

COPY 54

AdA 087604

NAVAL AIR TEST CENTER TECHNICAL REPORT

SY-33R-80
N62269/78/WR/00408
N62269/79/WR/00462
7 July 1980

REPORT NO:
WORK REQUEST NO:
DATE:

FINAL REPORT

EVALUATION OF A PILOT WORKLOAD ASSESSMENT DEVICE
TO TEST ALTERNATE DISPLAY FORMATS AND
CONTROL HANDLING QUALITIES

by

Dr. S. G. Schiflett



2006003284

Approved for Public Release; Distribution Unlimited.

Part I indicates a deficiency, the correction of which is necessary because it adversely affects:

- a. Airworthiness of the aircraft.
- b. The ability of the aircraft (or piece of equipment) to accomplish its primary or secondary mission (or intended use).
- c. The effectiveness of the crew as an essential subsystem.
- d. The safety of the crew or the integrity of an essential subsystem. In this regard, a real likelihood of injury or damage must exist. Remote possibilities or unlikely sequences of events shall not be used as a basis for safety items.

Part II indicates a deficiency of lesser severity than a Part I which does not substantially reduce the ability of the aircraft or piece of equipment to accomplish its primary or secondary mission, but the correction of which will result in significant improvement in the effectiveness, reliability, maintainability, or safety of the aircraft or equipment. A Part II deficiency is a deficiency which either degrades the capabilities of the aircraft or equipment or requires significant operator compensation to achieve the desired level of performance; however, the aircraft or equipment being tested is still capable of accomplishing its mission with a satisfactory degree of safety and effectiveness.

Part III indicates a deficiency which is minor or slightly unpleasant or appears too impractical or uneconomical to correct in this model, but should be avoided in future designs.

NAVAL AIR TEST CENTER
NAVAL AIR STATION
Patuxent River, Maryland 20670

N62269/78/WR/00408
N62269/79/WR/00462
SY-33R-80
7 July 1980

From: Commander, Naval Air Test Center, Patuxent River, Maryland 20670
To: Commander, Naval Air Development Center, Warminster, Pennsylvania 18974
Subj: NAVAIRTESTCEN Technical Report SY-33R-80, Evaluation of a Pilot Workload Assessment Device to Test Alternate Display Formats and Control Handling Qualities, Final Report; transmittal of
Ref: (a) NAVAIRDEVCEN Work Request N62269/78/WR/00408 of 28 Nov 1977
(b) NAVAIRDEVCEN Work Request N62269/79/WR/00462 of 31 Jan 1979
(c) NAVAIRTESTCEN Technical Report SY-27R-80 of 26 Mar 1980

1. References (a) and (b) requested the Naval Air Test Center to develop and evaluate a workload assessment device and operational procedure to support the Human Factors Test and Evaluation Methodology Development Program.
2. Reference (c), an annotated bibliography report on operator workload assessment techniques, and this final technical report complete the requirements of the referenced Work Requests.



R. R. BUEHLER
By direction

DISTRIBUTION:

Wright-Patterson AFB (AFFDL/FGR)	(1)
Wright-Patterson AFB (AMRL/TSZ)	(1)
Wright-Patterson AFB (AMRL/HEB)	(5)
Wright-Patterson AFB (Integrated Aircraft Systems - FDL)	(1)
Wright-Patterson AFB (ASD/ENECC)	(1)
U.S. Army AMRL, Fort Rucker	(3)
U.S. Army HEL, Aberdeen Proving Ground	(1)
U.S. Army Material Systems Analysis Activity, Aberdeen Proving Ground	(1)
NASA-AMES Research Center	(3)
PACMISTESTCEN (Code 1226)	(3)
NAMRL, New Orleans	(1)
NAVAIRDEVCEN (Code 6021)	(5)
NAVAIRSYSCOM (AIR-340F)	(1)
NAVOCEANSYSCEN (Code 823)	(1)
Brooks AFB (USAFSAM/VNE)	(3)
Brooks AFB (SAM/VNE)	(1)
Brooks AFB (AFHRL)	(1)
Edwards AFB (AFFTC/DOEEH)	(3)
Naval Post Graduate School, Monterey	(3)
Office of Naval Research (Code 455)	(3)
Office of Naval Personnel	(1)
White Sands Missile Range, New Mexico	(1)
HQ TCATA, Fort Hood	(1)
Systemetrics, Incorporated, Blacksburg	(1)
Systems Research Laboratories, Incorporated, Dayton (Mr. R. Spicuzza)	(3)
Honeywell Incorporated, Minneapolis (Dr. R. North)	(1)
System Technology, Incorporated, Hawthorne	(1)
Canyon Research Group, Incorporated, Westlake Village	(1)
Dunlap and Associates, Incorporated, LaJolla	(1)
Hughes Helicopter, Culver City (Mr. S. Thompson)	(1)
Bell Helicopter Company, Fort Worth (Dr. D. Strothers)	(1)
Vought Systems Division, Dallas (Mr. J. Burke)	(1)
Boeing Aerospace Corporation, Seattle (Mr. D. Jahns)	(1)
Massachusetts Institute of Technology, Cambridge (Dr. T. Sheridan)	(1)
RAF Institute of Aviation Medicine	(1)

Technion-Israel Institute of Technology, Haifa (Dr. D. Gopher)	(1)
Advanced Research Resources Organization	(1)
Analytics, Willow Grove (Mr. M. Streib)	(1)
Mr. W. Clement, Mountain View, California	(1)
University of South Dakota, Vermillion (LCDR W. Helm)	(1)
University of Illinois, Champaign (Dr. C. Wickens)	(1)
University of Toronto, Ontario (Professor J. Senders)	(1)
University of Dayton, Dayton (Dr. T. Schori)	(1)
University of New Mexico, Albuquerque (Dr. M. Brecht)	(1)
DTIC	(2)

ACKNOWLEDGEMENTS

The development and evaluation funds for this project were provided by the Naval Air Systems Command (AIR-340F) via the Naval Air Development Center (6021), Warminster, Pennsylvania, in support of the Human Factors Test and Evaluation Methodology Development Program.

The workload assessment device was designed and fabricated by the Systems Research Laboratories, installed by the Calspan Corporation, and integrated into the NT-33A avionics system by the General Electric Company.

The author wishes to express special appreciation for coordination efforts by Mr. F. C. Hoerner, Program Manager, Display Evaluation Flight Test Program, and Air Force Liaison Officers, LTCOL C. D. Eliason and Maj R. D. Stanley.

Professional recognition is attributed to Naval test pilots CDR S. J. Fitrell and LCDR C. C. Anderson for providing many insightful comments.

NAVAL AIR TEST CENTER
NAVAL AIR STATION
Patuxent River, Maryland 20670

7 July 1980

EVALUATION OF A PILOT WORKLOAD ASSESSMENT DEVICE
TO TEST ALTERNATE DISPLAY FORMATS AND
CONTROL HANDLING QUALITIES

by

Dr. S. G. Schiflett

ABSTRACT

This in-flight research project evaluated the utility of a Workload Assessment Device (WAD) to measure pilot workload for approach and landing tasks under simulated instrument meteorological conditions, alternate HUD formats and control stability variations. The flight tests were conducted in an NT-33A research aircraft, extensively modified for the Air Force and Navy by the Display Evaluation Flight Test program. The hardware, software, and test procedures associated with the WAD functioned efficiently with only minor discrepancies and minimum pilot distraction. The project established the feasibility of using an item-recognition task as a measure of sensory-response loading and reserve information processing capacity while flying precision approaches. In a descriptive statistical treatment of the data, the results indicate an appreciable increase in reaction time and errors with degraded handling qualities as compared to ground baseline measures and good handling qualities. The preliminary findings also reveal consistent trends toward the availability of more mental reserve capacity when flying predominantly pictorial/symbolic HUD configurations as compared to conventional HUD formats with scales and alphanumerics. It is recommended that further evaluations be conducted to establish the efficacy of utilizing the WAD to measure mental workload in a wide variety of aircrew tasks.

Distribution limited to U.S. Government agencies only; Test and Evaluation; May 1980. Other requests for this document must be referred to Commander, Naval Air Test Center, Patuxent River, Maryland 20670.

TABLE OF CONTENTS

	<u>Page No.</u>
INTRODUCTION	1
BACKGROUND	1
PURPOSE	2
DESCRIPTION OF AIRCRAFT/EQUIPMENT	2
SCOPE OF TESTS	5
METHOD OF TESTS	6
CHRONOLOGY	7
RESULTS AND DISCUSSION	8
GENERAL	8
PRIMARY FLIGHT MEASURES	8
SECONDARY TASK MEASURES	9
REGRESSION EQUATIONS (REACTION TIMES)	9
PERCENT ERRORS	13
CONCLUSIONS	17
RECOMMENDATIONS	18
APPENDIX A - REFERENCES	19
APPENDIX B - FUNCTIONAL CAPABILITIES AND HARDWARE/SOFTWARE DESCRIPTION	20
APPENDIX C - EVALUATION FLIGHT SCENARIO	28
APPENDIX D - HANDLING QUALITIES RATING SCALE	29
APPENDIX E - PRIMARY MEASURES DATA ANALYSIS	30

INTRODUCTION

BACKGROUND

1. New developments in cockpit display designs and integrated weapons system avionics have significantly altered the role of the pilot from that of a skilled, manual control operator to an executive manager of an integrated weapons system. Emphasis on psychomotor control has been augmented by an interest in more cognitive skills represented by such functions as short-term memory, information processing, and decision making. Few measurement techniques exist which are able to provide an objective, reliable, and valid estimate of the subtle differences in workload introduced by these new systems. To date, methodology for objectively quantifying workload has not been effectively applied to the flight test and evaluation of aircrew systems. New approaches to the measurement of workload are, therefore, required.

2. This project introduced a novel approach to the traditional manner of measuring pilot workload. Aircrew workloads are typically measured by subjective assessment rating scales which are based on pilot opinions that relate operational task demands to system response characteristics, e.g., Cooper-Harper Handling Qualities Rating Scale. The new approach applied in this project is an item-recognition task first identified by Sternberg (reference 1)⁽¹⁾ and modified by the Air Force (reference 2) to measure the reserve capacity of the pilot. The approach assumes that an upper bound exists on the ability of the pilot to gather and process information. As the pilot's workload increases on the primary task, i.e., flying the aircraft, reserve capacity for processing information decreases until a point of overload is reached by the pilot. At this point, the information processing demands of the task exceed the pilot's total workload capacity and is manifested by degradation in performance (i.e., increase in errors and response times) on the secondary item-recognition task.

3. The theoretical formulation of the item-recognition task, as proposed by Sternberg (figure 1), has several attractive features which make it ideally suited for evaluating the source of increase in task-loading in aircraft test environments. The theory assumes a least-squares, linear regression fit of the data where the intercept represents the input/output component and the slope depicts the mental information processing component of the item-recognition task. If, for example, the sensory-response mode (i.e., input/output, stages 1 and 3 of figure 1) are impaired by degraded display characteristics or multiple response overload, the theoretical expectation is a change in the y-intercept of the regression line with no change in slope. Conversely, if the source of task-loading was one which affected the pilot's mental information processing capabilities (e.g., short-term memory overload), the expectation is a change in the slope of the curve without a corresponding change in the intercept value. Either result would be a decrease in the pilot's reserve capacity for processing information.

