DOT/FAA/EM-81/14
DOT/FAA/CT-82/23

The Measurement of Pilot Workload

Earl S. Stein
Bruce L. Rosenberg

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

January 1983

Interim Report

U.S Department of Transportation
Federal Aviation Administration
Office of Systems Engineering Management
Washington, D.C. 20590

NOTICE

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof.

The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the object of this report.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/EM-81/14		

4. Title and Subtitle	5. Report Date
THE MEASUREMENT OF PILOT WORKLOAD	January 1983
	6. Performing Organization Code
	8. Performing Organization Report No.
7. Author(s)	
Earl S. Stein and Bruce L. Rosenberg	DOT/FAA/CT-82/23
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Federal Aviation Administration Technical Center Atlantic City Airport, New Jersey 08405	11. Contract or Grant No. 161-301-150
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Aviation Administration Office of Systems Engineering Management Washington, DC 20590	Interim Report
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The evaluation of pilot workload has represented a complex measurement problem since the earliest days of manned flight. Traditional approaches have included a series of techniques, many of which have not proved successful. The most popular method has employed post-flight questionnaires or interviews.

This current experiment was an attempt to measure workload during flight simulation, using two primary variables: the pilots' own evaluation sampled once per minute with a computer and the latency or delay of that response. This was supplemented by a post-flight questionnaire. Three levels of flight difficulty were established by subject matter experts. These were varied by controlling (1) initial clearance complexity, (2) level of air traffic control, (3) turbulence, and (4) inflight emergency. "Flights" were conducted in a General Aviation Instrument trainer and 12 pilots participated.

Results demonstrated that pilots were willing and able to make inflight workload evaluations which corresponded directly with the induced difficulty level. Response latencies increased in relationship to difficulty, but the intermediate and most difficult flights were not significantly different. Factor analyses of all measures produced two clusters for the easiest and intermediate flights (inflight and postflight) and four for the most difficult flight. In the latter case, inflight and postflight measures separated into two factors and the questionnaire split also into two segments. These separations indicated that within the current state of the art, both types of measures should continue to be collected.

Plans call for follow-on research in General Aviation Workload.

17. Key Words		18. Distribution Statement
Workload	Task Load	Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161
Pilot Workload	Task Analysis	
Cockpit Workload	Workload Rating	
Workload Measurement	Task Difficulty	
Inflight Workload		

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	55	

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

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

APPENDICES Page

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

PURPOSE.

This report represents the results of a continuing series of experiments designed to explore a workload assessment technique. The technique is intended for use in evaluating the potential impact associated with changes in cockpit procedures and instrumentation. The technique would serve in the role of an appropriate workload measurement method required to provide a common basis for assessing the results of future cockpit-oriented experiments. The purpose was to determine if a relationship exists between a 10-point subjective response scale and a predetermined objective level of workload. The test was conducted to determine if pilots could differentiate between three levels of difficulty associated with an inflight simulation. The research to be described in this report was accomplished as a portion of the Federal Aviation Administration's (FAA) program concerned with applied human factors, as part of a joint NASA/FAA research program. This is the second study in a series focusing on a preliminary evaluation of the acceptability and utility of inflight workload measures.

BACKGROUND.

Flying a modern aircraft involves a complex, multidimensional series of behaviors, only some of which can be observed directly. Pilots must communicate, navigate, control, and monitor (Sheridan & Simpson, 1979). They must accomplish some tasks simultaneously and set priorities on others to be accomplished in sequence. Since the advent of concern for man-machine relationships, investigators have been trying to evaluate how well equipment designs meet the capabilities, needs, and limitations of human operators.

Equipment is becoming increasingly reliable, and the weak link in man-machine systems is often the human operator, whose reliability can be a function of the load placed on him/her (Roscoe, 1978). In aviation, the relationship between flight safety and workload could have serious consequences. Unfortunately, the definition and specification of what is meant by the workload construct is not a simple task.

Researchers are coming to the realization that workload is multidimensional and that no simple definition may be universally acceptable (Eggemeir, 1980; Chiles, 1979). It may be not only reasonable, but also desirable, to tailor the definition to the research situation. Johannsen (1977) indicated that there are essentially three reference points in any man-machine environment. These include (1) the inputs to the operator, (2) how he/she processes this information, and (3) what he/she does with it in terms of performance. Some sort of measurement is possible at each of these reference points.

Development of workload measurement techniques has been complicated by the diversity of workload definitions and underlying theoretical formulations. Williges and Wierwille (1979) described 28 separate methods of measuring workload which have been cited in the research literature. These can be categorized under three generic headings: Performance, Physiological, and Psychological measures. Performance is the most directly measurable, but has had its problems as a workload assessment technique. Primary task performance refers to what the individual is supposed to be doing, such as flying an aircraft. Secondary tasks are those which are unrelated to the primary effort and are designed to load the individual's spare capacity. In theory, the operator's secondary task performance will decline as he/she becomes more loaded on the primary task. Physiological measures have included heart rate, galvanic skin response, sinus arrythmia, and blood chemistry among others. Both of these sets of measures are somewhat intrusive on the operator, and results have been varied across studies. The final category involves the psychological measures which focus on the evaluation of operator subjective responses.

Roscoe (1978) noted that it would be desirable if human beings could be measured with the same precision as mechanical or electrical systems. Unfortunately, that is not within the current state-of-the-art. While psychological measurement is concerned with performance and physiology, a great deal of emphasis has been placed, especially in aviation, on workload assessment using subjective, self-report data. This popularity has arisen, in part, from the relative ease of administration and low cost of these techniques. Cooper and Harper (1969) developed a scale to assess aircraft handling qualities which stimulated a great deal of interest. While designed to guage handling qualities and not directly applicable to workload measurement, its form and substance have been employed in several workload studies (Sheridan and Simpson, 1979; Katz, 1980). Self-report techniques are generally not intrusive when applied in a postflight context. However, they must depend on pilot/operator memory, which may be prone to some error as a function of leveling, sharpening, and assimilation.

The use of subjective measurement during task performance was tested at the FAA Technical Center using a nonflying critical tracking task. Rosenberg, Rehmann, and Stein (1982) asked 12 pilots and 12 nonpilots to hold a point of light centered in a CRT display using a joystick control. Every minute, the subjects responded to a query tone by pressing one button in a series numbered from one to ten. Each response was an assessment of their workload from very easy (1) to very hard (10). A measurement of response delay for each query tone was also recorded. Results indicated that participant responses were directly related to four objectively controlled difficulty levels. The tracking task experiment was the first study in the workload series at the Technical Center and served as a stimulus for the research described in this report.

This current research is based on the assumption that workload is multidimensional in character. It includes both overt and covert dimensions. Those which are not directly observable must be inferred. An individual's assessment of how hard he/she is working at any point in time is assumed to be directly related to some idealized, ultimate indicator of workload which has yet to be clearly defined (if it in fact exists) and precisely measured. The individual's response to such a question will no doubt encompass both overt physical effort and internal events, which have been referred to as information processing, planning, problem solving, decision-making, stress response, etc. How hard an individual believes he/she is working may turn out to be as relevant as the idealized measure when it is used to assess the influence of workload on the acceptability and utility of new systems.

A description of the method employed in this current experiment follows in the next section.

METHOD

RESEARCH DESIGN.

Since the basic objective of this study was to determine whether or not the workload measurement system was sensitive to differing levels of workload, a definition of what input variables would induce workload was necessary. Based on the advice of two subject matter experts (both high flight-time pilots), the following variables were selected: (1) level of air turbulence/wind, (2) initial clearance complexity, (3) frequency of air traffic control inputs, and (4) an inflight emergency. The reasoning of the research design focused on the development of an independent variable which could be called flight difficulty. This was to represent three distinct qualitative, rather than quantitative, levels which could potentially induce three levels of workload. "Difficulty" should not be confused with "workload" because it refers to the input stressors placed on the participants. "Workload" is used here to describe the pilots responses to these environmentally induced conditions. The independent variable was organized into three flights which were designed to be of increasing order of difficulty. These have been labeled flights A, B, and C, respectively. Each flight was flown on the same geometry from Millville to Atlantic City, New Jersey. Table 1 describes the basic contents of these flights, while a more detailed description of what occurred is available in appendix B under the labels of scenario guide and scenarios 1, 2, and 3.

TABLE 1. FLIGHT DIFFICULTY CONTENTS

Flight/Stressors	Turbulence	Wind	Clearance	ATC Freq.	Inflight Emerg.
A	None	None	Simple	Low	No
B	Low	Low	Complex	Moderate	No
C	Moderate	Low	Very Complex	High	Yes

PARTICIPANTS.

Twelve pilots completed this experiment, and their data were recorded for analysis. This does not include two pilots, who exercised their voluntary consent rights and terminated during the training phase of the experiment. It also does not include two other pilots, who should have been screened out for lack of familiarization with the aircraft configuration as flown. These two individuals completed the experiment, but their data was deleted. All participant pilots were volunteers. Criteria for participation included previous experience with multiengine instrument flying and local availability. A check ride/training flight was completed prior to any testing, and this lead to the two voluntary withdrawals because of admitted lack of familiarity with the equipment. The 12-pilot sample included personnel who ranged in total flight hours from 1,600 to 14,500 with a median of 4,000 hours. The range of their instrument time was from 175 to 2,500 with a median of 450 hours. Finally, they were asked how much they had flown in the past year as an indicator of currency. The range was from 10 to 375 with a median of 200 hours. Flight times of each participant are listed in appendix A. All participants were either FAA employees or members of the local Air Guard Unit.

Participants were carefully briefed on the experimental requirements prior to testing. Debriefing after the experiment was made available to everyone, but not all requested it. The pilots were asked not to discuss the experiment with their co-workers, who might yet participate.

While the decisions involved in developing the three flight scenarios were arbitrary, it was believed that they would induce a spectrum of workload from low to very high, which could effectively exercise the measurement system.

The experimental design was of the repeated measures type in which all participants were exposed to all conditions. After the training/check ride, each pilot had to fly all three test flights. A counter-balanced design was developed which included the six administration orders, as indicated in table 2.

TABLE 2. ADMINISTRATION ORDER

Pilot Numbers	Flight Order
08, 14	A B C
02, 10	C A B
03, 11	B C A
04, 12	A C B
05, 13	B A C
06, 09	C B A

Two pilots were assigned to each other. The purpose of counterbalancing was to control the potential effects of experience and fatigue.

Dependent variables (measures taken to determine the influence of the three levels of difficulty) included (1) inflight workload responses (subjective rating), (2) response delay on workload responses, (3) observer inflight task analysis, and (4) postflight workload/stress questionnaire.

EQUIPMENT.

The basic unit of equipment, upon which the entire experiment focused, was the Singer-Link General Aviation Trainer (GAT II). The FAA Technical Center GAT replicates the appearance and simulates the performance of a Cessna 421, a cabin class reciprocating twin-engine aircraft. It permits instrument flying only and has no visual display system. It is mounted on a motion platform having two degrees of freedom and is able to provide vestibular and kinesthetic pilot cueing for pitch, roll, and, to a certain extent, elevation changes. The cockpit is equipped with (1) Collins FD 109 Flight Director, (2) AP 106 Auto Pilot, (3) twin NAVCOMS, (4) transponder, (5) ADF, and (6) other standard instrumentation.

The GAT was equipped with one special feature that was not related to its flight performance. This was a workload response box which was mounted just below the throttles outside the pilot's primary visual scan. It contained 10 pushbutton switches placed in a semicircular array and a tone alert speaker. At the center of the switch array was a red light emitting diode, which was turned on each time there was a query tone requesting a workload response. This light was to remain on until the participant pushed any button. The use of this box will be explained in more detail in the procedures section of this report.

This hardware was driven by, and provided inputs to, several computer systems. An analog/digital system computed the equations of motion, controlled the motion platform, and drove some of the aerodynamic information displays. Guidance processing was accomplished with a NAV Systems Simulation Package or NSSP.

Finally, a DEC LSI-II computer was used to serve multiple roles. It provided flight track plotting, and stored the pilots' workload responses and their response delays. These delays were computed using an internal clock. These data were available in printout form at the end of each flight. This computer also provided query tones every minute to the pilot, which were used to request his/her workload responses.

