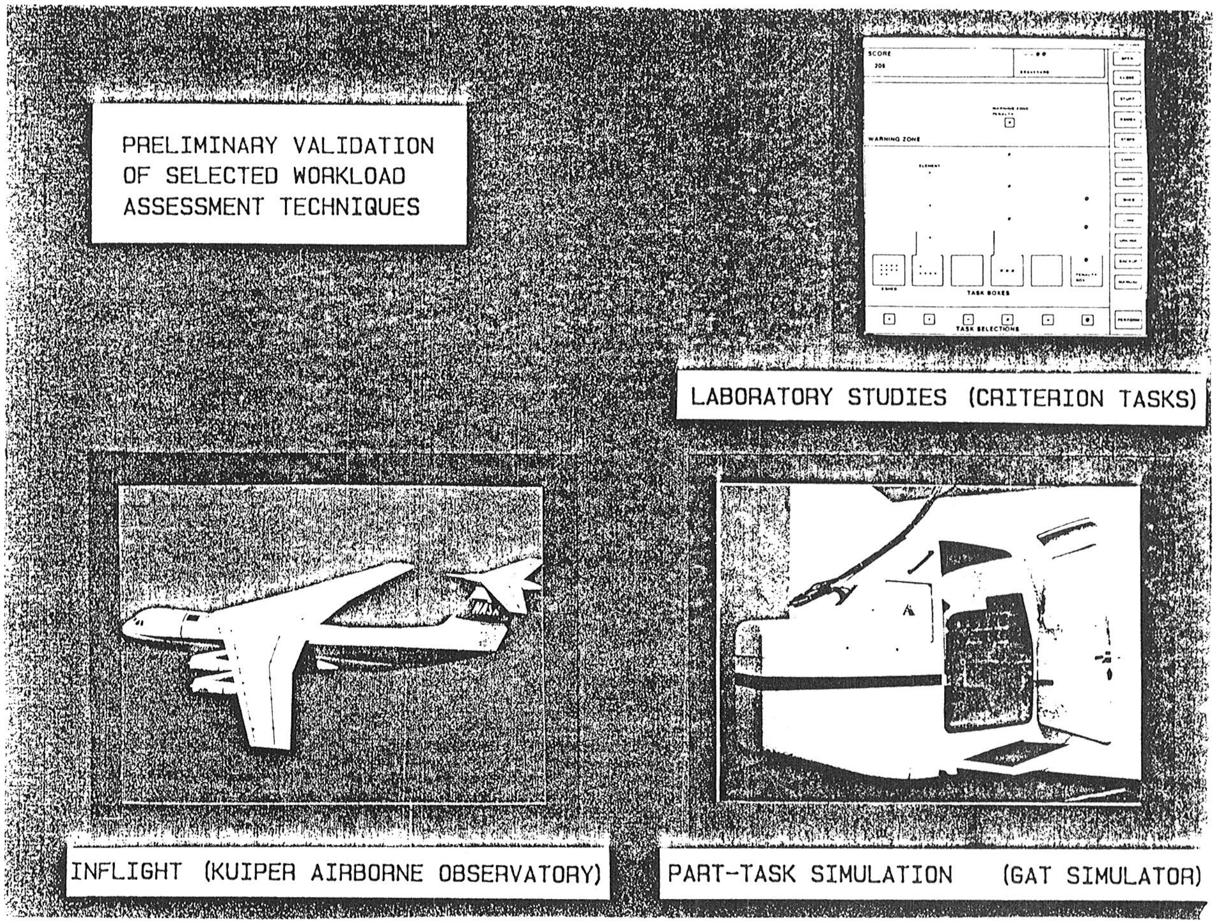
Own Performance, Frustration, and Effort.

Each of these scales alone provides useful, diagnostic, and often independent information about the sources of workload and the experiences of operators. By combining these individual scale values, weighted to reflect their importance in creating the level of workload imposed by a specific task, a global indicator of overall workload can be derived that is less variable between subjects and more sensitive to experimental manipulations than are existing rating technqiues.



A priori workload weights, which form the basis for several popular techniques, do not reflect the objective contributions of specific factors to the workload of a specific task. The model presented in this figure represents the conceptual framework of the rating technique that we developed. Objective demands are imposed on an operator, which are translated into psychological representations. These invoke behavioral and psychological responses from an operator. A weighted combination of the relevant factors - - both objective and subjective - - are integrated into a subjective experience of workload that may be translated in to a numeric or verbal evaluation. The key element of this model is that the integration represents a weighted combination of factors. The weights reflect the objective and subjective importance of the factors to the structure of that task and the ratings reflect the psychological magnitudes of each factors during that activity.

The bipolar rating scale that we propose is two dimensional: evaluations of the magnitude as well as the importance of each of six factors are obtained from subjects following specific tasks or task segments. The combined weighted average of the six factors provides a sensitive and stable measure of overall workload.



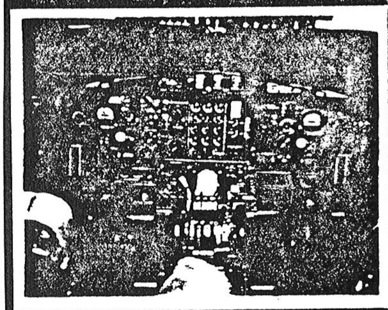
With this measure, as with all of the others, validation is accomplished in a variety of environments. Each measure is tested against criterion tasks that impose known, well-controlled levels of workload. Promising measures are then tested in part-task simulations within our lab. Finally, many measures have been applied - - piggy-back - - on a variety of operational activities to provide "real-world" validation.

VALIDATION OF NASA WORKLOAD ASSESSMENT MEASUREMENT BATTERY

OBJECTIVE: DETERMINE THE SENSITIVITY AND OPERATIONAL VALIDITY OF THE WORKLOAD MEASURES DEVELOPED AT NASA-AMES

APPROACH: CONSTRUCT SCENARIOS WITH WORKLOAD PREDICTIVE MODEL
PERFORM FLIGHTS IN B-727 SIMULATOR AND SH-3 HELICOPTER
COMPARE MODEL PREDICTIONS TO EMPIRICAL RESULTS

MEASURES: PERFORMANCE (COMMUNICATIONS, ERRORS, CREW COORDINATION, CONTROL VARIABILITY, SECONDARY TASKS)
PHYSIOLOGICAL (HEART RATE/VARIABILITY, EYE BLINK RATE/TIMING, SCAN PATTERN, AUDITORY EVOKED CORTICAL POTENTIALS)
SUBJECTIVE (NASA MULTI-DIMENSIONAL SCALE, REFERENCE TASK COMPARISON, MODIFIED COOPER-HARPER SCALE)



MVSRF 727 COCKPIT

SH-3 HELICOPTER

The final validation effort for our workload-assessment battery will be accomplished within the next year. We plan to conduct at least two full-mission studies in which all of the most promising measures will be applied in realistic environments. The test scenarios will be created with the workload predictive model. Two environments have been selected for these studies:

(1) The MVSRF 727 motion-base simulator

(2) A Sea-king (SH-2) helicopter.

Our goal is to provide as complete and as operationally relevant a validation of the measures as possible in a well-controlled and realistic series of flights.

Concurrent with this effort, the predictive model for Space Station application will continue, and it will be validated at JSC in 1987.