⁽¹⁾ All references are cited in appendix A.

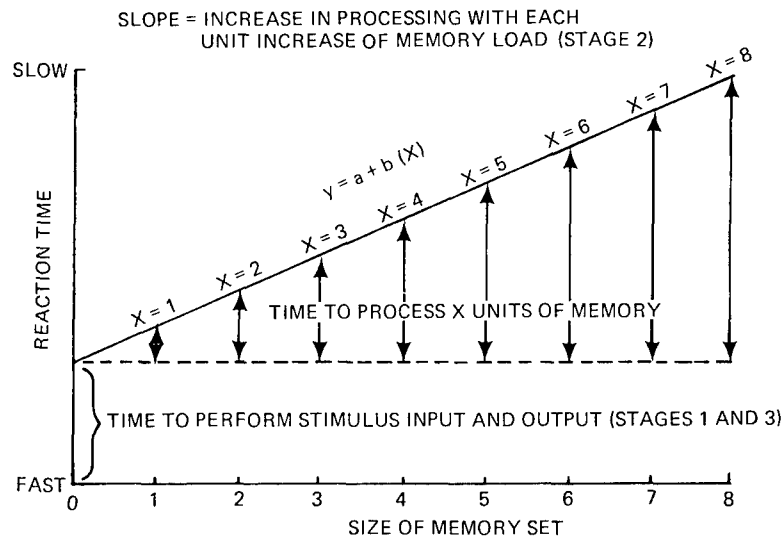


Figure 1
Theoretical Components of the Item
Recognition Task Proposed by Sternberg
where a = Intercept and b = Slope

4. The use of the item-recognition task to assess primary task workload is not a new concept in aircrew flight simulation studies (references 3 and 4). However, the uniqueness of its application in this project is that a Workload Assessment Device (WAD) that generates and controls the secondary item-recognition task was designed, fabricated, and installed in an NT-33A aircraft to measure and analyze the pilot's reserve workload capacity for the Display Evaluation Flight Test (DEFT) program.

PURPOSE

5. The purpose of this project was to evaluate the utility of the WAD to measure pilot workload for approach and landing tasks under simulated Instrument Meteorological Conditions (IMC's) for alternate HUD formats and aircraft control stability variations.

DESCRIPTION OF AIRCRAFT/EQUIPMENT

6. The NT-33A variable stability aircraft is an extensively-modified, T-33 jet trainer. The elevator, aileron, and rudder controls in the front cockpit were disconnected from their respective control surfaces and connected to separate servo-mechanisms that make up an "artificial feel" system. In addition, the elevator, aileron, and rudder control surfaces were connected to individual servos which were driven by a number of different electrical inputs. This arrangement, through a response-feedback system, allowed the normal T-33 stability derivatives to be augmented to the extent that the handling qualities of the hypothetical research configurations were simulated. A more comprehensive description of the NT-33A can be found in reference 5.

7. The DEFT system provided a fully software-programmable display system to complement the variable stability features of the host-modified NT-33 aircraft. Relative to the aircraft configuration, the DEFT system provided the capability of changing display formats and changing the algorithms and dynamics of the display driving signals. The display system consisted of a HUD, two digital computers, a magnetic tape system, an INS, sensors to augment the existing aircraft sensors, and a display repeater and mode control unit for the aft cockpit.

8. The software programs provided an in-flight choice of two uniquely different display configurations for use in the approach and landing phases of flight. These displays were of a conventional F/A-18-based format (figure 2) and the predominantly symbolic Klopstein format (figure 3). As depicted in the figures, the F/A-18 display used a conventional format with a flight path ladder, scales, and alphanumeric readouts of various flight parameters. The Klopstein display, however, is predominantly symbolic depicting the horizon, an artificial runway overlaying the actual runway, and other flight guidance symbols.

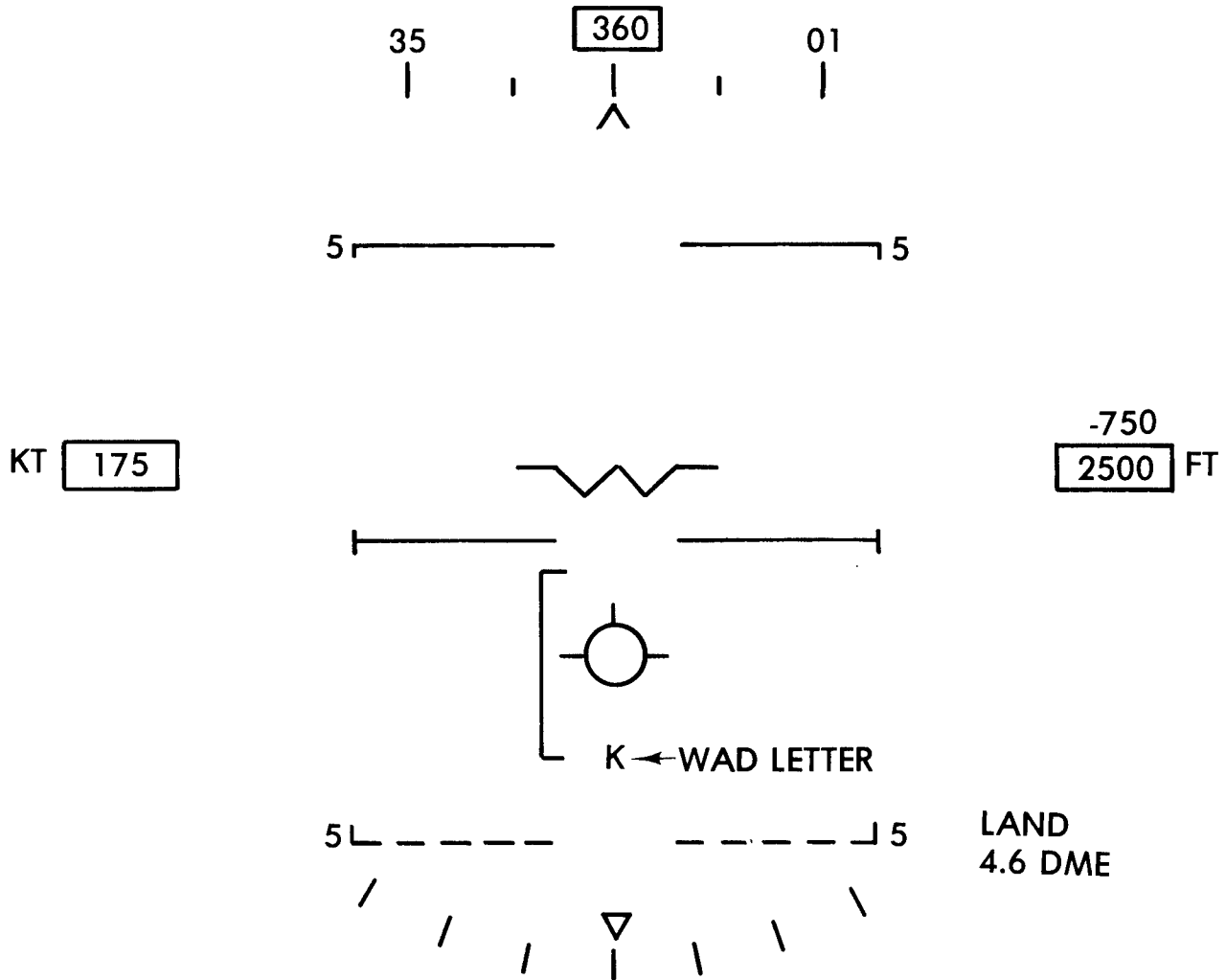


Figure 2
F/A-18 Based HUD Format

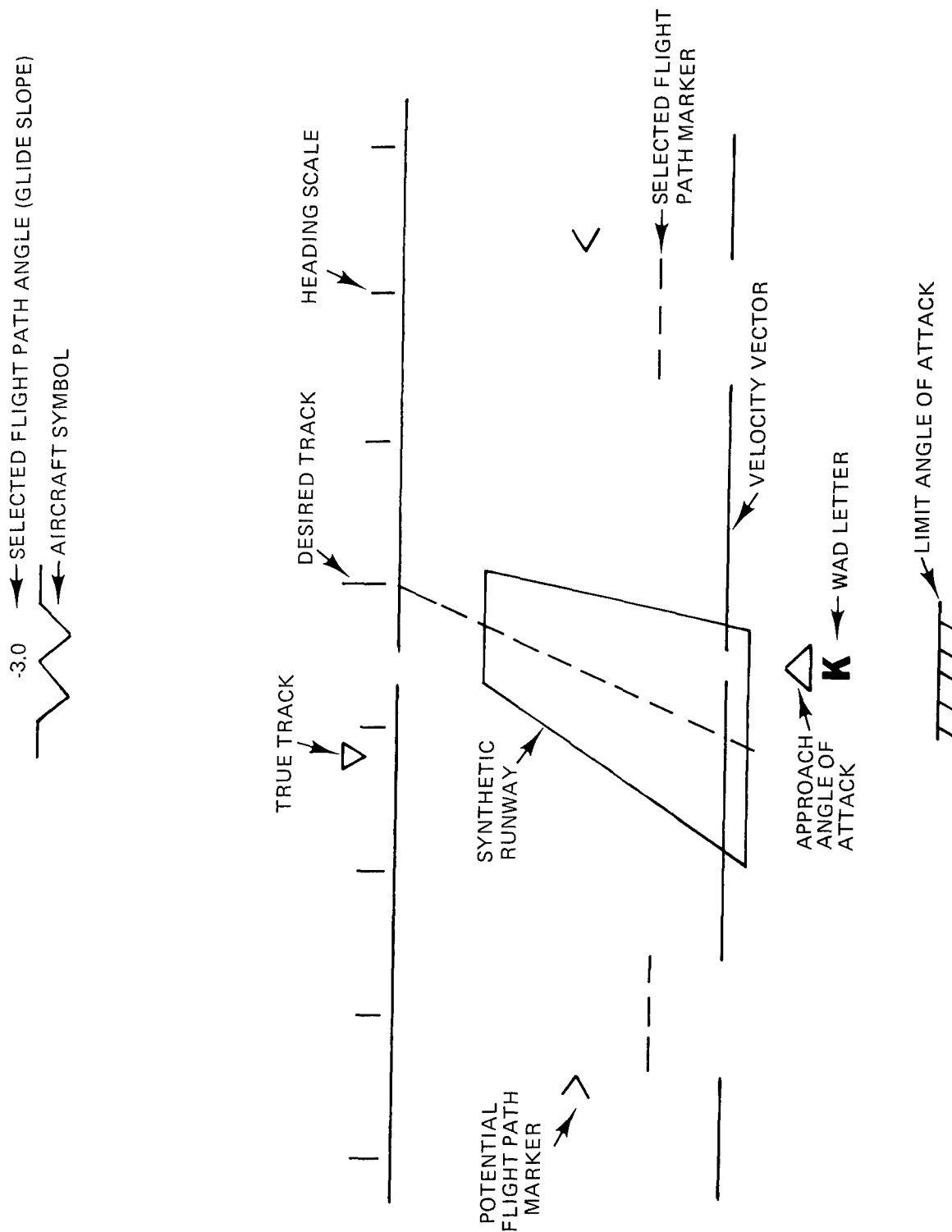


Figure 3
Klopstein HUD Format

9. The WAD consists of two basic units: the airborne controller and the ground-based analysis center. The controller is configured for installation in the front avionics bay of the NT-33A aircraft. The unit provides the electronics, power supplies, software, interfaces to the HUD and the aircraft intercom, rear cockpit initialization switches, control stick response switches, and data recording system necessary to perform a complete series of item-recognition experiments. In addition, the controller can operate as a standalone laboratory system capable of performing the same tasks as when airborne. The ground-based data analysis center is used to initialize several software options of the controller and to reduce and analyze response time data. A description of the functional capabilities of the hardware and software is discussed in appendix B. A detailed description of the complete WAD system is contained in reference 6.