The final element of equipment in this experiment was the instructor's console. This was located in a separate room from the simulator and served as the work station for the air traffic controller. This console has a repeater panel, which provided a portion of the same information that the pilot had available. It provided control over the atmospheric environment of the simulated flight and over aircraft systems operations. This device permitted simulated flight problems and failures to be induced. Communication with the cockpit could be used to provide ATC influence. A plotter, located as part of the console, was not employed for this experiment. A more accurate Hewlett-Packard plotter, linked to the LSI-11, was used in its place and made available a real-time track of each flight, which the controller could see from his position at the console.

PROCEDURE.

During the training/check ride phase, the entire effort was devoted to familiarization with the Cessna 421 as configured in the GAT. No briefing was provided on the research itself until the test phase of the experiment, when the three key flights were flown. This training was done by three separate individuals because none of them were available for the entire experiment. While training "flights" were accomplished in the vicinity of Atlantic City, none were flown on the same route as employed in the three test flights. The experimenter asked the trainers to orient pilots on the equipment and to determine whether the pilot was adequately proficient to participate in the study. As indicated earlier, two individuals slipped through this screening, and this may have been a function of differing standards across the three trainers. The briefing employed by the trainer who worked with the majority of the pilots is listed in appendix C, as the "maneuvers briefing." The familiarization period for all the pilots lasted approximately 1 hour and 30 minutes. Each was then randomly assigned to one of six administration orders.

At this point, the air traffic controller briefly described to the pilot the route he/she would be flying from Millville to Atlantic City. The pilot was issued an 8- by 11-inch locally drawn air route map and a note pad with which to copy clearances and changes to clearances. The locally produced map was required because it was easier to produce than to alter flight geometry in the GAT computer, which differed slightly from standard. With the pilot in the left seat of the cockpit, the experimenter sat in the right seat and briefed the pilot on the research and his/her tasks regarding the workload response box. The pilot was instructed to respond as quickly as possible every minute to the workload query tone and was told to rate his/her workload from a low of 1 to a high of 10. Pilots were reminded of their rights to privacy and anonymity. Detailed instructions, as read to the pilot, are presented in appendix D-1. At the conclusion of these instructions, the experimenter informed the pilot that he/she could call Millville Flight Service for a clearance and proceed with his/her flight.

The air traffic controller preset wind and air turbulence into the instructor's console and provided a clearance, as indicated in scenarios 1, 2, and 3. In flight C, the controller pushed a button at a predetermined point (just after the aircraft came out of an instructed holding pattern) which caused the right engine to fail. While the majority of interchanges between ATC and the pilots were based on the scenarios, the controller retained the flexibility to respond to pilot questions and unforeseen circumstances.

During each flight, the experimenter performed a task analysis which amounted to a frequency tally of overt pilot behavior. Four categories of behavior were tallied: Control, Navigation, Communication, and Nontask Appropriate. The latter category referred to movements (i.e., head scratching) and verbalizations that had nothing to do with flying the aircraft. The experimenter informed the pilot before the first test flight that the task analysis was not an evaluation and was designed to determine pilot activity level.

After each "flight," the experimenter immediately administered the flight workload questionnaire. This included four scales: Workload, Busyness, Thinking, and Feeling. (See appendix D-2 for actual questions.) When the questionnaire was completed, the GAT was reset to Millville via the magic of the computer, and the pilot was informed to call for a new clearance. Each pilot completed his/her three flights in about 1 hour and 30 minutes, which included administrative time. Since piloting style differs from one individual to the next, flight times varied in kind. Median flight times and ranges are presented in table 3. The next section of this report will describe the results of this experiment.

TABLE 3. MEDIAN FLIGHT COMPLETION TIMES (MIN.)

Flight	Median	Range	
		Low	High
A	14.5	12	16
B	16.0	13	22
C	22.0	21	27

RESULTS

This was a preliminary experiment and any conclusions which are made should not exceed the level of precision of the sampling and data collection procedures. Participants in this experiment were local volunteers and therefore, may or may not represent General Aviation at large. Four types of data were collected during the experiment. These included the task frequency tallies, the inflight workload responses and delays, and the postflight questionnaire. Results will be reported in this section as data summaries and statistical analyses. Discussion of these results will be deferred for the most part to the next section. A results summary is presented below for the benefit of readers not technically interested in statistical analyses. (Those readers may then wish to skip to the conclusion section.)

RESULTS SUMMARY.

A task frequency tally of pilot behavior showed an increase in pilot activity across flights A, B, and C in three behavioral categories: (1) navigation, (2) communication, and (3) nontask appropriate behavior. Inflight task tallies documented differences in activity level in two categories, navigation and communication. For navigation, these differences existed between flight A and the first part of flight C (C_1) prior to the inflight emergency (C_2). A difference also existed in pilot activity level between C_1 and C_2, with C_2 being considerably less active (see table 5). In the communication category, the only significant difference was between flights A and C_1, with C_1 being the more active.

Analyses of inflight workload responses and delays showed that there were significant differences across flights for both variables. Using the workload response variable, all flights were significantly different from each other, but segments C_1 and C_2 were not. The order of the mean workload responses A--B--C, was directly in line with the hypothesis that pilots could accurately separate their workload evaluations over three flights of increasing difficulty. The results for the delay variable indicated significant separation of flights A from B, and A from both C_1 and C_2. Flights C_1 and C_2 were also not significantly different.

Postflight questionnaires containing four response items showed some significant differences for each question. The workload question was the least useful in separating the three flights, with significant differences between flights A and C only. The other three questions, related to busyness, thinking, and feeling, were significantly different between all pairs of flights. The more difficult the flight was, the higher was the mean numerical rating which pilots assigned.

Factor analyses indicated that for flights A and B, there were essentially two clusters of variables--one composed of the two inflight measures (workload and delay) and the other composed of the four postflight questions. In flight C, however, four factors appeared. Workload and delay loaded on separate factors and the busyness postflight question broke away from the remainder of the postflight questionnaire.

Predictions of postflight questionnaire responses, using inflight data, were only moderately successful, confirming a basic difference between the nature of the two sets of measures.

TASK ANALYSIS

The purpose of the task frequency tally was to obtain a coarse measure of pilot activity level across the three flight difficulties. Only one observer was employed (a psychologist who is a nonpilot) and no attempt was made to assess measurement reliability. The frequency tally procedure was intended only as an indicator of pilot activity rather than as a major measurement system. This was a preliminary experiment and follow-up experiments will pursue the question of observer reliability. The observer's purpose was to tally pilot activities accurately and verify that the data are in general agreement with the assumption that three levels of flight difficulty were generated.

Prior to becoming "airborne," three categories of pilot behavior were tallied. These included navigation, communication, and nontask appropriate behaviors.

Figure 1 presents the total number of tasks performed divided by the number of pilots. Because no clock was run during the preflight preparation, these are average totals per pilot that do not take into account the preparation time. The total amount of effort prior to becoming airborne was of more concern than how long it took. There is an apparent increase across the three flight levels A, B, and C in the total number of preparation tasks accomplished. This was directly in line with the level of clearance complexity generated by the experimental design.

At the point the wheels left the ground, time was measured in minute increments by the computer. This meant that an average frequency of tasks per minute per pilot could be computed. These are presented in figure 2.

The reader will note that flight C has been separated into two segments--before and during the inflight emergency. The pattern of change in the four categories reported in figure 2 was not as clear as it appeared in the preflight data. Part of this may be due to the introduction of the time variable.

FIGURE 1

PREFLIGHT TASK SUMMARY

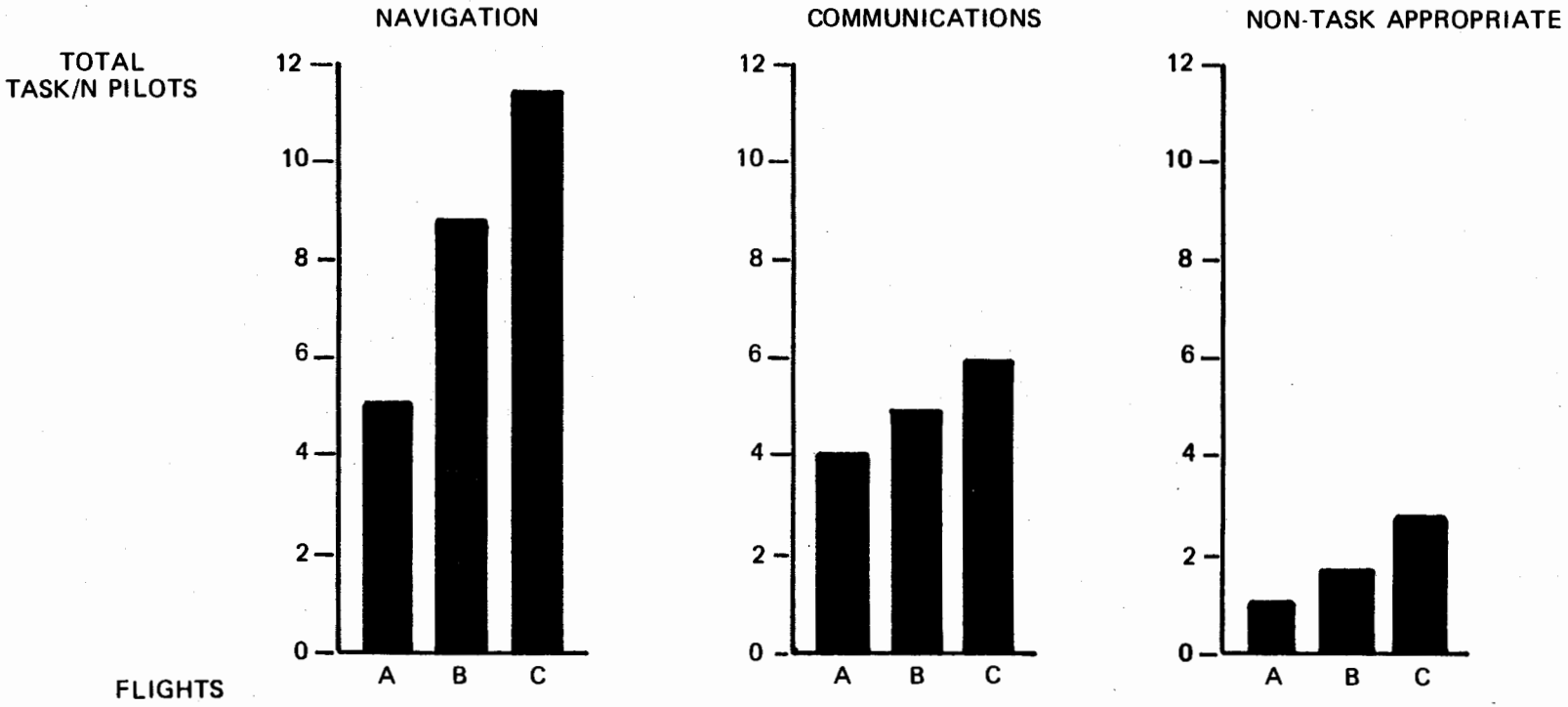
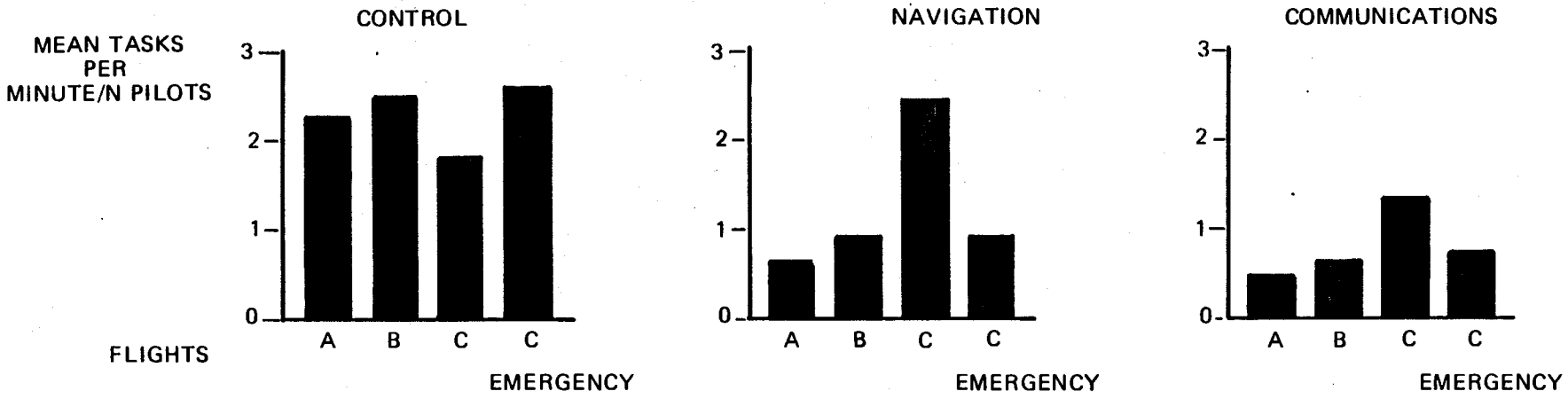
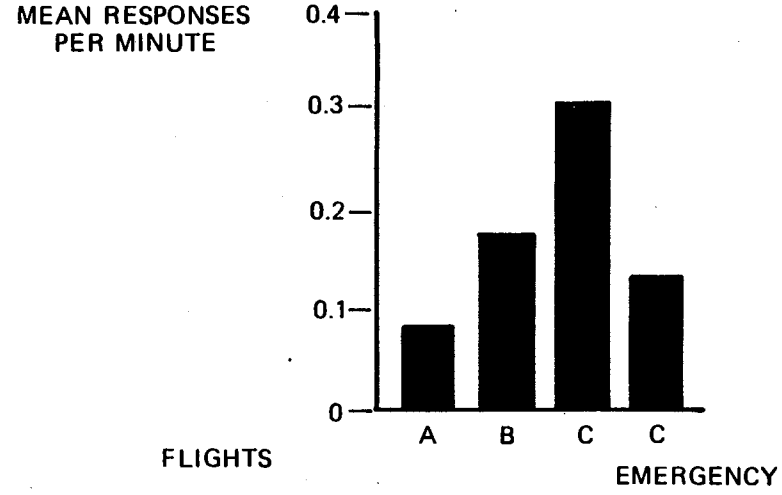