SCOPE OF TESTS

10. The tests were conducted in Buffalo, New York, at the Calspan Corporation, with on-site support from the General Electric Company and Systems Research Laboratories, Incorporated. A total of 10 flights was flown by two Navy pilots at Niagara Falls International Airport or Buffalo International Airport. The flights consisted of two orientation flights and eight evaluation flights of approximately 1 1/2 hr each. Each pilot flew two evaluation flights using the F/A-18-based HUD format and two with the Klopstein format.

11. During each evaluation flight, a pilot performed eight approaches terminating in either a low approach or touch-and-go landing for a total of 32 approaches per pilot. One-half of the approaches for each flight were made using "good" handling qualities, the other half were made using either "fair" or "poor" handling qualities. The handling qualities were manipulated by changing the pitch response of the aircraft after every four approaches. The response of the roll and yaw axes was held constant throughout the tests. The handling qualities for the pitch characteristics are shown in table I.

Table I
Pitch Configuration Characteristics

Quality	ω_{sp}/ζ_{sp}^*	n_z/a^{**}	Time Delay msec
Good	2.6/0.7	5.6	0
Fair	2.6/0.7	5.6	150
Poor	2.6/0.7	5.6	200

*Longitudinal short period frequency to damping ratio (rad/sec).

**Longitudinal response sensitivity (G/deg).

METHOD OF TESTS

12. After several practice sessions and prior to the start of the evaluation flights, a baseline measurement was obtained on the item-recognition task. Each pilot was given the item-recognition task for each memory set size while sitting in the cockpit of the aircraft stationed on the ground. The task required the pilot to memorize sets of one, two, or four letters, i.e., A, RJ, ZPNW. The pilot was then instructed, prior to testing with each memory set, which set of letters would be presented for memory recall. The prememorized letters (positive) or other letters (negative) were presented on the HUD one at a time every 7 sec. The positive and negative letters were presented with a .5-probability of occurrence. Each letter appeared on the HUD until the pilot responded or 5 sec elapsed. The pilot responded to a letter presentation by pressing one of two designated buttons on the control stick. One button indicated that the letter was a member of the prememorized set (positive) and the other indicating it was not a member of the prememorized set (negative). Positive letters never appeared as negative letters and the same positive letter sets were used throughout the test. A total of 30 letters, 15 positive and 15 negative, was presented for each memory set for the baseline conditions.

13. The same procedures were used in flight as during the baseline test conditions with the exception that the pilot was flying the aircraft while performing the secondary task. An additional experimental control allowed one approach per handling quality/display format combination to be flown without any letter presentations to evaluate the impact of the secondary task on the primary task of flying the aircraft.

14. The reaction times and response errors were collected and analyzed by the WAD controller and ground-based analysis system. After each response, the reaction time was measured from the onset of a letter to the physical response of pressing the correct button. The reaction times for both the positive and negative letters were stored on cassette tapes. The reaction times for the correct responses were then averaged and plotted as a function of the memory set sizes. The response errors were coded, tabulated, and categorized by type of error and frequency of occurrence. A response was considered an error if the pilot pressed the wrong key (reversal error), responded correctly but after 1,500 msec (out-of-bound error), or did not respond before 5 sec (time-out error).

15. The basic flight scenario for each approach and touch-and-go was as follows. The Evaluation Pilot (EP) was given control of the T-33 by the Safety Pilot (SP) with the desired display-aircraft handling quality combination. The EP then flew on instruments while using an orange filter over the windscreen and a blue visor attached to the helmet to simulate IMC⁽²⁾. After intercepting the glide slope, the EP descended to 1,800 ft MSL to intercept the localizer at 8 nmi. At this point, the SP turned on the digital recorder and the WAD controller which were used to record the primary flight measures and the secondary task measures, respectively. The EP proceeded to fly the glide slope and the localizer to perform the approach. The outer marker was at approximately 4 miles. At 200 ft AGL and approximately 1/2 nmi from the runway threshold, the EP "broke out" (i.e., he lifted the blue visor) and flew visually for the remainder of the low approach (20 ft AGL). If conditions permitted (fuel state, crosswind, etc.), the EP then performed the touch-and-go landing, minimizing the sink rate on touchdown to less than 3 ft/sec. The touchdown point was a 500-ft zone, 1,500 ft from runway threshold. After liftoff and at approximately 200-ft AGL,

⁽²⁾ Overlaying the two complementing colors produced a perceptual environment similar to night IMC when the pilot attempted to view the external world.

the SP turned off the WAD controller and the digital recorder. After four approaches, the SP assumed control of the aircraft, then changed the pitch handling quality to the next desired setting and again released control of the aircraft to the EP. A schematic plan view of the evaluation flight scenario is presented in detail in appendix C.

16. After each block of four approaches was completed under the same pitch handling quality, the EP and SP rated the approach and flare/landing segments of the flight profile using the Cooper-Harper pilot rating scale as shown in appendix D. Additional commentary data were gathered from the EP and SP throughout the flight tests by use of an audio tape recorder, e.g., comments on degree of air turbulence.

CHRONOLOGY

17. The major milestones and associated dates are as follows:

- | | |
|-----------------------------------|-------------------|
| a. Work Request received | 7 December 1977 |
| b. Contract awarded | 21 July 1978 |
| c. Device delivered | 1 April 1979 |
| d. Functional checkouts completed | 27 August 1979 |
| e. Flight evaluations started | 10 September 1979 |
| f. Flight evaluations completed | 19 September 1979 |
| g. Data analysis completed | 21 February 1980 |

RESULTS AND DISCUSSION

GENERAL

18. The test and evaluation paradigm used in this project was a repeated measures design in which type of display format (F/A-18-based versus Klopstein), flight handling quality (good versus poor), and secondary task difficulty (memory set sizes, 0, 1, 2, and 4) were fractionally combined to form 16 different conditions. It was planned that the two EP's would be exposed to each of the 16 conditions twice. However, due to schedule conflicts, funding constraints, high crosswinds, and inoperative equipment, each EP was able to complete all combinations of the test conditions only once.

19. Out of a total of eight 1.5-hr evaluation flights, a complete set of secondary task data was analyzed for only four flights. A partial set of data was gathered for two additional flights but, due to temperature-related failures of the write-out mechanism of the WAD tape transport system, these data were not analyzed in this report. When it became apparent that repeated measures could not be obtained for all test conditions, the remaining two flights were utilized in exploring the effect of a "fair" pitch-handling quality on the Klopstein display for each subject.

20. The results showed that the general procedures established for the conduct of the evaluation flight tests of the WAD were acceptable to the pilots. The in-flight procedures provided the EP's and SP's with reliable guidelines for efficient crew coordination during successful approaches and during incidents of equipment malfunction. Pilot comments aided in the investigation of the most salient characteristics of the item-recognition task including the selection, location, and timing of the letters as presented on the HUD. A thorough testing of the WAD procedures during the project resulted in only minor software changes and hardware replacements and clearly established the feasibility of using the recording item-recognition task. This project also yielded recommendations for improved alterations to the WAD software for subsequent projects which will include establishment of an improved criterion for a cross-coupled, adaptive secondary task paradigm.

PRIMARY FLIGHT MEASURES

21. The primary flight measurement data taken by Calspan Corporation from the digital recorder was divided into two defined subtasks of approach and flare/landing. The statistical data parameters selected to describe the investigative areas of interest are shown in appendix E. Because of the length and complexity of the analyses of the primary flight measurement data, the results will be published under separate cover by Calspan Corporation.

22. However, preliminary results of these analyses from Calspan Corporation indicate that the primary flight performance parameters and Cooper-Harper ratings showed a general inconsistency between displays and handling qualities during the approach and flare/landing phases of the flight task. Lack of systematic differences in the primary flight measures and Cooper-Harper ratings suggests that pilot performance remained the same for all conditions.

SECONDARY TASK MEASURES

23. Due to the small subject sample size ($N=2$) and the lack of repeated measures as noted in paragraphs 18 and 19, the use of an Analysis of Variance (ANOVA) statistical treatment of the data was precluded. Analysis was restricted to exploring trends in the graphic and tabulated reaction time and response error data.

REGRESSION EQUATIONS (REACTION TIMES)

24. The reaction times associated with each correct response were averaged for the complete flight profile for each m-set size (letters 1, 2, or 4), handling quality (good or poor), and display format (F/A-18-based or Klopstein). Linear regression equations were then calculated to indicate the slope and intercept of the plotted data as shown in figures 4 and 5. The data reveal that both the intercept and slope of the curves for each pilot increased from baseline conditions when the handling qualities were degraded. These results indicate that the WAD is sensitive to the increased sensory/response and mental processing requirements imposed by the addition of a primary task and to the level of difficulty of that task. For example, the largest intercept and slope changes occurred between each subject's relative baseline and poor handling quality condition.

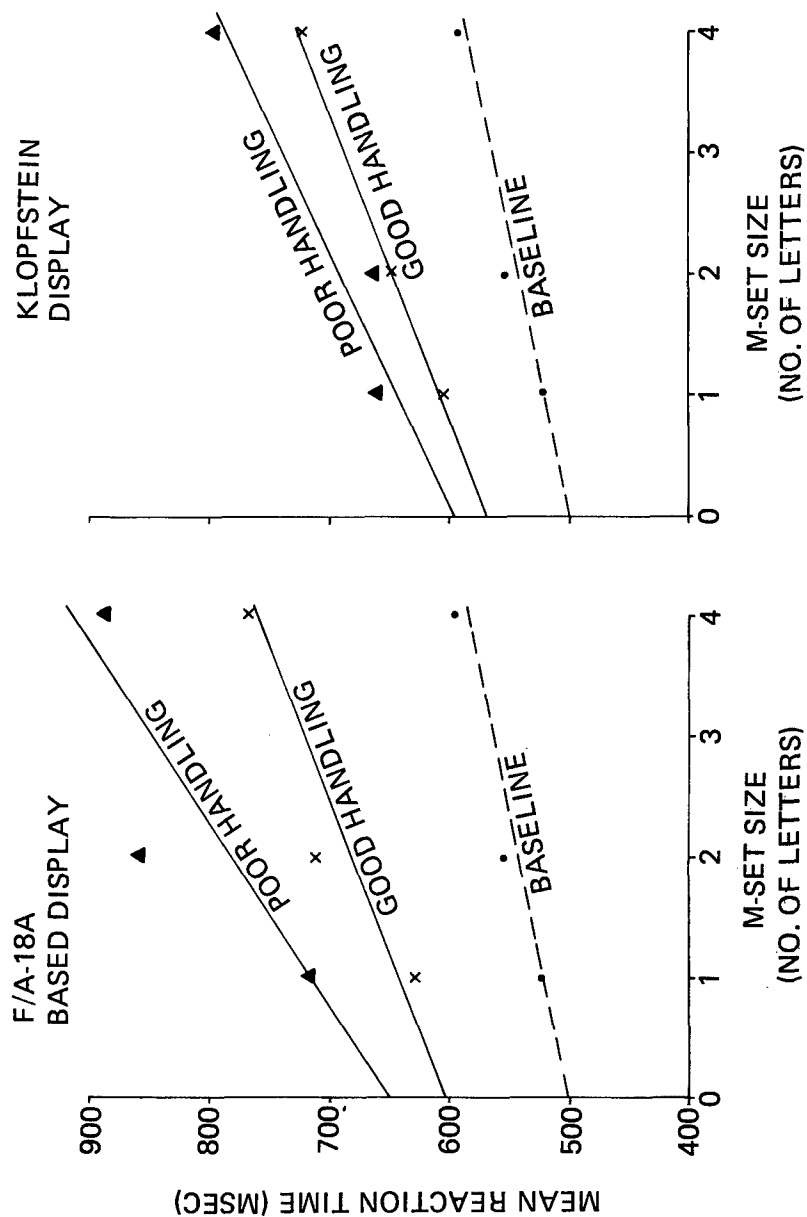


Figure 4
 Linear Regression Lines for Evaluation Pilot No. 1
 Plotted as a Function of Reaction Time and Memory Set Size
 for Flight and Baseline Conditions