FIGURE 2

INFLIGHT TASK SUMMARY



NON-TASK APPROPRIATE RESPONSE/MINUTE



Each category of the task tally was analyzed separately to determine whether the frequencies across the four flight levels (A, B, C_1, and C_2 - (emergency) differed by more than would be predicted by chance. Because the task tally was a nonstandardized measure, at best, and because the population parameters upon which it was based were unclear, a nonparametric analysis, Friedman's Test, was employed (Linton and Gallo, 1979). Results are shown in table 4.

TABLE 4. FRIEDMAN'S TEST - INFLIGHT TASK FREQUENCIES

Category	Chi Square
Control	6.98
Navigation	30.43**
Communication	201.88**
Nontask Appropriate	7.78

**Significant $P<.01$

Two categories demonstrated variability across the four flight levels. The nontask appropriate category approached significance at the 0.05 level but did not quite make it. The bar graph in figure 2 may be somewhat misleading in the nontask appropriate category. It would appear that if communication and navigation were significant, the nontask appropriate category should also have been. However, examination of the Y axis will show that the mean responses per minute/pilot are much lower than that for the other two categories. There were, in fact, many ties in ranks across the four levels which brought the Chi Square down below the critical cutoff.

A significant Friedman's Test, like its parametric counterpart, the analysis of variance, indicates that somewhere between the levels of the variable, significant differences exist. However, it does not indicate where they are. This requires an additional analysis called Nemenyi's Test.

Friedman's Test required the analyst to rank the mean tasks per minute for each pilot across the four flight levels. So for example, if flight A was the least busy, then many 1's should be assigned, and if flight C_2 was most busy, it should receive many 4's. When these ranks are averaged across pilots, a mean sum of ranks is produced. Nemenyi's Test produces a critical difference between flight levels. Then, all the differences between the flight levels are computed and compared against the critical difference. Those which exceed it are said to differ by more than chance would predict.

Table 5 presents the computations for the navigation category. There are some apparent surprises. The cutoff to be exceeded was 1.775. Two significant differences occurred.

TABLE 5. NEMENYI'S TEST/NAVIGATION FREQUENCY

FLIGHT LEVEL		A	C_2	B	C_1
	MEAN SUM OF RANKS	1,208	1.916	2.958	3.916
A	1.208		.708	1.750	2.708**
C_2	1.916			1.042	2.000**
B	2.958				.958
C_1	3.916				

**Significant P<.01

C_1, or that portion of flight C prior to the emergency, had a significantly higher frequency of tasks/minute/pilot than did flight A. It also had a significantly higher frequency than C_2 (or the emergency portion of the flight). While this will be discussed in a latter section, it is important to note here that this finding was most likely an artifact of experimental design. Navigation tasks may have actually become simpler after the emergency during which the pilot had to make only one turn and then stay on the Instrument Landing System (ILS) localizer beam. The results for the communication category are presented in table 6.

TABLE 6. NEMENYI'S TEST/COMMUNICATION FREQUENCY

FLIGHT LEVEL		A	B	C_2	C_1
	MEAN SUM OF RANKS	1.416	2.208	2.542	3.833
A	1.416		.792	1.126	2.417**
B	2.208			.334	1.625
C_2	2.542				1.292
C_1	3.833				

**Significant P<.01

The critical cutoff here was again a difference of 1.775. Only one difference exceeded this level, and that was between flights C_1 and A. The difference between C_1 and C_2 approached significance, but did not quite reach it.

INFLIGHT WORKLOAD RESPONSE AND DELAY.

The primary purpose of this experiment was to determine whether pilots were able and willing to make workload responses during flight, and whether these responses and their delays corresponded in some manner with the difficulty levels produced by the experimental design. Because this was viewed as a preliminary experiment, flights were not segmented into components with the exception of flight C, in which a clear separation between two elements could be seen. Each flight was treated as an entity and the arithmetic means of pilot's workload responses and the delays of those responses were selected as the numbers to represent the entire flight and as the data points for further analysis.

The reader will recall that the pilot was asked every minute to respond to the query tone with an answer to the question of how hard he/she was working. The pilot pushed a button from one (very easy) to ten (very hard). The delay of response was measured via the computer. It was assumed, for the purpose of this experiment, that that response delay was related to workload, and if the pilot failed to make a response within 1 minute, this was also an indicator of high workload. When the pilot did not respond, the computer automatically recorded a response of ten and a delay of 60 seconds at which point the pilot was again queried. The mean frequency of missed responses across pilots is reported in table 7.

TABLE 7. MEAN MISSED WORKLOAD RESPONSES

	Flight			
	A	B	C_1	C_2
Mean	.75	3.42	3.25	1.67
Standard Deviation	.97	2.97	2.60	1.15

The mean workload response for each flight, with the inflight emergency treated as a separate flight, is described in table 8.

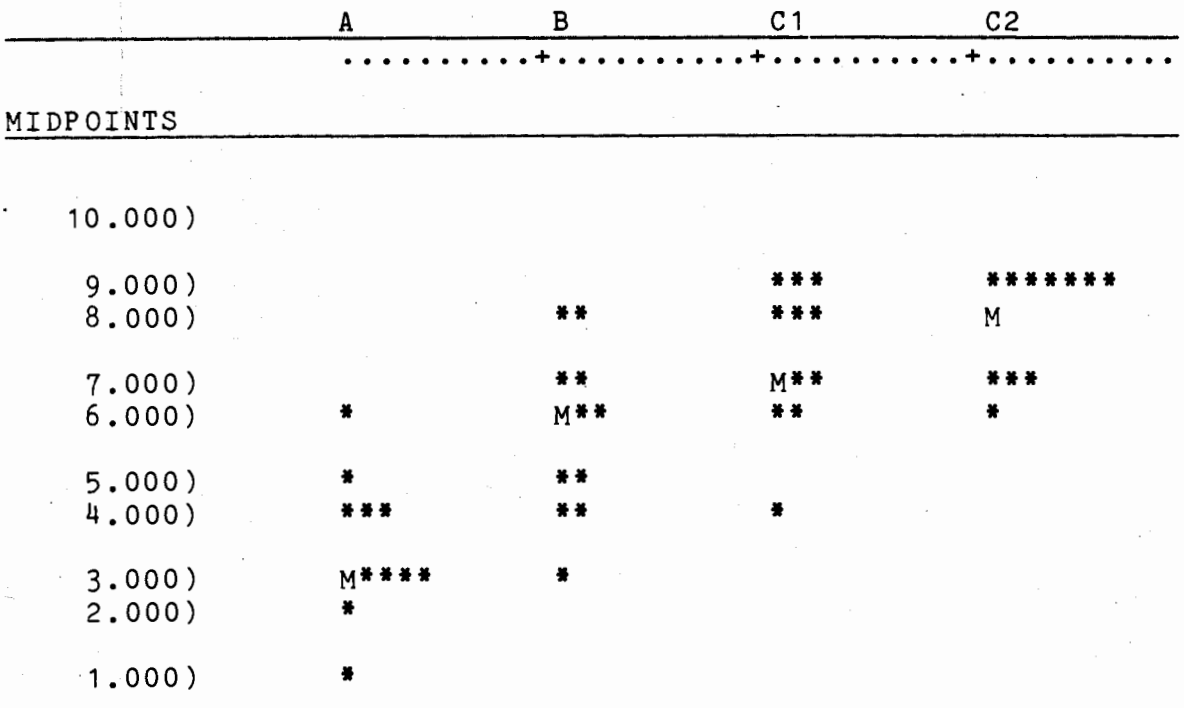
TABLE 8. MEAN WORKLOAD RESPONSES

	Flight			
	A	B	C_1	C_2
Mean	3.41	5.75	7.22	8.11
Standard Deviation	1.33	1.54	1.41	1.18

The data that were used to generate these means are available in the scatterplot of figure 3.

FIGURE 3

WORKLOAD RESPONSE SCATTERPLOT



Since there appeared to be variability across the flight difficulty levels, further analysis was required. Analysis of variance (ANOVA) was selected as the technique to be applied. Before this was accomplished, Hartley's Fmax Test was applied to determine if the four flight levels were relatively similar in internal variability across pilots. This was, in fact, the case with Fmax equal to 1.73, which did not exceed the cutoff. The analysis of variance was accomplished, and the results were significant ($F=62.3$ $(3,33df)$ $P<.001$). This meant that for the workload response, there was significant variability across

flights, but identification of which specific pairs of flights were different beyond chance would require additional testing.

This testing took the form of a Newman-Keuls Analysis, which is similar to Nemenyi's Test used previously. The mean or average workload rating for each flight was placed in order of its size from lowest to highest, as depicted in the table below.

TABLE 9. NEWMAN-KEULS COMPARISON OF MEAN WORKLOAD RESPONSES

		Flight			
		A	B	C_1	C_2
	Mean	3.41	5.75	7.22	8.11
A	3.41		2.34**	3.81**	4.70**
B	5.75			1.47**	2.36**
C_1	7.22				.89
C_2	8.11				
			** Significant P<.01		
Cutoff for Significance			1.01	1.15	1.24

The difference between each pair of means was computed and compared against a cutoff listed at the bottom of table 9. Those differences which exceeded the cutoff were marked with two asterisks. The table indicates that the only pair of flights which were not significantly different were C_1 and C_2, indicating that the inflight emergency did not increase the pilot's perceived workload.

The time delay between the query tone and each pilot's response was measured and recorded. Mean delays for each flight are presented in table 10.

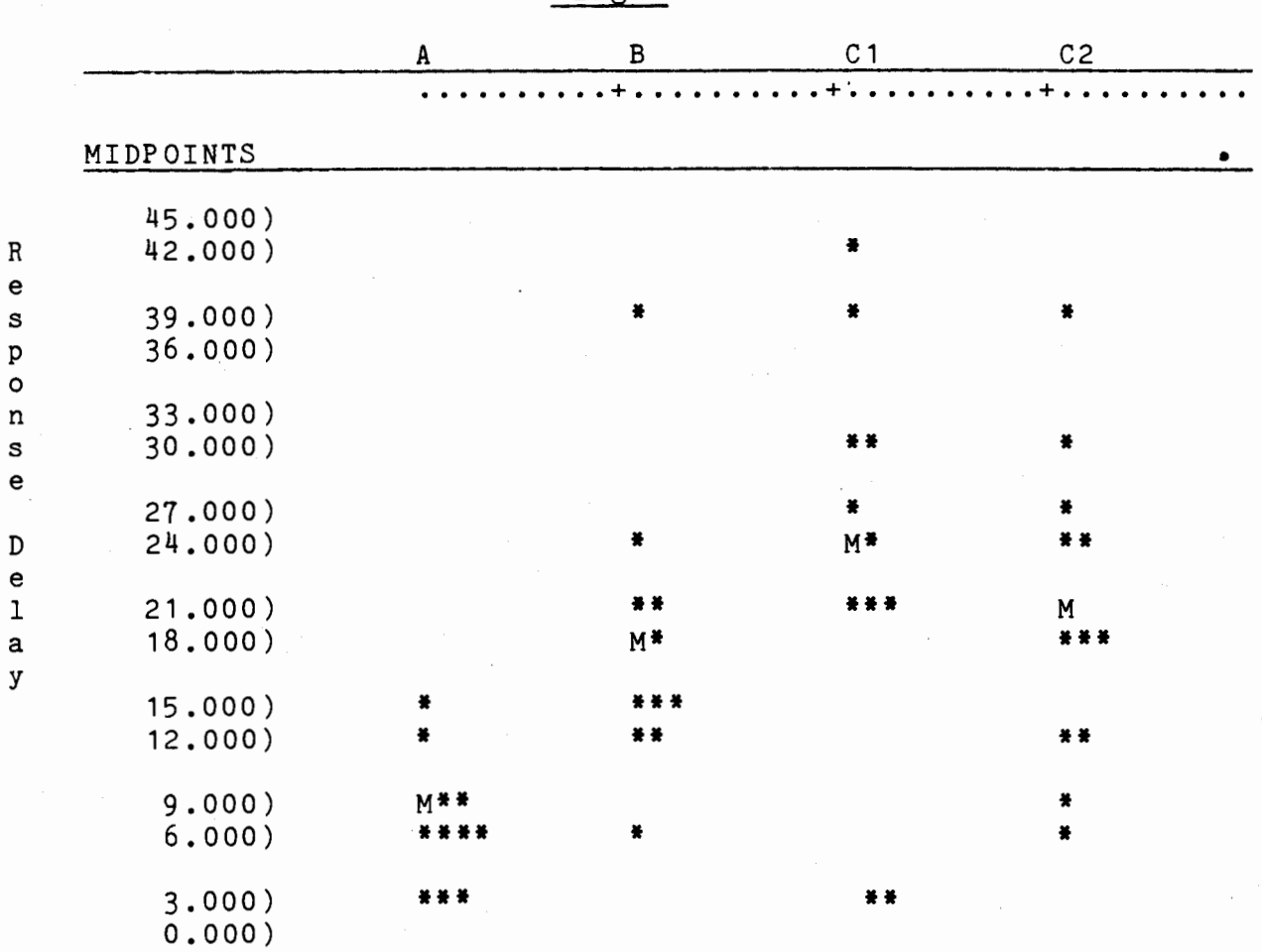
TABLE 10. MEAN RESPONSE DELAY (SECONDS)

	Flight			
	A	B	C_1	C_2
Mean	7.51	18.00	23.79	19.67
Standard Deviation	4.10	8.07	11.25	9.70

The data upon which these means were based are available in the scatterplot of figure 4.

FIGURE 4

RESPONSE DELAY SCATTERPLOT (SECONDS)



(Group Means Are Denoted By M's)

The upward trend across the flight difficulties holds for segments A, B, and C_1, but then a downturn appears in segment C_2. Hartley's Fmax Test was applied to the data to determine if the four flight levels differed significantly in internal variability. The Fmax was 7.53, which was significant (P .05). This meant that special precautions had to be taken with the analysis of variance of response delay data, or an incorrect finding might occur. The analysis of variance produced an F=9.78. If Hartley's Fmax had not been significant, this would have been tested against a cutoff equal to 4.51 (df-3,33), which it did exceed. In order to make the analysis of variance more conservative, the degrees of freedom used to select the cutoff from table 9 were halved to 1,16. The cutoff chosen was 8.53, which the F value still exceeded (Greenhouse - Geiser Test, See Morrison, 1976, P214). This meant that there was significant variability across the flights, and the Newman-Keuls Analysis was again applied. This procedure has already been explained. The results are presented in table 11.

TABLE 11. NEWMAN-KEULS COMPARISON OF MEAN RESPONSE DELAYS

		Flight			
		A	B	C_2	C_1
	Mean (sec)	7.51	18.00	19.67	23.79
A	7.51		10.49**	12.16**	16.28**
B	18.00			1.67	5.79
C_2	19.67				4.12
C_1	23.79				
Cutoff for Significance			8.61	9.86	10.63

**Significance $P<.01$

The differences between flights were not as clear as they had been for the workload response data. The results indicate a difference in delay between flights A and B and between A and both segments of C. Other than that, the remainder of the pairs were not significantly different. This included the pair C_1 and C_2. The reversed order of their means has little importance since one cannot conclude that they are different for reasons other than chance. This meant that C_1 and C_2 could be pooled for further analysis. One explanation of this finding could be that the "emergency" caused an increase in the pilot's level of arousal or activation and produced more rapid responses.

A comparison of the relationship between the two inflight measures of workload response and delay was made for flights A, B, and C. Since C_1 and C_2 were not found to differ on either variable, they were not separated for this comparison. The Pearson Product-Moment Correlations of the two variables are reported in table 12. What they indicate is that as the flights became more difficult, the relationship between workload response and the delay of that response broke down, then, at the most difficult level, the two variables began measuring different aspects of the pilot's workload experience.

TABLE 12. CORRELATION OF WORKLOAD AND DELAY

Flight	r_p
A	.748*
B	.585*
C	.211

*Significant from zero ($P<.05$)

The nature of these relationships can be brought into focus by examining the plots in figures 5, 6, and 7, respectively. The least squares regression line is presented for flights A and B. The amount of scatter around the regression line is inversely proportional to the magnitude of the correlation; i.e., the more scatter, the weaker the relationship. No regression line is plotted for flight C because of the weakness of the relationship.