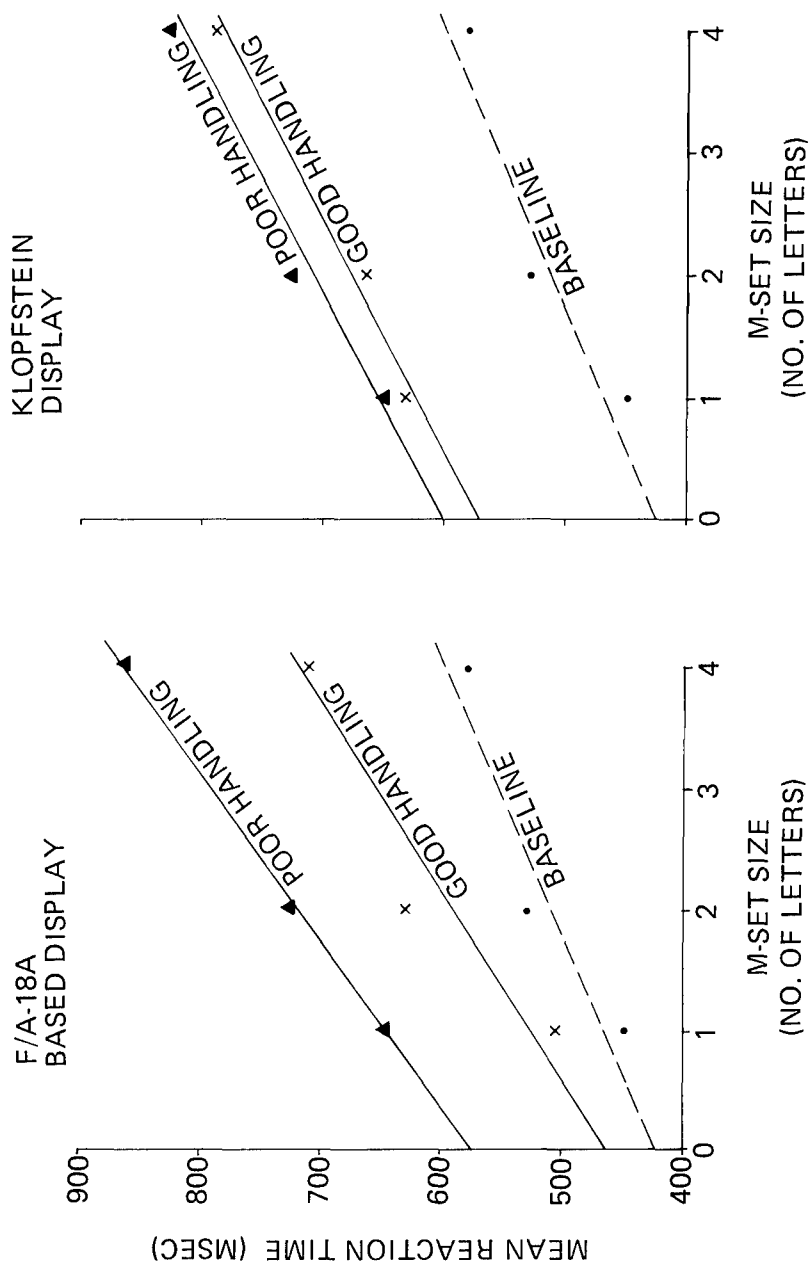


Figure 5
 Linear Regression Lines for Evaluation Pilot No. 2
 Plotted as a Function of Reaction Time and Memory Set Size
 for Flight and Baseline Conditions

25. A closer examination of the data, as tabulated in table II, reveals that the differences in the magnitude of change in the slopes were consistently larger for the conventional F/A-18 HUD format than the pictorial Klopstein HUD format under either good or poor handling qualities. This trend, relative to each subject's shift in slope magnitude, suggests that more mental reserve capacity was available to process information while flying the Klopstein display than the conventional F/A-18 HUD format and while flying good handling qualities independent of the type of display format used.

Table II.a

Linear Regression Slopes and Intercepts
for Evaluation Pilot Number 1

Linear Regression	Baseline	Handling Quality			
		Good Display		Poor Display	
		F/A-18	Klopstein	F/A-18	Klopstein
Slope	24.78	43.77	39.58	66.40	48.08
Intercept	502.29	602.25	569.80	651.40	594.75

Table II.b

Linear Regression Slopes and Intercepts
for Evaluation Pilot Number 2

Linear Regression	Baseline	Handling Quality			
		Good Display		Poor Display	
		F/A-18	Klopstein	F/A-18	Klopstein
Slope	42.16	66.65	53.69	72.78	56.49
Intercept	423.35	460.52	572.22	575.30	604.92

26. Reviewing the resulting changes in intercepts revealed a similar trend with regard to the handling quality parameter. The average increase in the magnitude of change in intercept was less for conditions of good handling qualities than for poor handling qualities. However, with regard to the display variable, the trend was reversed from that observed for the changes in slope; i.e., the average intercept value changed less for the F/A-18-based format than for the Klopstein. Assuming the observed trends would persist in a larger data sample, the results indicated that degrading the handling qualities had a consistent effect on the input/output stages of the item-recognition task, whereas the effect of the display format variable on the input/output stages of the task was subject to inconsistent individual differences. The lack of consistent trends in the changes in intercept relative to the display variable may be due to: (1) individual differences in establishing a time-error tradeoff⁽³⁾, (2) locations of the letter in relationship to differences in eye scan patterns, and/or (3) different strategies of memory recall.

27. These results suggest that degrading handling qualities had a consistent and predominant effect of reducing the pilot's reserve capacity for all three stages of the information-processing, secondary task. Changing the display formats appeared to yield similar results but are subject to the influences of individual differences with regard to the mental component of the information processing task.

28. The reader is reminded that these data reveal only trends and were gathered from a sample of two pilots. Additional flight data are required with a larger pilot sample and more replications of test conditions before definitive conclusions can be made concerning the reliability of the results of these measures.

PERCENT ERRORS

29. The WAD provided an accumulative record of the number of errors, sequence of occurrence, type of error, and reaction time associated with each error for both positive and negative letters. The combined percent of secondary task errors for both pilots is shown in figure 6. The error data show that as the difficulty of the secondary task was increased, i.e., as the m-set size increased, a corresponding decrease in response accuracy was observed which supports the expectation of increased error rate under conditions of task overloading.

⁽³⁾The EP's were only instructed to respond as quickly and accurately as possible to the secondary task while flying a precision approach and landing.

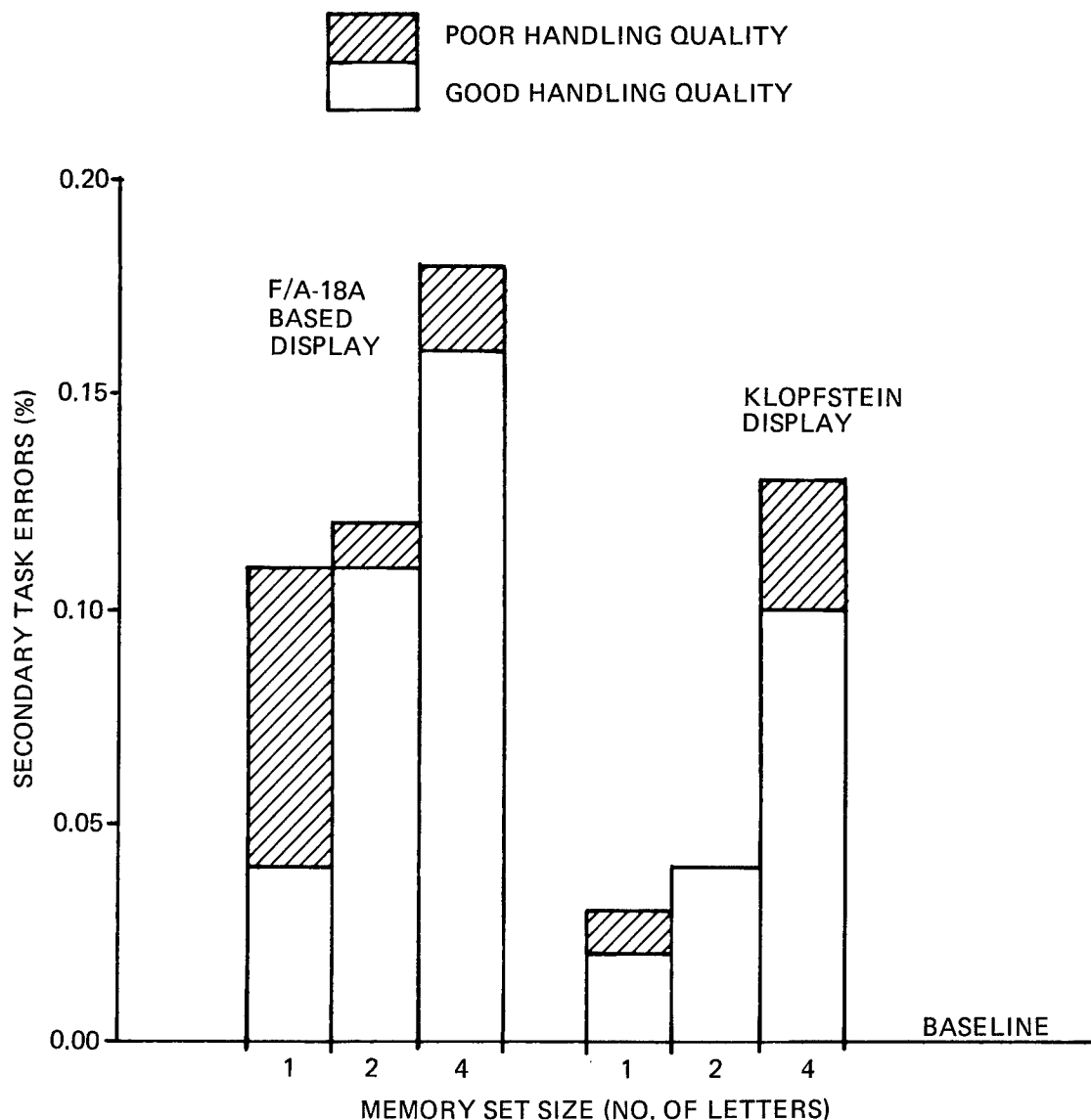


Figure 6
Mean Percent Secondary Task Error for Memory Set Size
(Number of Letter) by Display Format and Handling Quality

30. The increases found in secondary task response errors under conditions of poor handling qualities for both display formats are consistent with the results of the slope and intercept reaction time data with regard to the influence of degraded handling qualities. That is, under test conditions producing a reduction in reverse capacity, a corresponding increase in response errors occurred.