FIGURE 5
DELAY X WORKLOAD SCATTERPLOT FLIGHT A

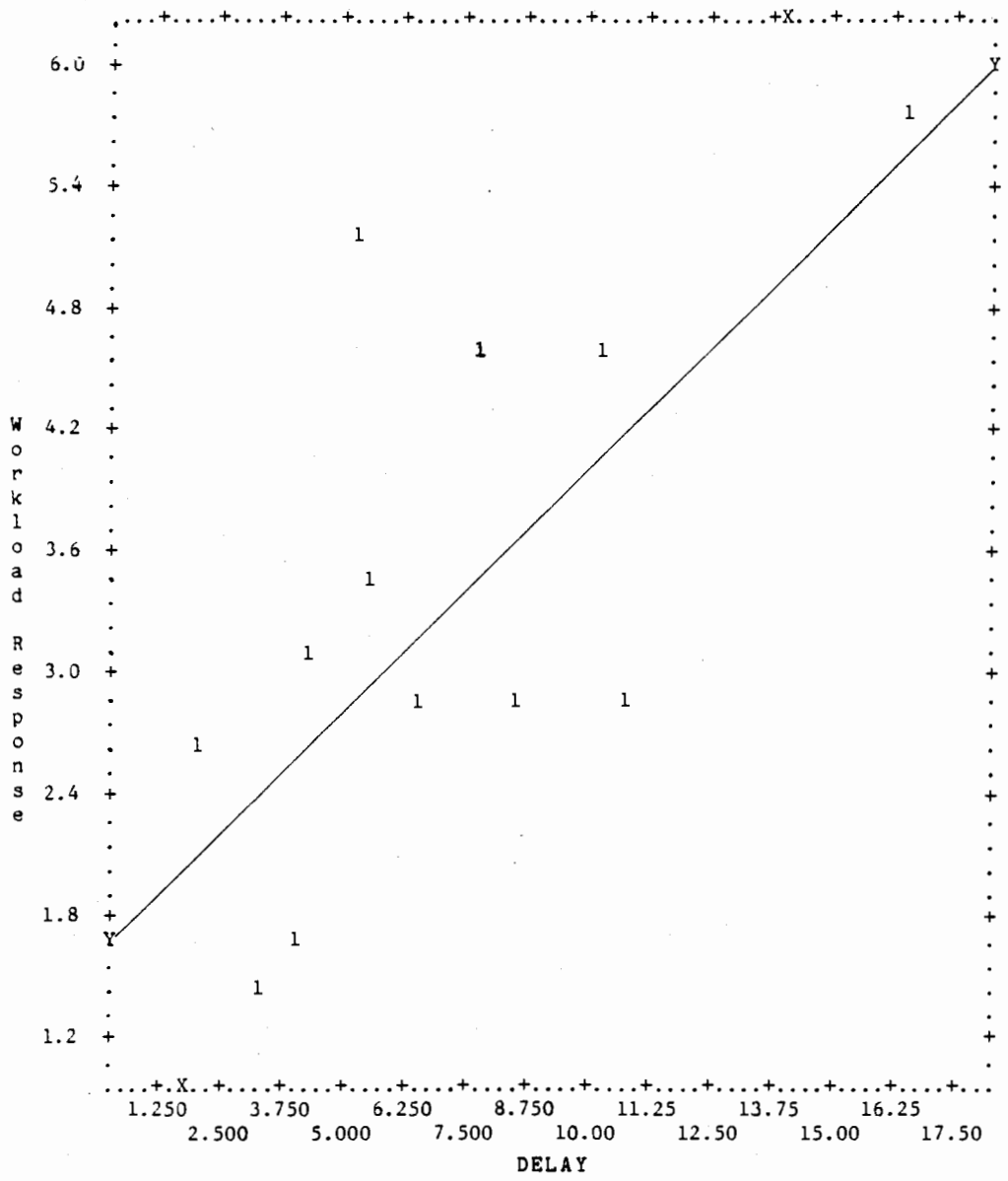


FIGURE 6

DELAY X WORKLOAD SCATTERPLOT FLIGHT B

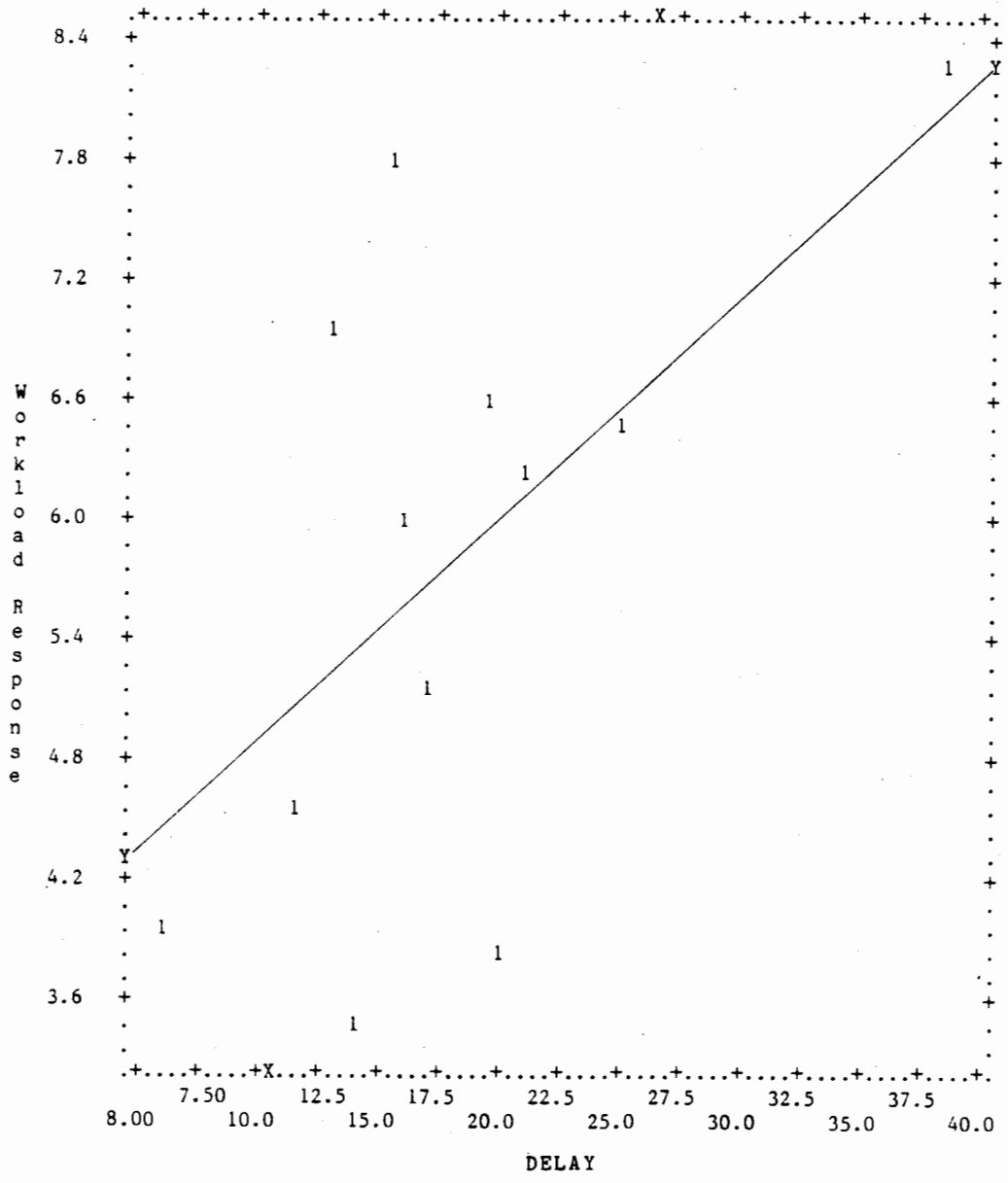
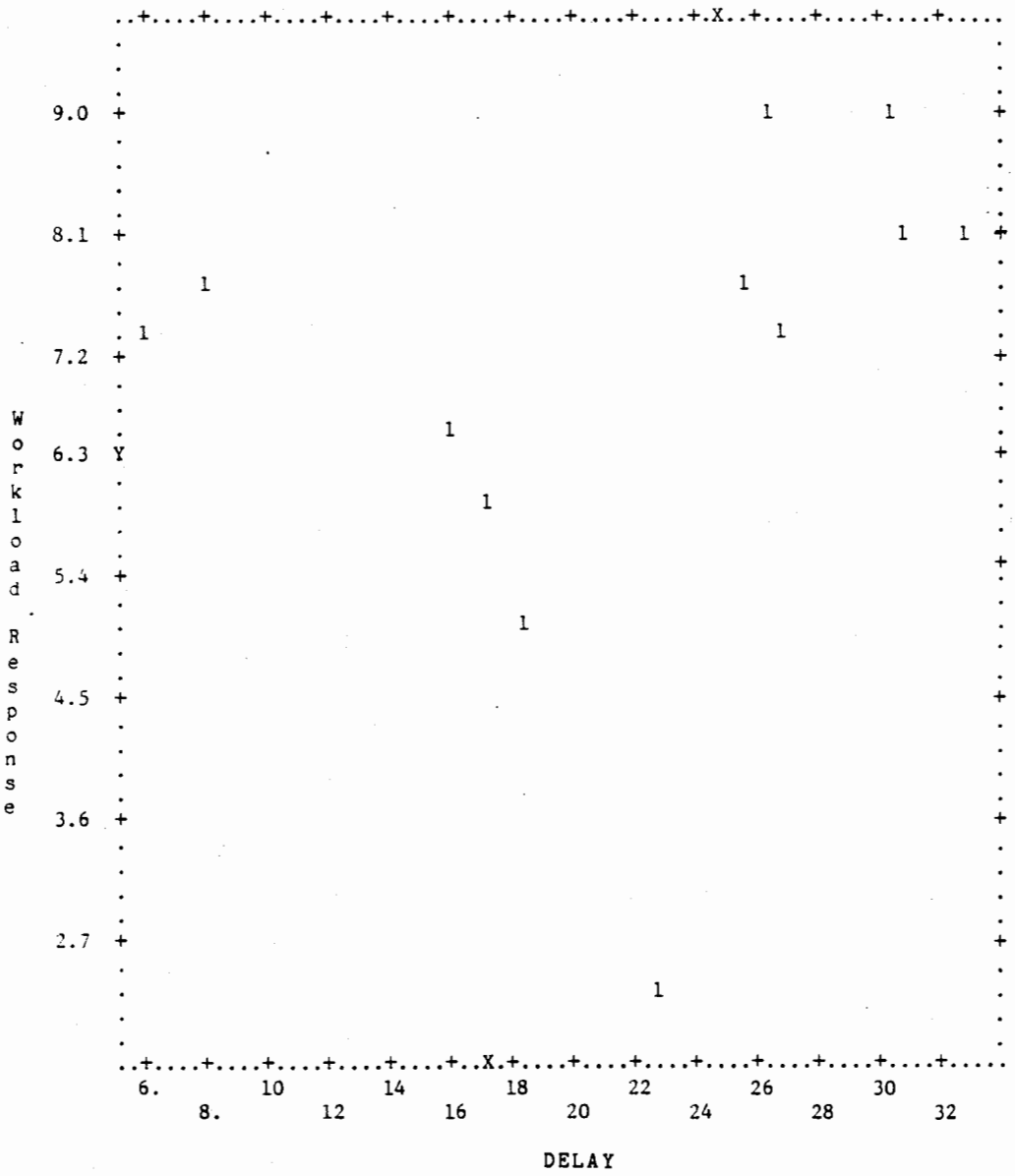


FIGURE 7
DELAY X WORKLOAD SCATTERPLOT FLIGHT C



POSTFLIGHT QUESTIONNAIRE

Immediately upon landing, each pilot was handed a questionnaire which contained four questions. These questions asked the pilot to evaluate how hard he/she had been working, how busy he/she had been, how hard he/she had to think, and finally, how he/she felt (stress) during the flight. The mean responses for each of these questions on each flight are presented in table 13.

There appeared to be an increase in the means across the three flights from A to C. The standard deviations across flights were very similar and Hartley's Fmax was not significant for any of the four questions. Analysis of variance was computed for each of the questions and the next table (14) presents the computed F values with their respective levels of significance.

TABLE 13. MEAN RESPONSES TO POSTFLIGHT QUESTIONNAIRE

Question	Workload			Busy			Think			Feel		
Flight	A	B	C	A	B	C	A	B	C	A	B	C
Mean	3.33	6.17	9.50	3.75	6.25	8.58	3.08	5.67	8.25	3.42	5.75	8.25
Standard Deviation	1.61	2.17	1.44	1.66	1.66	0.90	1.62	1.97	1.42	1.78	2.30	1.42
Hartley's Fmax	2.27			3.41			1.92			2.62		

TABLE 14. RESULTS OF ANOVA ON POSTFLIGHT QUESTIONS

Question	F	Level of Significance
Workload	5.045	$P < .05$
Busy	43.950	$P < .01$
Think	55.350	$P < .01$
Feel	41.560	$P < .01$

All questions showed significant variability across the three flights. The workload question was the weakest in separating the flights. Newman-Keuls Analyses were completed on each question.

Results for the workload question indicated a significant difference ($P<.05$) between flights A and C only. The other three questions provided difference between all pairs of flights ($P<.01$). This meant that pilots discriminated across flights A, B, and C in increasing order for the questions concerning busyness, thinking, and feeling, but only separated flights A and C for the workload question. These analyses were based on data indicated in the scatterplots in figures 8, 9, 10, and 11. One can clearly see the upward trend on all four questions using these plots.

FIGURE 8

QUESTIONNAIRE SCATTERPLOT - WORKLOAD

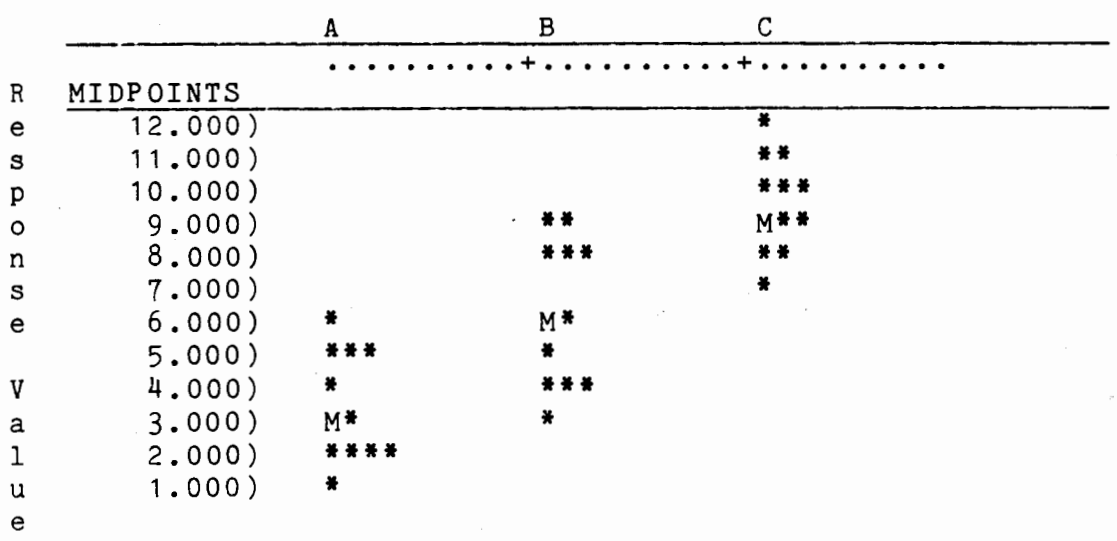


FIGURE 9

QUESTIONNAIRE SCATTERPLOT - BUSY

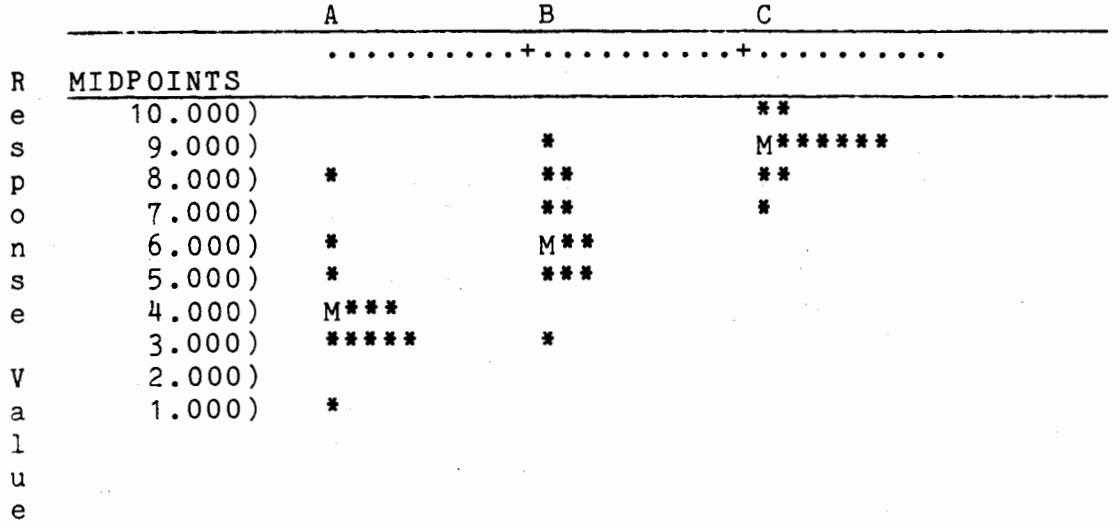


FIGURE 10

QUESTIONNAIRE SCATTERPLOT - THINK

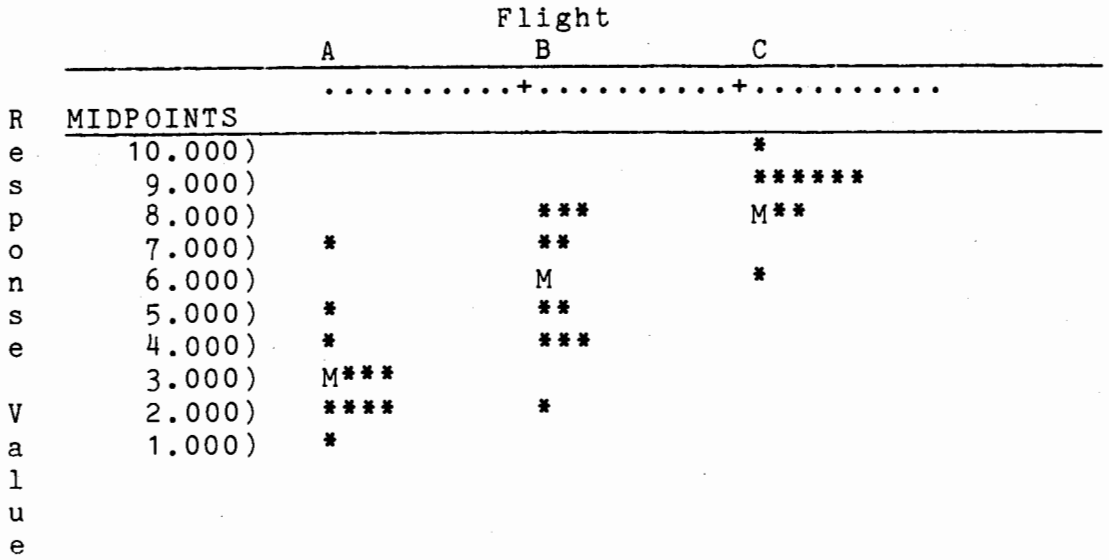
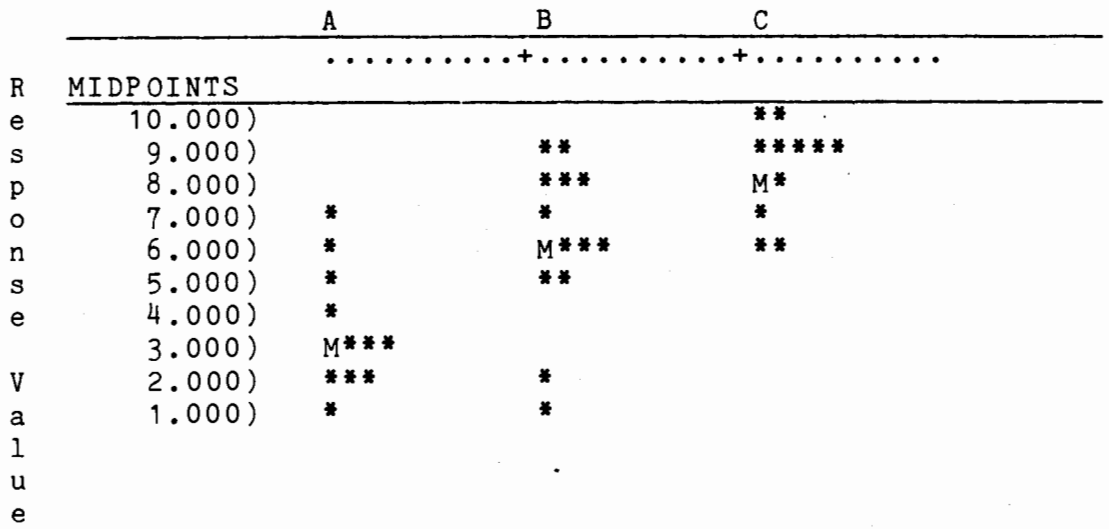


FIGURE 11

QUESTIONNAIRE SCATTERPLOT - FEEL



Factor analytic techniques were applied to the data in order to determine the degree to which the different measures were redundant. In other words, the question to be resolved was whether or not there were really four separate postflight questionnaire measures and two inflight measures (workload response and delay). Using the biomedical package for software, principle components analyses were applied to each flight, followed by verimax rotation. This provides the best orthogonal (the factors do not correlate) fit of a factor matrix to the data. A factor matrix shows the correlations of the variables in the experiment with

the factors which represent new variables made up of the data in the experiment. If two or more variables correlate well with a factor, it indicates that both may be measuring the same thing and could be combined in the future (assuming that the experimenter is comfortable with the sampling technique to begin with). A preanalysis criterion for factor rotation was set so that factors would cease being rotated out when approximately 90 percent of the variance was accounted for.

Tables 15 and 16 describe the factor structures of flights A and B, respectively, with factor loadings below 0.50 suppressed to zero for ease of interpretation and because they did not differ significantly from zero.

In flight A, two factors account for 89 percent of the total variance, indicating that there are primarily two measures, inflight and postflight.

TABLE 15. FACTOR STRUCTURE FLIGHT A

VARIABLE	FACTOR 1	FACTOR 2
Think	.967	----
Feel	.947	----
Busy	.932	----
Workload (Postflight)	.762	----
Delay	----	.936
Workload (Inflight)	----	.886

The questionnaire and inflight measures each cluster about themselves. The same is true for flight B, where two factors describe 89 percent of the variance. On both flights, there was some overlapping of the postflight workload question, but it was still a reasonable conclusion that there were essentially two measures taken during the lowest and the moderate difficulty flights.

TABLE 16. FACTOR STRUCTURE FLIGHT B

VARIABLE	FACTOR 1	FACTOR 2
Busy	.947	----
Workload (Postflight)	.901	----
Feel	.888	----
Think	.883	----
Delay	----	.942
Workload (Inflight)	----	.792

The results for flight C were distinct from those in A and B. The next table (17) lays out the factor structure for the most difficult flight. The first two factors accounted for only 60 percent of the variance, while four factors accounted for 89.4 percent. The two inflight measures loaded on separate factors and the busy scale separated onto its own factor. It was apparent that the changes introduced in flight C by the experimental design led to a measurement system which operated somewhat differently than it had for flights A and B.

TABLE 17. FACTOR STRUCTURE FLIGHT C

VARIABLE	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
Think	.914	----	----	----
Workload (Postflight)	.735	----	----	----
Feel	.732	----	----	----
Delay	----	.948	----	----
Workload (Inflight)	----	----	.967	----
Busy	----	----	----	.986

A final analysis employed multiple linear regression in an attempt to predict postflight responses from those made inflight. The independent variables in the regression were delay and workload response, while the dependent variables were four questions administered after the flights. Table 18 presents the multiple correlations and coefficients of determination (r^2).

TABLE 18. PREDICTION OF POSTFLIGHT MEASURES

PREDICTED VARIABLE	MULTIPLE r	MULTIPLE r^2
Feel	.79	.63
Workload	.78	.61
Think	.75	.57
Busy	.73	.55

These data were computed by pooling the data for all three flights and by rescaling the data to remove variability within subjects. This procedure did not change the nature of the data itself, but merely put it on a different base. The results of the regression analysis indicated moderate positive relationships in which from 55 to 63 percent of the variability was accountable. This also meant that sizeable proportion of the variability between inflight and postflight measures was not accountable. This was in line with the results of the factor analysis and meant, quite simply, that inflight measures are made by pilots with perhaps a different perspective than those after the flight is over.

DISCUSSION

The goals of this preliminary experiment were to determine whether or not workload could be reasonably measured in two words and whether this type of measurement was different from the traditional postflight questionnaire. Based on the results of this experiment, one can conclude that pilots are willing and able to make workload responses in flight and that these responses correspond to the difficulty level of the flight. As hypothesized, the more difficult flights tended to generate higher mean workload responses and, to a certain extent, longer response delays. Measures taken in flight appear to be different from those collected after the flight is over. These conclusions are based on the interpretations which follow.

The results of the task frequency tally indicated certain differences in pilot activity level across flights. This confirmed that the difficulty level of the flights could be varied using what amounted to qualitative distinctions and some educated guesswork concerning what would make the pilots work harder. The apparent differences in the preflight period were produced by the complexity of the initial flight clearance. Several pilots were heard to comment about the simplicity of flight A after they were given their radar-vector clearance. Flight C, in contrast, produced comments of a completely opposite nature and many pilots felt the need for repetitions of the clearance. This was reflected in the higher frequency of navigation adjustments (i.e., setting the NAVCOM frequencies, adjusting horizontal situation

indicator) and the elevated communication rate. The primary drivers of inflight pilot activity were again navigation and communication. Control movement frequencies (i.e., trim, throttle adjustment, pitch adjustment) did not differ across the flights. The flight scenario, in its entirety, did not drive pilot activity, but rather, certain elements of the scenario may have had more influence than others. As mentioned earlier, the drop in activity level after the emergency began in flight C was a design artifact. It was assumed that the emergency itself would keep the pilot busy. It was not known before the experiment began that communications and navigation complexity would be the most relevant categories. Once the emergency began, ATC inputs decreased considerably and navigation requirements only called for one turn onto the localizer beam and the final approach.

Pilot activity level is only one element of workload. It is also the most directly observable. It does not, however, account for all the internal processes that occur within the pilot and was never meant as more than a rough check on the selection of the flight difficulty levels.

There appeared to be an impact of the design artifact on the mean frequency of missed responses during the flights. It was assumed that missed responses were indicative of high workload level and yet, when the workload was expected to be highest during the emergency in flight C, the mean for missed responses was only slightly higher than flight A. Once the initial reaction to the failed engine situation had stabilized, pilots, for the most part, performed very methodically and found the time to make the responses. Pilots were never given a set priority on the workload responses other than to make them "as quickly as possible." Increased activation level caused by the emergency was posed in an earlier section as one possible explanation for the lack of delay in workload responses.

Pilots' responses during the emergency segment were not significantly higher than they had been during the first part of flight C. They were also not significantly lower, even though the mean frequency of missed response (with automatic 10- and 60-second delays) was lower. This meant that when they made a response during C_2, it was a high response. Although the activity level decreased for navigation and communication in C_2, the workload responses did not. This indicated that some nonobservable elements may have been driving these responses. Although not significant, there appeared to be some decrease in nontask appropriate behavior (especially verbalization) during C_2. This too may have been an indicator that pilots were focusing on the task of safely landing the aircraft.

The separation between flights that was demonstrated by the inflight measure of workload documented the utility of the technique. Pressing buttons every minute in response to a query tone is not what one would call an unobtrusive measure. This was

known from the beginning. It was apparent, in a previous study reported by Rosenberg, Rehmann, and Stein (1982) that the task itself probably contributed somewhat to the workload. However, it was felt that this was acceptable if the method provided something unique to the measurement of pilot workload.

There was an ordinal relationship between the three levels of flight difficulty (as determined before the experiment by pilot experts) and the mean workload responses for each flight. On the average, increases in difficulty led to corresponding increases in perceived workload. This occurred despite the fact that the flights were presented in counter-balanced order to control for learning, experience, and habituation. To carry this logic somewhat further, it was apparent that the manipulation of certain preflight and inflight variables, primarily initial clearance and level and type of ATC, had dramatic effects on perceived workload. To the extent that the inflight task tally can be accepted, despite its short-falls, one could speculate that the major elements of workload may have been those involved in planning, navigating, and communicating, rather than in the actual control (without autopilot) of the aircraft. This is obviously a concept which goes well beyond the data and should be researched further.

The results for the response delays were not as definite as those for the responses themselves. It was assumed that delay or latency was the more objective measure, depending less on what the pilot held in consciousness at any point in time and more on the basis of primal elements of the construct "workload." Using delays, flights A and C were clearly separated by pilots' behavior, with C having the longer delays. Flights A and B were also separated, but flights B and C were not sufficiently different to reach significance. We expected workload responses and the delays to be correlated, but also felt that the relationship would not be perfect. Had the relationship been optimal, then the measures would have been largely interchangeable. The first indication that flight C was somewhat different in terms of inflight measurement came from the scatterplots of workload and delays (figures 5, 6, and 7). A moderate positive relationship existed between the two variables for flights A and B, but not for C. In flight C, knowing what workload responses were would not greatly facilitate prediction of how long pilots took to make them.

The factor analysis showed that, although workload and delay were only moderately correlated, they were related strongly enough to cluster together in the same factor for flights A and B. Further, postflight measures, which were collected with a questionnaire, produced their own distinct factor. In flight C, workload and delay were each contained in their own separate factors. What this all meant was that inflight and postflight measures are different. By itself, the factor analysis does not prove that one set of measures is superior over the other. It means that they capture independent aspects of the test and, therefore, both should be collected.