31. In contrast, the reaction time data indicated that the type of display format differentially influenced both the input/output and mental stages of the information processing task, whereas response error data showed a consistently higher degree of response accuracy under conditions of the pictorial Klopstein display format.

32. To further explore these results, the total percent errors were classified into the type of error for each handling quality and display format. The secondary task errors in table III reveal that the total percent errors were evenly distributed between incorrect responses (reversals), late responses (out-of-bounds), and no responses (time-outs). However, when the total percent errors are differentiated between display type and handling quality, it clearly shows that three times as many reversal errors were committed by the EP's flying the F/A-18-based HUD than the Klopstein display format. Degrading the handling qualities increased the percentage of time-out errors for the EP's flying with the F/A-18 display and increased the out-of-bounds under the Klopstein display format. Since it was assumed that a time-out error would reflect a greater decrement in reserve capacity than an out-of-bounds error, these results would imply that the EP's had less reserve capacity while flying under the F/A-18-based HUD and degraded handling qualities than the Klopstein display format.

Table III

Total Percent Error for Each Type Secondary Task
Error by Display Format and Handling Quality

Type of Error	Total Percent Error	F/A-18 Display Handling Quality			Klopstein Display Handling Quality		
		Good	Poor	Total	Good	Poor	Total
Reversal	.12	.04	.05	.09	.01	.02	.03
Time-out	.10	.01	.05	.06	.03	.01	.04
Out-of-Bounds	.14	.04	.04	.08	.01	.05	.06
Total Percent Errors	.36	.09	.14	.23	.05	.08	.13

NOTE: Reversal - Incorrect response (wrong key).

Time-out - No response (>5,000 msec).

Out-of-Bounds - Late response (>1,500 msec, <5,000 msec).

33. An additional implication that has a direct impact on the results of this project and may involve a potential improvement to the WAD scoring strategy is the selection of an appropriate criterion for an error. The empirically-derived criterion for defining an out-of-bounds error was set at 1,500 msec. This boundary for error measurement was utilized to reinforce explicit instructions to respond as quickly as possible, identify correct but delayed response strategies, and to reduce the variability associated with extreme reaction times. However, if one selected a less stringent or more restrictive criterion for error measurement, then it could have produced alterations in the slopes and intercepts of the regression equations due to the skewed distribution of the percent of errors, i.e., four-letter set contained a larger percent of errors. Since a possible erroneous interpretation could result from the use of an inappropriate boundary criterion, it is recommended that additional empirical analysis be conducted to assess the credibility of this technique of error measurement.

34. In summary, the percent of secondary task errors increased whenever the memory set size increased, the handling qualities were degraded, and the task was performed in flight under the F/A-18-based display format conditions. Poor handling qualities primarily induced errors of delay or no response while the type of display affected accuracy of response.

CONCLUSIONS

35. The hardware, software, and test procedures associated with the Workload Assessment Device (WAD) functioned efficiently with only minor discrepancies and minimum pilot distraction (paragraph 20).

36. The project established the feasibility and sensitivity of using a secondary item-recognition task as a measure of sensory/response loading and reserve information processing capacity while flying precision instrument meteorological conditions approaches (paragraphs 20 and 24).

37. The pilots showed an appreciable increase in reaction time and percentage of errors on the secondary task flown under poor handling qualities as compared to good handling qualities and ground baseline conditions (paragraphs 24, 27, and 30).

38. The WAD revealed that the pilots had less secondary task errors, more mental reserve capacity, but longer reaction times attributed to sensory/response delays while flying with pictorial/symbolic HUD configurations (Klopfstein) than conventional HUD formats (F/A-18-based) (paragraphs 26 and 34).

RECOMMENDATIONS

39. It is recommended that additional flight data be collected with a larger sample of pilots and more replications of test conditions before specific definitive conclusions can be made concerning the validity and reliability of these measures (paragraphs 23 and 28).

40. It is recommended that additional investigations be conducted to evaluate the credibility of using an out-of-bounds error measurement technique (paragraph 33).

REFERENCES

1. Sternberg, S.; High Speed Scanning in Human Memory, Science, 1966, 153, 652-654.
2. Spicuzza, R. J., Pinjus, A. R., and O'Connell, R. D.; Development of Performance Assessment Methodology for the Digital Avionics Information System, Systems Research Laboratories, Dayton, Ohio, 1974.
3. O'Donnell, R. D.; Secondary Task Assessment of Cognitive Workload in Alternate Cockpit Configurations, Wright-Patterson Air Force Base, Ohio, Aerospace Medical Research Laboratory, AMRL-TR-75-49, 1975 (AGARD-CPP-181).
4. Crawford, B. M., Pearson, W. H., and Hoffmann, M.; Multi-Purpose Digital Switching and Flight Control Workload, Wright-Patterson Air Force Base, Ohio, AMRL-TR-78-43, 1978 (AGARD-AG-246).
5. Huber, R. W. and Parrag, M. L.; Flight Manual Supplement for the United States Air Force/Calspan NT-33A 51-4120 Variable Stability Airplane, Calspan Corporation, Buffalo, New York, T.O. 1T-33A-1 Supplement (Revision B), Jun 1979.
6. Spicuzza, R. J.; Workload Assessment Device Model WAD 8085 Manuals: Volume I - Installation/Operation, Volume II - Hardware/Software Description and Theory of Operation, Systems Research Laboratories, Dayton, Ohio, 1979.

FUNCTIONAL CAPABILITIES AND HARDWARE/SOFTWARE DESCRIPTION

The SRL WAD 8085 Workload Assessment Device consists of two basic units - the airborne controller (figure 1) and the ground-based data analysis center (figure 2). The controller is configured to fit into the front end of an NA-T33A Navy aircraft (figures 3 and 4). The unit provides the electronics, power supplies, software, interfaces to the General Electric (GE) HUD and the aircraft intercom, rear seat initialization switches, response stick switches, and data recording system required to perform a complete series of item recognition experiments. In addition, the controller can operate as a standalone laboratory instrument capable of performing the same tasks as when airborne. The ground-based data analysis center is used to initialize several software options for the controller and to reduce and analyze response time data.

HARDWARE DESCRIPTION

A functional block diagram of the WAD 8085 is shown in figure 5. The CPU Board (2) in the airborne box (1) presents the stimulus parameters to the subject and collects and records his responses for later data analysis. The heart of the unit is an Intel 8085 microprocessor utilizing software routines located in erasable programmable read only memory (EPROM). The processor is mounted on a printed circuit board using a standard S-100 bus design. Circuits to provide 4 levels of vectored interrupts, 6144 bytes of EPROM, a 14-bit real-time clock, 256 bytes of random access memory (RAM), two 8-bit latched and buffered output ports, one 8-bit unlatched input port, one 6-bit input/output port, and two RS 232 serial I/O ports are also located on the processor board.

One of the vectored interrupts (RST 7.5) is used for a real-time clock input with resolution of 1 millisecond per input pulse.

The 6144 bytes of EPROM are used to hold software for a debug monitor and routines for the operation of the task. Coresident with the EPROM are 256 bytes of RAM which are used as a buffer space and processor stack space.

One 8-bit latched output port interfaces to the GE HUD computer along with one of the 6 bits of an input port. This scheme requires the microcomputer to load an ASCII character into the output port connected to the GE computer. When the GE computer reads the 8 bits from its input lines, it processes the signal for presentation on the HUD. When the character appears on the display, a pulse is sent back to the microcomputer indicating that the character is valid. The two most significant bits of the 8-bit code sent to the GE computer contain information for locating the character on the display.

Located on an adjacent printed circuit card in the same rack is a memory card (3). It contains 1024 bytes of RAM and 16380 bytes of EPROM. The RAM is used to store data and to provide some buffer space while the EPROM contains extensions of the run-time routines. Also located in the same bank of memory chips are the parameters for the speech synthesizer. The letters of the alphabet require 9100 bytes of memory.

The next printed circuit card is the speech synthesizer from Computalker Corporation (4). It contains all the circuitry for speech synthesis. It has one output to be connected to the aircraft intercom system.

An interface board connection (5) to the MFE Digital tape transport (6) is provided for reading experimental parameters and writing data files. Also located on the cassette interface board is a real-time clock that provides a highly accurate 1 kHz time base for measuring subject reaction times.

The response switches, located on the aircraft control stick, are interfaced to 2 bits of a parallel input port. One response switch is the stick trigger switch and the other is the stick thumbswitch. The rear cockpit controls consist of a master reset switch and a toggle switch for starting the task, aborting it, and restarting the same task or ending data collection, with all data being saved. Similar switches are provided with the alternate ground-based laboratory version of the SRL airborne box.

To initialize a digital tape for the airborne controller and to reduce the collected data, a ground-based data analysis center is provided. The device contains a printer circuit CPU card (2) identical with the one used in the airborne controller, making them interchangeable. Two other cards contain 8K each of RAM (13). The third printed circuit card holds 16K of EPROM (14). A fourth card contains the Basic Interpreter in 16K of RAM (15). The fifth card has a cassette interface identical to the one in the airborne controller minus the real-time clock (5). The printed circuit cards are contained in a standard package with the mother board back plane (16), two power supplies (9), and the digital magnetic tape recorder (6). A standard video terminal (11) is interfaced to the package along with a printer (10).

This unit also uses the S-100 bus making it compatible with the airborne controller and many other off-the-shelf peripherals.

SOFTWARE DESCRIPTION

The software for the airborne controller was written in Intel 8080 assembly language. The basic scheme was to establish certain options and then loop through the task repeating each pass with new control parameters. Each time a cassette tape is initialized from the ground-based data analysis unit, the experimenter is required to enter some optional parameters. These options are then used by the airborne controller for the experimental run.

1. The inter-stimulus interval (ISI) can be fixed or random. The experimenter is required to enter the fixed interval during the tape initialization sequence prior to any experimental run. A random sequence of ISI's ranging from 5 to 30 sec can also be entered at the time of tape initialization.
2. The number of probe items for the positive and negative sets combined can be any combination up to a total of 100 letters for any single m-set presentation. The experimenter is required to enter a random sequence of probe items during tape initialization.
3. The maximum number of positive letters in any m-set cannot exceed four (positive probe items = 1, 2, 3, or 4). Each trial consists of presenting one sequence of positive and negative letters and recording the results on digital tape. Any number of trials can be presented in an experimental session but each trial must be initialized and a header file placed on the digital tape prior to the session.