The means of pilots' responses to all the postflight questions were in the same order as the three difficulty levels of flight; the more difficult-- the higher the response. The weakest question was the one directly related to workload. It only separated flights A from C. This question was a local adaptation of Sheridan and Simpson's (1979) Cooper-Harper type scales. These scales employ a multiple anchoring system so that every scale point has some verbal description attached to it. This was in contrast to the busyness, thinking, and feeling scales which were only anchored at the end points. These questions provided significant differences between all pairs of flights, with A, B, and C showing responses of increasing magnitude. The factor analysis indicated that although the workload question was weaker in separating the flights, it loaded on the same factor as the other postflight questions for flights A and B. In the low-to-moderate difficulty flights, it did not matter what the question was called, a similar pattern emerged. There was a general postflight response where pilots did not discriminate very much between the questions. The fact that the inflight measures clustered in one factor indicated that regardless of how they were defined initially, both measures captured the same aspects of the test (at least for flights A and B).

When considering the most difficult flight (flight C), something changed. This was foreshadowed by the lack of correlation between delay and workload. The factor structure became more complex. The inflight measures each went their own way and one postflight measure (busyness) split off onto its own factor. Given that flight C had the highest mean frequency of missed responses, it would have been feasible that the resulting "10" workload values and 60-second delays assigned by the computer could have driven the workload response/delay relationship to a new high. That this did not occur was further evidence of the distinction between response and delay. A high workload response was no longer preceded by a long delay. The relationship of these two variables had taken on a different meaning than it had for the easier flights. There was evidence that under the heaviest workload, pilots became somewhat more discriminating in their evaluation of their experience. This was indicated by the fact that the busyness scale no longer clustered with the responses on the rest of the questionnaire.

It appears that at the lower difficulty levels, participants' perception of their workload more closely approximated a unidimensional space with respect to time. When they were actually flying, subjective impressions and objective delays formed a unified entity. When the flight was over, a separate impression took hold which was based on the remembered experience. At the highest level of difficulty, the multidimensional character of workload, which has been so often cited in the literature, made its appearance. Unless it is known precisely where a given flight is on a difficulty continuum, then there is justification for collecting both inflight and postflight workload measures.

It will take a great deal more research before a firm relationship is established between one or both sets of measures and before an idealized, ultimate indicator of how hard a person is working is clearly defined.

While this program at the FAA Technical Center has generated the beginnings of an inflight workload measurement system, more refinement will be required before a reliable tool is available. A workload measurement system will some day be linked to an effective pilot performance index, such that the impact of new airborne concepts can be adequately addressed.

CONCLUSIONS

1. Given the opportunity, pilots were willing and able to make inflight workload judgments.

2. Workload judgments were directly related to the experimentally induced difficulty level of the flights.

3. Response latencies were ordinaly related to difficulty level, but they did not separate the intermediate from the most difficult flight.

4. Factor analysis of all dependent variables indicated that both inflight and postflight measures are necessary to obtain a complete view of pilot workload.

5. Workload measurement research should continue in order to refine the tools necessary to evaluate the impact of new systems.

REFERENCES

1. Chiles, W. and Alluisi, E.A., Human Factors, On the specification of operator or occupational workload with performance measurement methods. 1979, 21 (5), 515-528.

2. Cooper, G.E. and Harper, R.P., NASA Tech Note (Tn D-5153), The use of pilot rating in the evaluation of aircraft handling qualities. Washington, D.C., 1969.

3. Eggemeir, F.T., Some current issues in workload measurement. Proceedings of the Human Factors Society - 24th Annual Meeting, 1980, 669-673.

4. Johannsen, Workload and workload measurement. (In N. Moray Ed.). Mental Workload: Its Theory and Measurement. New York: Plenum Press, 1977, pp 3-11.

5. Katz, J.A., Pilot Workload in the Air Transport Environment: Measurement, Theory, and the Influence of Air Traffic Control, Flight Transportation Laboratory, MIT, Cambridge, Mass., May 1980. (FTL Report R80-3)

6. Linton, M. and Gallo, P. The Practical Statistician. Wadsworth: Belmont, CA., 1975.

7. Morrison, D.F., Multivariate Statistical Methods. McGraw Hill: N.Y. 1976.

8. Roscoe, A.H., Introduction. IN AGARD Monograph. Assessing Pilot Workload. Hartford House, London, 1978.

9. Rosenberg, B., Rehmann, J., and Stein, E.S., The Relationship Between Effort Rating and Performance in a Critical Tracking Task. FAA Technical Center Technical Report (DOT/FAA/EM 81/13), Atlantic City, N.J., October 1982

10. Sheridan, T.B. and Simpson, R.W., Toward the Definition and Measurement of the Mental Workload of Transport Pilots. Technicalp dpr1t, Flight Transportation Man/Machine Lab, MIT, Cambridge, Mass., DOT. OS-70055, 1979.

11. Willeges, R. and Wierwille, W., Behavioral measures of aircrew mental workload. Human Factors. 1979, 21, 549-574.

12. Winer, B.J., Statistical Principles in Experimental Design. McGraw Hill: New York, 1962.

APPENDIX A

PILOT EXPERIENCE AND MEAN WORKLOAD RATING

PILOT			MEAN WORKLOAD RATING		
			Flight		
Number	Total Hours	Instrument Hours	A	B	C
02	12,000	2,500	2.93	3.92	5.00
03	7,200	450	5.15	6.21	7.44
04	2,600	200	4.50	7.77	7.91
05	2,250	175	2.67	5.73	6.71
06	4,000	250	2.86	8.13	8.04
08	5,000	600	4.50	6.93	8.91
09	2,000	200	2.92	4.56	7.88
10	-----	----	3.31	6.00	7.79
11	1,600	300	1.40	3.86	7.41
12	1,600	200	1.61	3.50	6.00
13	4,000	1,500	5.73	6.43	8.90
14	14,500	1,400	3.53	6.54	8.05

APPENDIX B

SCENARIO GUIDE B - 1

PURPOSE.

This is a workload measurement task with ATC involvement to add realism and workload to the subject pilot.

Scenario 1 is intended mainly for background realism with little actual ATC workload for the subject.

Scenario 2 provides background realism and moderate ATC workload for the subject.

Scenario 3 provides background realism and a rapid fire sequence of events, typical of those found in a busy terminal area.

INSTRUCTIONS.

1. The ATC specialist utilizing these scenarios should read the lines entitled "ATC."

2. The lines entitled "GAT" (General Aviation Trainer) are known or assumed responses from the subject pilot.

3. Any "ad lib" questions from the subject pilot will have to be responded to in like fashion by calling upon your own experience.

4. Listen carefully to any "readbacks" to insure accuracy, since any erroneous wording will invalidate the experiment.

5. Flight Environment

   a. Scenario 1: wind - calm, turbulence - none
   b. Scenario 2: wind - 090015, turbulence - 1/3 of maximum
   c. Scenario 3: wind - 090015, turbulence - 2/3 of maximum

SCENARIO #1

Flight A

GAT Millville radio, N1477D IFR to Atlantic City

ATC 77D, roger, clearance on request, runway 28 in use, wind calm, altimeter 30.06

ATC (30 seconds later) N1477D Millville, clearance

GAT Go ahead

ATC ATC clears N1477D to the Atlantic City Airport via radar vectors to the ILS runway 13 final approach course. Maintain 2,000 feet. Fly runway heading after departure, squawk 0253, contact Atlantic City approach on 124.6 when airborne, clearance void if not off by (issue time 10 to 15 minutes from now) time is now . Issue present time.

GAT Millville radio, N1477D departing runway 28

ATC N1477D roger, no reported traffic, wind calm

GAT Atlantic City approach N1477D's with you

ATC 77D radar contact, altimeter 30.06

ATC (when aircraft reaches 2,000)
77D turn right heading 070

ATC (continues to adjust flight path as necessary)

ATC (give intercept heading when appropriate) 77D, position 7 from the marker, turn right heading 100, maintain 1,600 or above until the localizer, cleared for the ILS approach, tower eighteen nine at the marker.

GAT Tower, 77D with you at marker

ATC 77D cleared to land, wind calm

SCENARIO #2

FLIGHT B

GAT Millville Radio, N1477D's IFR to Atlantic City

ATC N1477D, Roger standby

ATC (one minute later) N1477D, Millville I've got your clearance

GAT Go ahead

ATC ATC clears N1477D to the Atlantic City Airport, as filed. Maintain runway heading after departure until leaving 1,000 feet, then turn right heading 100, intercept V166. Intercept the Coyle 230R at 2,000 feet or below. Maintain 3,000. Squawk 0112, contact Atlantic City approach control on 124.6 after departure clearance void if not off by . (issue time 10 minutes from now)

GAT (will read back clearance and might advise Millville Radio that they are departing Millville now)

ATC (listen closely to readback to insure its correctness)

GAT Atlantic City approach, N1477D with you

ATC 59437 you squawking 0213

ATC 1477D radar contact altimeter 29.98

ATC (approximately 1 minute later) 77D I'm gonna have to amend your clearance limit, you're now cleared to Tragg intersection hold south on V-1, 1 minute pattern, left turns, maintain 3,000 feet, expect further clearance at . (issue time 20 minutes from now)

GAT (will read back)

ATC And 77D, it should only involve about one trip around the pattern

ATC 59437 I'm getting your beacon now, push the ident for me

ATC Acom 166 you're cleared for the VOR-A approach to Bader, give me a check out of 4

ATC 437 radar contact, fly heading 340 vectors to the VOR-A final approach course at Hammonton

ATC (before 77D gets to Tragg) 77D you,re cleared on course as previously cleared, maintain 2,000 feet now

ATC 77D I've got an Allegheny commuter inbound to Millville, did you pick up any turbulence in the vicinity of the airport?

ATC (acknowledge for the information) then Acom 184 did you get that OK?

ATC Acom 184 you're 7 from Ladie, turn right heading 130, maintain 2,000 or above til on the localizer, cleared localizer runway 10 approach.

ATC Roger Acom 184 and I've got your clearance to Atlantic City when your ready

ATC Acom 184 cleared to the Bader Field airport via radar vectors, maintain 3,000 feet, call Atlantic City approach when airborne, check with Millville radio for release

ATC Acom 184 that's correct and you can keep the same transponder code.

ATC (2 miles before the localizer) 77D, position 6 from Naada, turn right heading 100, intercept the localizer at 1,600 feet or above, cleared for the ILS

ATC Roger 437 understand you have Hammonton in sight and you're cancelling instruments now

ATC Atlantic City Tower 77D's with you at the marker

ATC 77D cleared to land wind 110 at 5

ATC (1 minute later) 77D are you going to need progressive or are you familiar with the airport?

ATC (acknowledge whatever response you get)

ATC (after touchdown) 77D it will be a right turn at "J" ground point 9 when you clear

SCENARIO #3

FLIGHT C

GAT Millville Radio, N1477D IFR to Atlantic City

ATC 77D, roger, standby

ATC (1 minute later) N1477D Millville, clearance

GAT Go ahead

ATC ATC clears N1477D to the Leeah intersection via V166. Depart runway 28, maintain runway heading until crossing the Cedar Lake 226 radial, then turn right heading 100 to join V166. Cross the Cedar Lake 226 radial at or below 1,000 feet, climb so as to reach 3,000 feet by Leeah. Maintain 3,000 feet. Squawk 0144, contact Atlantic City approach control on 124.6, clearance void if not off by . (issue time 10 minutes from now)

GAT (will read back and might advise departing Millville)

GAT Atlantic City approach N1477D with you off Millville

ATC 77D standby

ATC (5 seconds later) 77D off Millville roger we've just lost the radar; report intercepting V166

GAT 77D's intercepting V166

ATC Roger 77D, you landing Bader or Atlantic City Airport

GAT Atlantic City

ATC Roger 77D and what's your altitude now

GAT Out of (whatever)

ATC 77D I think we're getting the radar back, what code are you squawking

GAT 0144

ATC OK, push the ident for me

ATC 77D radar contact, maintain 3, altimeter 30.06.

ATC (5 miles before Leah 77D cross the Cedar Lake 170 radial at 3,000 then descend and maintain 2. You're cleared to Tragg intersection via V166 and V1, hold southwest on V1, 1 minute pattern left turns, I say again left turns, expect further clearance at . (issue time 30 minutes from now)

GAT (will read back clearance)

ATC US Air 176 you're cleared straight-in ILS-13 approach via the Cedar Lake 100 radial

ATC 77D the radar's getting kind of shaky again, report entering holding at Tragg

ATC 77D we've got some reports that Kenton is off the air, can you tune Kenton and let me know what radial you're on

ATC 77D what's your time en route from Tragg to the Marker?