During a data trial, the GE HUD or the speech synthesizer will present the letter. The reaction time from stimulus onset will always be 2 sec less than the ISI. Responses occurring outside this window will be tagged for the maximum allowable time within range of the given ISI. When the task is not in use, a minus (-) sign will be presented on the HUD. The duration of any visual presentation on the HUD will be 2 sec less than the ISI or time of response occurrence.

In the rear seat of the NA-T33A is a three-position switch for the experimenter. The functions of this switch are described below.

1. When in the center (run) position, no task will run. A minus (-) sign will be displayed on the HUD during this time.
2. When in the up (set/end) position, the task is initialized and the next m-set in the sequence will run after the switch is returned to the center (run) position.
3. A momentary position (down) is provided so that any trial can be aborted. When this switch position is used, the aborted trial will be reloaded and used as the next trial in the sequence. The momentary position is deactivated between trials.
4. An end-of-trial position is also provided. Only after a trial has been started as per 1 and 2 above is the END feature activated. If the experimenter wishes to end a data collection trial and save the data, the toggle switch can be moved up to the END position. After a delay no longer than one letter presentation, a checkmark (✓) will appear on the HUD indicating that the data have been recorded. The experimenter must then move the switch down to its center (run) position and repeat the above sequence for the next trial.

When the data have been collected, the tape is returned to the ground-based analysis center. All the data analysis routines are written in BASIC because of its ability to handle the data analyses and format the output. Several routines are linked with assembly language subroutines capable of writing to the magnetic tape recorder.

The data analysis software options permit each of the following:

1. Printout of the raw reaction times including marking errors and no-response conditions.
2. Printout of the grouped raw positive reaction times and negative reaction times.
3. An option to specify what range of reaction time to use in the final analysis (i.e., 1000 milliseconds).
4. An option to specify what m-sets to use in the analysis.
5. A printout of the mean (X) and standard deviation (SD) of each m-set.
6. Plots of Xs versus m-sets by combining all Xs in a session.
7. A least squares regression line fit to the data giving slope and intercept of the equation.

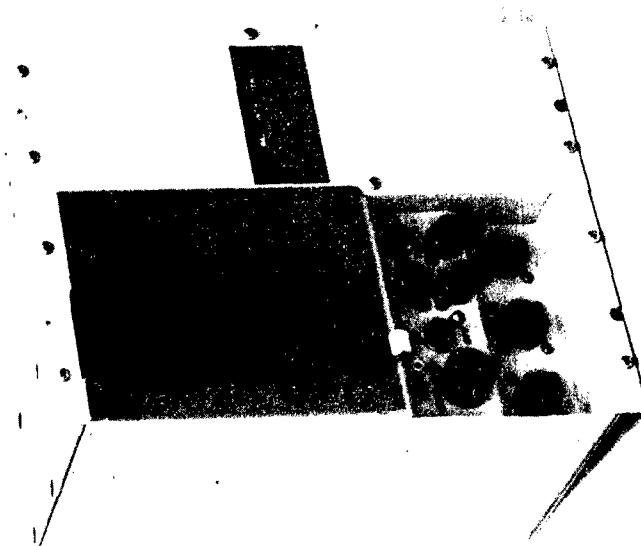
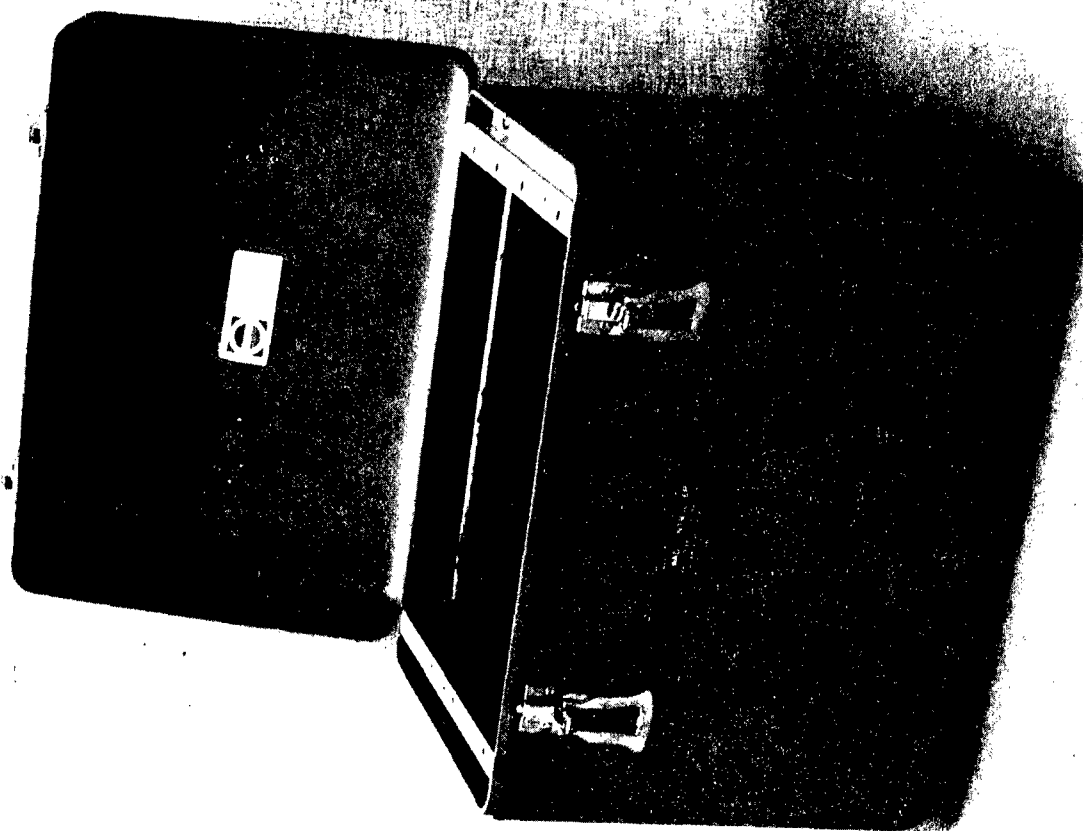