GAT (response)

ATC 77D amend your altitude to read climb in the holding pattern to maintain 4,000 feet

ATC US Air 176 tower eighteen nine

ATC N3177E you're cleared to the Tragg intersection via V-1, hold southwest on the airway maintaining 5,000 feet, expect further clearance at _ _ _ _. (30 minutes from now)

ATC Acom 184 off Millville radar contact, fly heading 160, vectors for the VOR-A approach at Badar, maintain 3,000

ATC All aircraft on the frequency, the latest special now gives 300 over and 1 mile with thunderstorms approaching from the west. Wind 090 variable 150, 12 to 18 at Atlantic City

ATC 77D you might see traffic off your left side 9 o'clock 3 miles opposite direction, do you have him?

GAT (no)

ATC Roger 77D I'll keep you advised.

ATC 77D what's your speed on final going to be?

GAT (response)

ATC Roger and 77D what's your final approach speed?

ATC Acom 184 turn left heading 140.

ATC 77D you picking up any turbulence?

GAT (response)

ATC 77D you're clear of that previous traffic

ATC (after aircraft completes holding pattern and is inbound) 77D cleared on course. (At this point, fail the right engine)

ATC All aircraft landing Atlantic City Airport, we've just had wind shear reported off the approach end of runway 13 by a Boeing 727

ATC Acom 184 turn left heading 120, you're cleared for VOR-A approach to Bader

ATC 77D descend to 2,000, report leaving 4

GAT 77D leaving 4

ATC 77E cleared to 4

ATC 77D give me a check out of 3 also

GAT 77E cleared to 3,000 now

ATC (2 miles before 77D intercepts localize) 77D position 6 from the marker, turn right heading 100, intercept the localizer at or above 1,600, cleared ILS approach

ATC Aircraft on the ground Bader standby, I have traffic on the approach

ATC Eastern trainer 405 radar contact, 15-minute delays landing AC, unable practice approach unless you're full stop

ATC 77D tower eighteen nine at the marker

At some point the subject will recognize the engine failure and advise ATC

GAT Approach 77D my right engine just quit

ATC (a typical reply might be) 77D roger you are cleared straight-in approach runway 13, are you requesting any equipment?

ATC (if reply is yes) Roger 77D, the equipment is on its way and how many souls on board?

ATC (if reply is no) Roger 77D, call the tower now on 118.9

GAT Tower, 77D marker inbound

ATC 77D cleared to land, wind 160, 12 gusting 18

ATC 77D you gonna be able to make the airport OK?

ATC (after landing) 77D any right turn, ground point 9 when you clear.

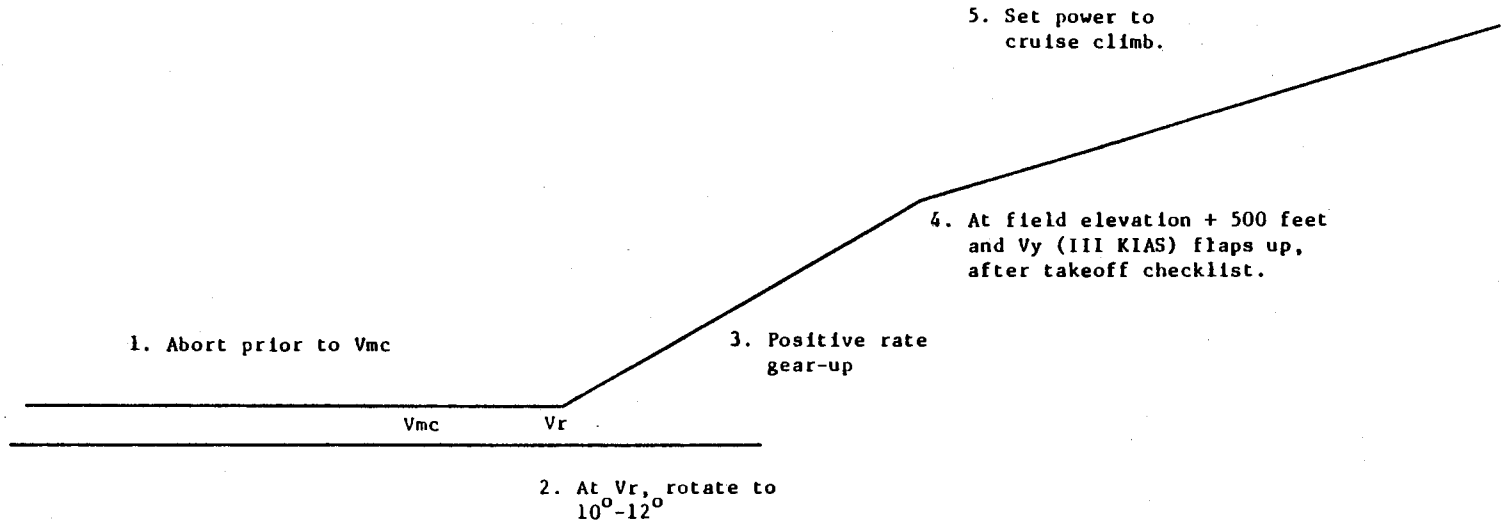
APPENDIX C-1

MANEUVERS BRIEFING

(TRAINING/SCREENING OF PILOTS)

NORMAL TAKEOFF	At V_r, rotate to 10° - 12°. When a positive rate of climb is indicated, retract the landing gear. Maintain 10 - 12 and accelerate to V_y (111 KIAS) (minimum) to 500' on the radio altimeter. At 500', retract the flaps, adjust power to cruise climb (32.5 in hg and 1900 RPM) and complete the after take-off checklist. Complete climb to altitude at this power setting (speed 120 to 130 KIAS).
EN ROUTE	After leveling off adjust power to 31 to 32.5 in hg and 1800 RPM's (65% power 75 ± 5 KIAS) trim the mixtures, switch, and complete the cruise checklist.
HOLDING	When 3 minutes or less from a clearance limit, start a speed reduction to 140 KIAS (23 in hg, 1800 RPM and 15° flap). Enter all holding patterns using either a parallel, teardrop, or direct entry, as appropriate.
EN ROUTE DESCENT	At approximately 3 times your altitude, from destination, or when cleared by ATC, establish a 3° to 5° descent (approx. 1000 FPM) adjust the power to 21 to 23 in hg and 1800 RPM (45% power 165 ± 5 KIAS) and complete the descent in range checklist to flaps.
NORMAL LANDING (2 ENGINES) 55° FLAPS	Fly 130-140 downwind with flaps 15°. Drop gear prior to turning baseleg. Flaps 40° turning base. Maintain 110 to 115 KIAS on final until landing is assured.
MISSED APPROACH/ GO AROUND	Apply max power and rotate to 10° to 12° when a positive rate of climb is indicated on the radio altimeter, gear-up, flaps 15° and continue with the normal take-off procedure.

DEPARTURE PROFILE

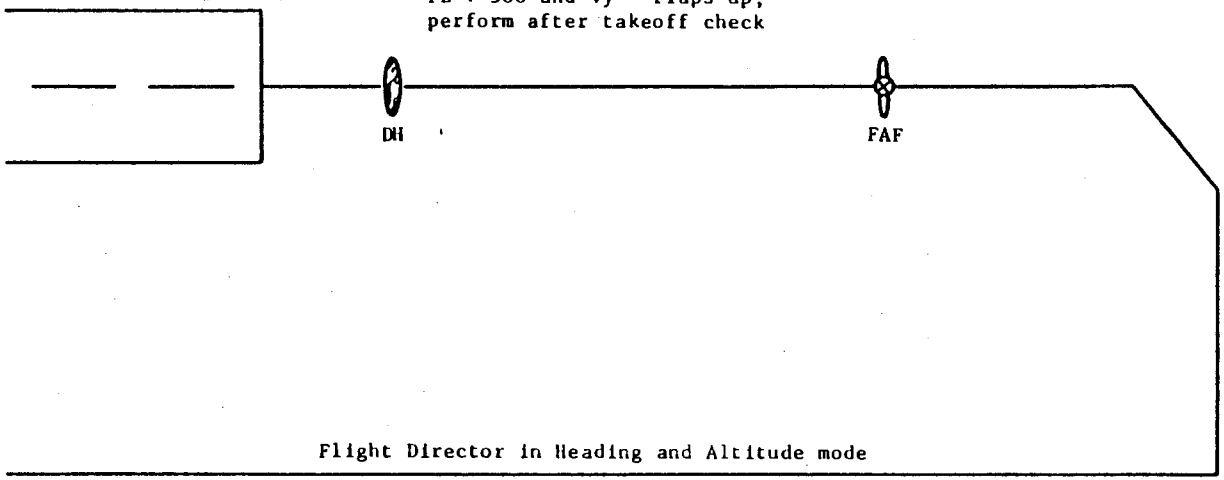


ILS and FLIGHT DIRECTOR PROFILE

6. At minimums:
   a. Runway in sight - land
   b. Runway not in sight:
   Max power - flaps 15
   Push GA button - rotate to command bars.
   Positive rate - gear up
   FE + 500 and Vy - flaps up, perform after takeoff check

5. At FAF:
   1. Time
   2. Landing checklist

4. 1 dot prior to G/S Intercept - gear down.
   At G/S intercept - full flaps (110 KIAS)



3. On LOC Intercept confirm IIDG light out - VOR/LOC light on.

Flight Director in Heading and Altitude mode

1. Cruise descent (45% power)
   Perform descent checklist
   Complete landing checklist to flaps.

2. When vectored to base leg (less than 90 degrees to final)
   Flaps 15 and select approach mode on the flight director.

APPENDIX D-1

WORKLOAD SCALE INSTRUCTIONS FOR THE PILOT

The purpose of this research is to obtain an honest evaluation of pilot workload or how hard the pilot is working. By workload, we mean all the physical and mental effort that you exerted in order to fly this aircraft. It includes planning, thinking, navigation, communication, and controlling the airplane. The way you will tell us how hard you are working is by pushing the buttons numbered from 1 to 10 on the box mounted below the throttles. I will review for you what these numbers mean in terms of workload. At the low end of the scale, 1 or 2, your workload is low--you can accomplish everything easily. As the numbers increase, your workload is higher. Numbers 3, 4, and 5 represent increasing levels of moderate workload, where the chance of error is still low, but getting higher. Numbers 6, 7, and 8 reflect relatively high workload, where there is some chance of making mistakes. At the high end of the scale, are numbers 9 and 10 which represent the very high workload, where it is likely that you will have to leave some tasks incomplete. All pilots, no matter how proficient and experienced, can be exposed to any and all levels of workload. It does not detract from a pilot's professionalism, when he states that he is "working hard" or "hardly working". Feel free to use the entire scale, and tell us honestly how hard you are working. You will hear a tone and the light on the box will come on. Push the button of your choice as soon as possible, after you hear the tone. Then the red light will go out. Remember that this data is not being collected by name, and your privacy is protected.

APPENDIX D-2

PARTICIPANT CODE DATE

FLIGHT WORKLOAD QUESTIONNAIRE

INSTRUCTIONS: The four questions which follow are to be completed at the end of each flight. Your responses should concern only the flight you have just completed. Disregard all others. Your name is not recorded on this form and we would appreciate it if you would be as accurate as you can. Your answers are being used for research purposes only.

1. Circle the number below which best describes how hard you were working during this flight.

Description of Workload Category	Rating (Circle One)
Workload Low - All Tasks Accomplished Quickly	1 2
Moderate Workload - Chance of Error or Omission is Low	4 5 6
Relatively High Workload Chance of Error or Omission is Relatively High	7 8 9
Very High Workload Not Possible to Perform All Tasks Properly	10 11 12

2. What fraction of time were you busy during the flight?

Seldom Have Much To Do 1 2 3 4 5 6 7 8 9 10 Fully Occupied At All Times

3. How hard did you have to think during this flight?

Activity is Completely Automatic Minimal Thinking and Planning	1	2	3	4	5	6	7	8	9	10	A Great Deal of Thinking, Planning, and Concentration Was Necessary

4. How did you feel during this flight?

The Experience Was Relaxing	1	2	3	4	5	6	7	8	9	10	The Experience Was Very Stressful

Thank you for your accurate answers.