Figure 1
Airborne Controller and Carrying Case

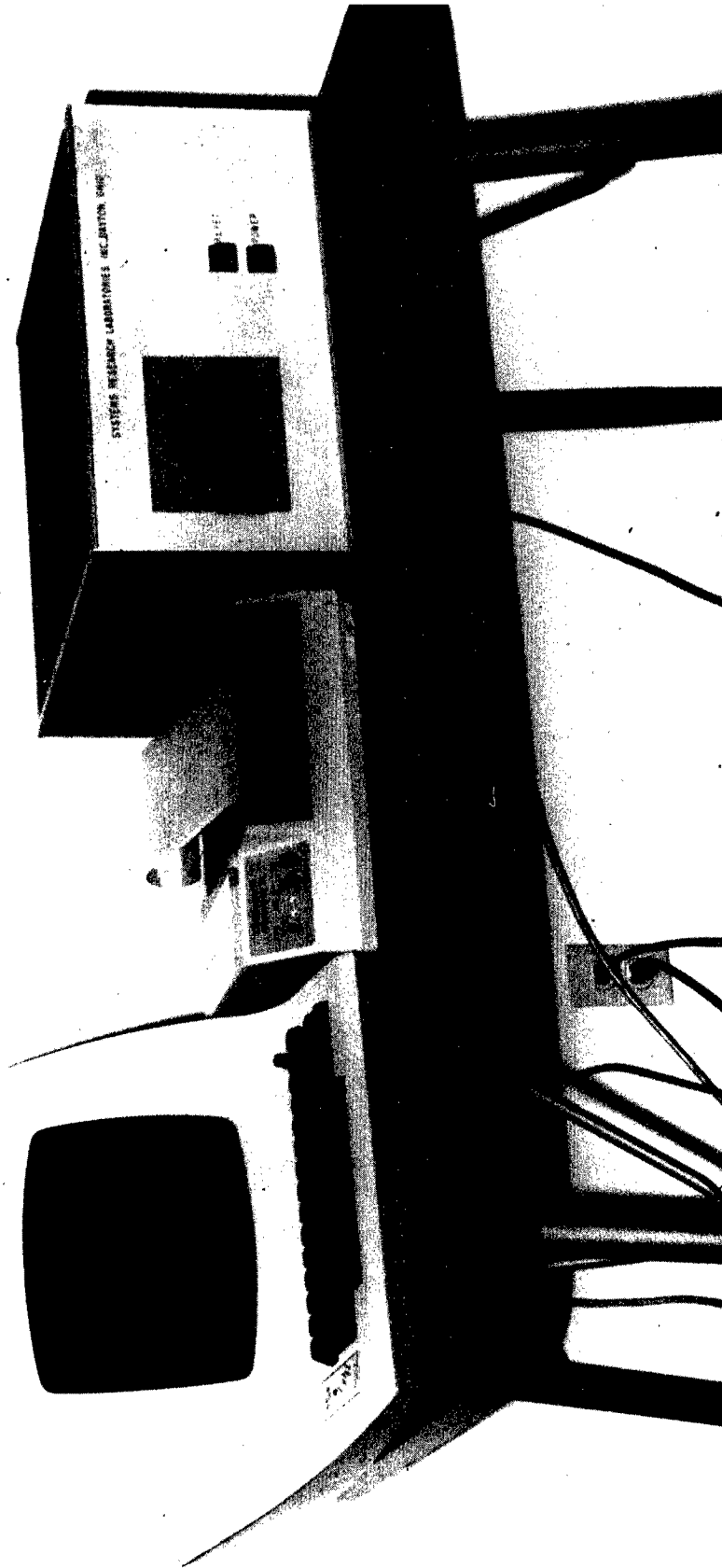


Figure 2
Ground-Based Data Analysis Center

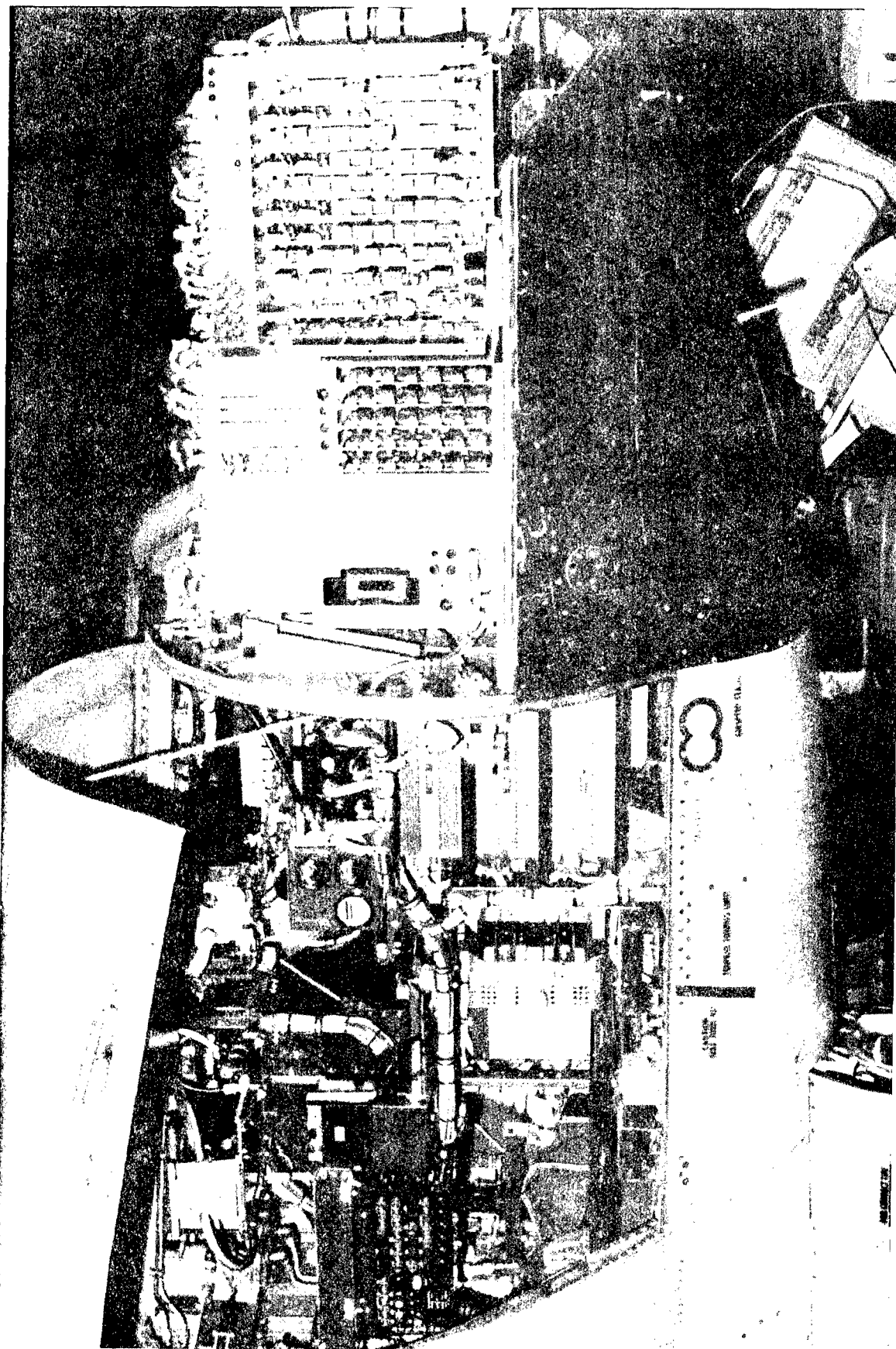


Figure 3
Airborne Controller Installed in NA-T33A Aircraft

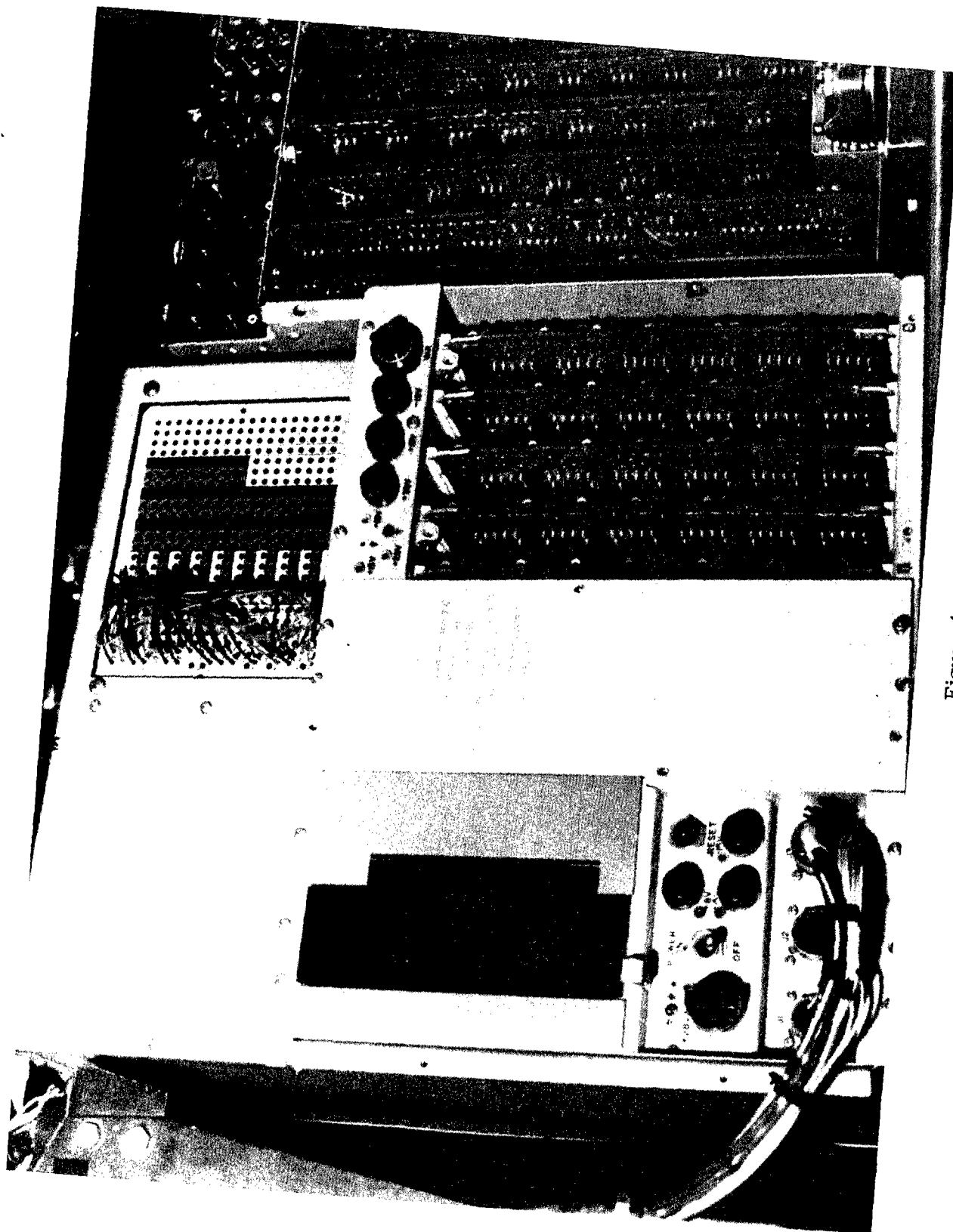


Figure 4
Close-Up View of Installed Controller

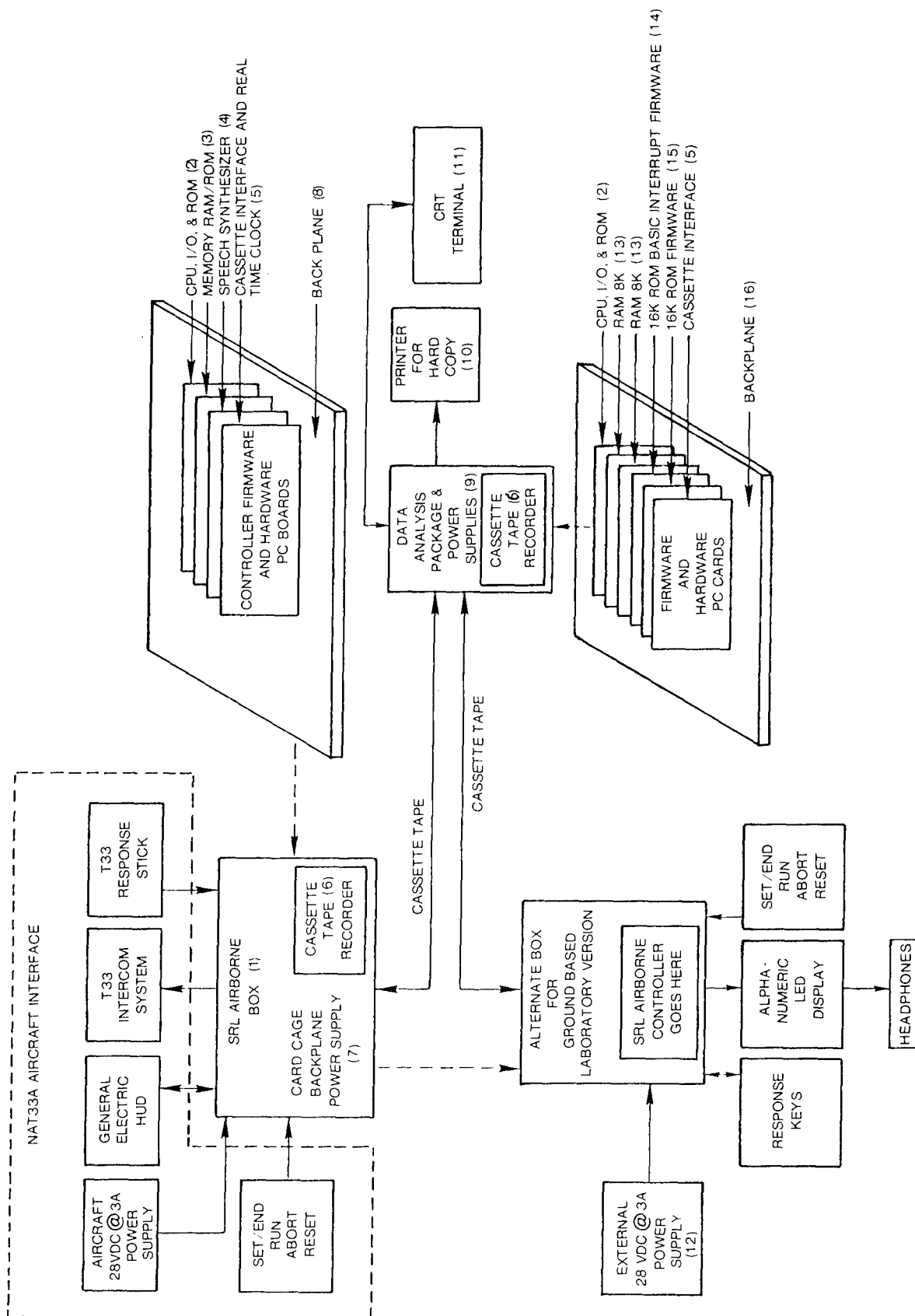
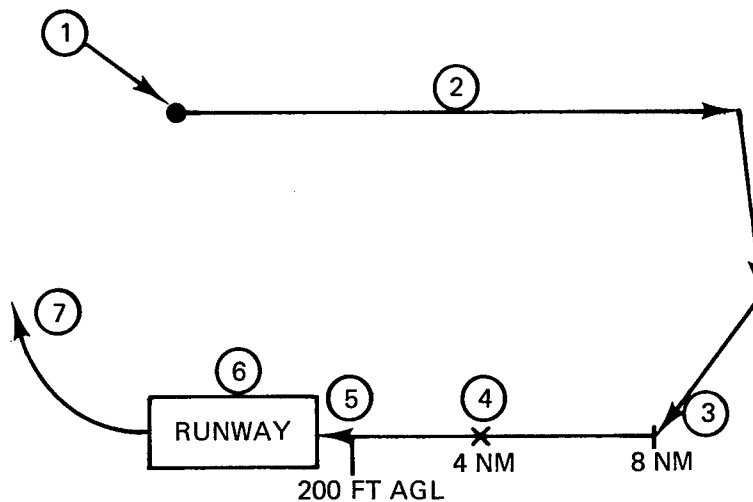


Figure 5
Block Diagram Showing Major Components in the SRL WAD 8085 Workload Assessment Device

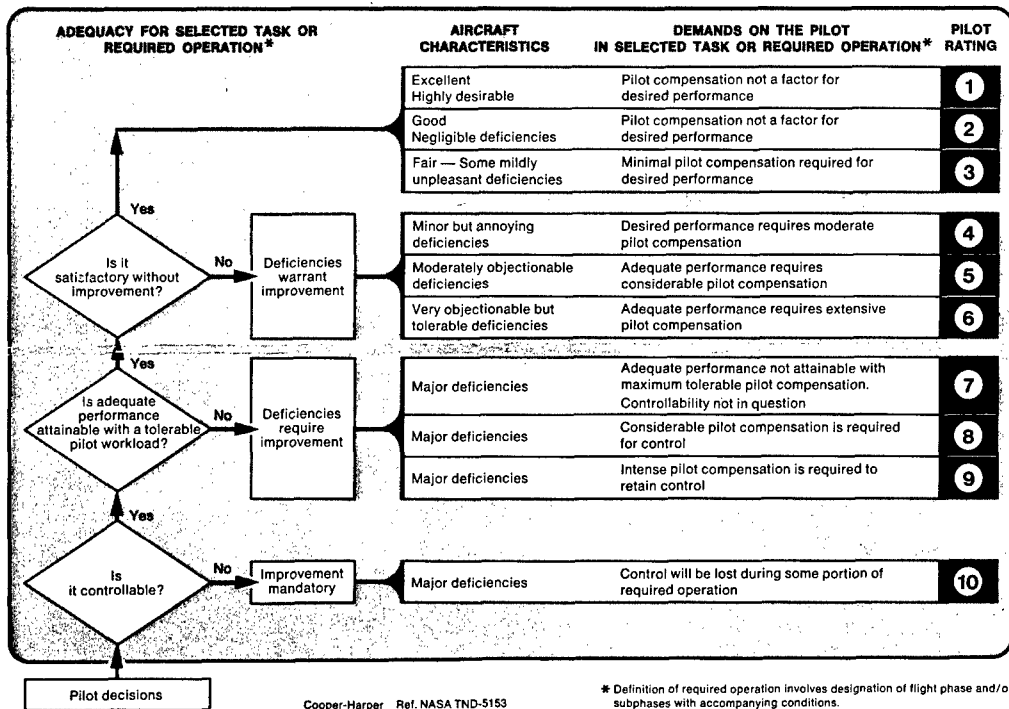
EVALUATION FLIGHT SCENARIO

<u>Task Event</u>	<u>Details</u>
(1)	Evaluation pilot (EP) is given control of aircraft with desired display/aircraft characteristics combination. When organized, he proceeds to join the downwind leg at 2,500 ft MSL.
(2)	EP goes on "instruments" using blue visor and follows safety pilot (SP) directions to intercept ILS beam.
(3)	EP descends to 1,800 ft MSL to intercept localizer. SP selects CAL records, Workload Assessment Device (WAD), and GE records (if desired).
(4)	EP flies glide slope and localizer to perform approach; outer marker is at approximately 4 mi.
(5)	At 200 ft AGL (or a suitable agreed upon height above that point), EP "breaks out" and flies visually for the rest of the task (lifts blue visor). Range is approximately 1/2 nmi from runway threshold.
(6)	EP performs touch and go landings, minimizing the sink rate on touchdown to less than 3 ft/sec, if below landing weight; otherwise, a low wave-off will be performed (from 20 ft AGL). Touchdown zone will be carefully defined for the task; essentially, it will be a 500 ft zone, 1,500 ft down the runway.
(7)	After liftoff and at a safe altitude, the SP will turn the recorders and the WAD off. The EP will fly to join downwind for another approach. After the fourth and last approach for each evaluation configuration, the SP will take control to allow the EP to make his comments and ratings.



HANDLING QUALITIES RATING SCALE

HANDLING QUALITIES RATING SCALE



DEFINITIONS FROM TN-D-5153

COMPENSATION

The measure of additional pilot effort and attention required to maintain a given level of performance in the face of deficient vehicle characteristics.

HANDLING QUALITIES

Those qualities or characteristics of an aircraft that govern the ease and precision with which a pilot is able to perform the tasks required in support of an aircraft role.

MISSION

The composite of pilot-vehicle functions that must be performed to fulfill operational requirements. May be specified for a role, complete flight, flight phase, or flight subphase.

WORKLOAD

The integrated physical and mental effort required to perform a specified piloting task.

PERFORMANCE

The precision of control with respect to aircraft movement that a pilot is able to achieve in performing a task. (Pilot-vehicle performance is a measure of handling performance. Pilot performance is a measure of the manner or efficiency with which a pilot moves the principal controls in performing a task.)

ROLE

The function or purpose that defines the primary use of an aircraft.

TASK

The actual work assigned a pilot to be performed in completion of or as representative of a designated flight segment.

PRIMARY MEASURES DATA ANALYSIS

Subtask	Record Portion	Investigation Area	Data Parameter and Statistics
ILS Approach	100 Sec Ending at 300 ft AGL	Approach Performance	Standard Deviation of Glide Path Error (GPE σ)
			Standard Deviation of Localizer Error (LOCE σ)
			Standard Deviation of True Angle of Attack ($\alpha_T \sigma$)
			Standard Deviation of True Track ($\psi_{TR} \sigma$)
		Approach Controllability	Standard Deviation of Longitudinal Stick Force ($F_{es} \sigma$)
			Standard Deviation of Pitch Angle ($\theta \sigma$)
			Standard Deviation of Bank Angle ($\phi \sigma$)
	2 Sec Ending at 300 ft AGL	Decision Height Performance	Standard Deviation of Throttle Displacement ($\delta_{TH} \sigma$)
			Mean of Glide Path Error (GPE)
			Mean of Localizer Error (LOCE)
			Mean of Track Error ($\Delta \psi_{TR}$)
			Mean of True Angle of Attack (α_T)

Subtask	Record Portion	Investigation Area	Data Parameter and Statistics
Flare and Landing	10 Sec Ending at Touchdown	Flare Controllability	Mean Pitch Angle (θ) Mean of Bank Angle ϕ F_{es} $\theta\sigma$ $\phi\sigma$
	1 Sec Ending at Touchdown	Touchdown Performance	Mean of Indicated Airspeed (V_i)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SY-33R-80	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) EVALUATION OF A PILOT WORKLOAD ASSESSMENT DEVICE TO TEST ALTERNATE DISPLAY FORMATS AND CONTROL HANDLING QUALITIES		5. TYPE OF REPORT & PERIOD COVERED FINAL JUNE 1978 - JANUARY 1980
7. AUTHOR(s) DR. S. G. SCHIFLETT		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS NAVAL AIR TEST CENTER NAVAL AIR STATION PATUXENT RIVER, MARYLAND 20670		8. CONTRACT OR GRANT NUMBER(s)
11. CONTROLLING OFFICE NAME AND ADDRESS NAVAL AIR SYSTEMS COMMAND (340F) NAVAL AIR SYSTEMS COMMAND HEADQUARTERS WASHINGTON, D.C. 20361		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N62269/78/WR/00408 N62269/79/WR/00462
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) NAVAL AIR DEVELOPMENT CENTER (6021) WARMINSTER, PENNSYLVANIA 18974		12. REPORT DATE 7 JULY 1980
		13. NUMBER OF PAGES 42
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
WORKLOAD PILOT WORKLOAD MENTAL WORKLOAD INFORMATION PROCESSING ASSESSMENT	TEST EVALUATION SECONDARY TASK DISPLAYS CONTROLS	DEVICE TECHNIQUE
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This in-flight research project evaluated the utility of a Workload Assessment Device (WAD) to measure pilot workload for approach and landing tasks under simulated instrument meteorological conditions, alternate HUD formats and control stability variations. The flight tests were conducted in an NT-33A research aircraft, extensively modified for the Air Force and Navy by the Display Evaluation Flight Test program. The hardware, software, and test procedures associated with the WAD functioned efficiently with only minor discrepancies and minimum pilot distraction. The project established the feasibility of using an item-recognition task as a measure of sensory-response loading and		

20.

reserve information processing capacity while flying precision approaches. In a descriptive statistical treatment of the data, the results indicate an appreciable increase in reaction time and errors with degraded handling qualities as compared to ground baseline measures and good handling qualities. The preliminary findings also reveal consistent trends toward the availability of more mental reserve capacity when flying predominantly pictorial/symbolic HUD configurations as compared to conventional HUD formats with scales and alphanumerics. It is recommended that further evaluations be conducted to establish the efficacy of utilizing the WAD to measure mental workload in a wide variety of aircrew tasks.

Naval Air Test Center, Patuxent River, Maryland

EVALUATION OF A PILOT WORKLOAD ASSESSMENT
DEVICE TO TEST ALTERNATE DISPLAY FORMATS
AND CONTROL HANDLING QUALITIES; FINAL
REPORT

7 July 1980
SY-33R-80
N62269/78/WR/00408
N62269/79/WR/00462

Approved for Public Release; Distribution Unlimited.

1. Workload
2. Pilot Workload
3. Mental Workload
4. Information Processing
5. Assessment
6. Test
7. Evaluation
8. Secondary Task
9. Displays
10. Controls
11. Device
12. Technique
13. Schiflett, S. G., Dr.

UNCLASSIFIED

Naval Air Test Center, Patuxent River, Maryland

EVALUATION OF A PILOT WORKLOAD ASSESSMENT
DEVICE TO TEST ALTERNATE DISPLAY FORMATS
AND CONTROL HANDLING QUALITIES; FINAL
REPORT

7 July 1980
SY-33R-80
N62269/78/WR/00408
N62269/79/WR/00462

Approved for Public Release; Distribution Unlimited.

1. Workload
2. Pilot Workload
3. Mental Workload
4. Information Processing
5. Assessment
6. Test
7. Evaluation
8. Secondary Task
9. Displays
10. Controls
11. Device
12. Technique
13. Schiflett, S. G., Dr.

UNCLASSIFIED

Naval Air Test Center, Patuxent River, Maryland

EVALUATION OF A PILOT WORKLOAD ASSESSMENT
DEVICE TO TEST ALTERNATE DISPLAY FORMATS
AND CONTROL HANDLING QUALITIES; FINAL
REPORT

7 July 1980
SY-33R-80
N62269/78/WR/00408
N62269/79/WR/00462

Approved for Public Release; Distribution Unlimited.

1. Workload
2. Pilot Workload
3. Mental Workload
4. Information Processing
5. Assessment
6. Test
7. Evaluation
8. Secondary Task
9. Displays
10. Controls
11. Device
12. Technique
13. Schiflett, S. G., Dr.

UNCLASSIFIED

Naval Air Test Center, Patuxent River, Maryland

EVALUATION OF A PILOT WORKLOAD ASSESSMENT
DEVICE TO TEST ALTERNATE DISPLAY FORMATS
AND CONTROL HANDLING QUALITIES; FINAL
REPORT

7 July 1980
SY-33R-80
N62269/78/WR/00408
N62269/79/WR/00462

Approved for Public Release; Distribution Unlimited.

1. Workload
2. Pilot Workload
3. Mental Workload
4. Information Processing
5. Assessment
6. Test
7. Evaluation
8. Secondary Task
9. Displays
10. Controls
11. Device
12. Technique
13. Schiflett, S. G., Dr.

UNCLASSIFIED