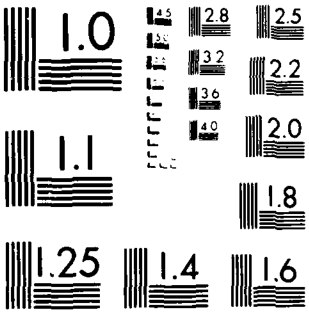
AD-A185 559 STRESS EFFECTS AND THE OPTIMUM FORMAT OF A SCROLLING SEVEN-LINE VDT TEXT WINDOW(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH S G ATKINS 1987 1/1

UNCLASSIFIED AFIT/CI/NR-87-83T F/G 23/2 NL

END
DATE
FILMED
12 87
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DTIC FILE COPY

AD-A185 559

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

(1)

1. REPORT NUMBER AFIT/CI/NR 87-83T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Stress Effects and the Optimum Format of a Scrolling Seven-Line VDT Text Window		5. TYPE OF REPORT & PERIOD COVERED THESIS/~~DISSERTATION~~ 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stephen Grant Atkins		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS AFIT STUDENT AT: Arizona State University		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFIT/NR WPAFB OH 45433-6583		12. REPORT DATE 1987 13. NUMBER OF PAGES 58
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DTIC
SELECTE
NOV 0 4 1987
S D
D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE: IAW AFR 190-1

Lynn Wolaver
LYNN E. WOLAVER 23 Sept 87
Dean for Research and
Professional Development
AFIT/NR

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

ATTACHED

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

Several aerospace applications have been identified for small text windows incorporated into existing computer VDT display screens. These text windows will provide system designers and software engineers with a means of providing real-time interactive plain English instructions on the same VDT screen as graphic or numerical data to which the instructions pertain. In aerospace applications, the screen space available is expected to be very small. For purposes of this report, on-screen windows are presumed to allow for no more than seven lines of text. This report evaluates the effect of time pressure or stress on a key formatting decision which designers must now make concerning these text windows. The NASA Human Factors Laboratory (in conjunction with Lockheed) has run an experiment to determine the most appropriate location (within a seven-line text window) for the current operative instruction (i.e. the current "open check-item). This experiment presents a simplified version of a proposed Shuttle/Space Station VDT screen text window with the current operative line-item at the top, middle, or bottom of the inserted text window. This student engineering report centers around a modification to the original NASA/Lockheed experiment. An additional factor (three levels of time stress) has been applied to the experiment. Appropriate background material is included (in the introduction to this report) to support the student's contention that time stress may have a significant interaction effect on optimal location of the current operative line-item within a seven-line text window. The data from this new modified experiment does, in fact, partially support this contention. It is hoped that when the data from this new experiment is combined with NASA's previous data, selection of an optimum format can be made. With some reservations, the experiment analysis described herein supports placement of the current open check-item at the top of an inserted VDT text window.

STRESS EFFECTS AND THE OPTIMUM FORMAT OF A SCROLLING SEVEN-LINE VDT TEXT WINDOW

by

Stephen Grant Atkins

Engineering Report
submitted in partial fulfillment
of requirements for the degree
Master of Science in Engineering

ARIZONA STATE UNIVERSITY

May 1987

This work is dedicated to the three people who made my efforts possible -- my mother, father, and wife.

Dorese Causey Atkins   Harvey Barnes Atkins   Erica Reneé Atkins

Excerpt from "THE MEASURE OF MAN"

There remains ... the cheerful possibility that we actually know less about the Science of Man than we do of the less difficult sciences of matter and that we may, just in time, learn more. Perhaps Hamlet was nearer right than Pavlov. Perhaps the exclamation "How like a god!" is actually more appropriate than "How like a dog! How like a rat! How like a machine!" Perhaps we have been deluded by the fact that the methods employed for the study of man have been for the most part those originally devised for the study of machines or the study of rats, and are capable, therefore, of detecting and measuring only those characteristics which the three do have in common.

J. W. Krutch, 1954
Indianapolis

ACKNOWLEDGEMENTS

This project would not have been possible without the support of Professors Hewitt Young, James Bailey, Lee Knight, and Mark Edwards, and the marvelous education they have given me in the field of human factors and industrial engineering. Also essential was the enthusiastic support of Dr. Michael Burns and Mr. Ted Kell of the NASA/Johnson Space Center Human Factors Lab. Fellow graduate student Neil Morrison did a great deal of work in support of a joint class project which also utilized data generated by the experiment described herein. The joint effort was related to comparing tendencies for error among people with different levels of relevent computer experience. Neil's advise and efforts made completion of my tasks a great deal more reasonable. Doctoral student John Bouse was of great assistance in checking randomly selected calculations (from among several hundred). The companionship of Neil and John during the period of time of this effort will always be a cherished memory. And, of course, words are not enough to thank my wife, Erica Renee', for her support during this time (so I'll have to think of something else -- hmmmmm).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

alpha	Probability of falsely rejecting the null hypothesis
ANOVA	Analysis of variance
CPPS	Correctly performed procedure step
DF	Degrees of freedom
LSD	Least significant difference
MS	Mean square = sum of squares divided by DF
P-value	Probability of arriving at a test statistic (F-value) greater than that just achieved if the null hypothesis is true (factor is not affecting)
SAS	Statistical analysis software accessed through mainframe batch jobs
s_{ij}	Time stress index
SS	Sum of squares
t,m,b	Top, middle, or bottom of text window
VDT	Computer monitor/visual display tube

Chapter I

INTRODUCTION

The NASA Human Factors Lab at Johnson Space Center is investigating the human performance impacts associated with using VDT's for display of procedural checklists versus the more standard practice of having these checklists displayed on the printed page.. This effort is being funded by the Computer Information Systems Branch of the NASA Space Station Program Office. Therefore, the study is limited, at this time, to evaluation of Space Station (on-orbit) applications of VDT checklist displays. This effort is further limited by it's proposed application to Space Station in that this VDT checklist display is intended to be simultaneously displayed with graphical and alphanumeric data related to the procedure being executed in the checklist. This will limit the screen space available for checklist display. For purposes of this report, screen space is assumed to be limited to seven text lines. The scope of this student's effort is limited to an evaluation of which text lines (from the checklist) should be displayed, given the seven line limitation. Alternative formats considered are: current checklist line-item (also referred to as procedure step) plus the six following procedure steps, or the current procedure step plus the six preceding steps, or the current step plus the three preceding steps and the three following steps. The scope of this student report is specifically limited to analysis of effects (on two measures of human performance) of varying levels of time stress under each of the three formats just described. Since this effort specifically studies checklist procedures for diagnosis of space station hardware failures, the human performance measures used are **number of mistaken diagnoses** and **procedure step completion time.**

Potential users of my results are cautioned to consider the artificial conditions under which this experiment was conducted (see Chapter 1, Subpart A). These non-operational conditions further limit the scope of this effort. The need for caution in attempting to transfer these results to an on-orbit workstation cannot be over-emphasized. The student hopes that an actual workstation, where such an on-screen checklist may be used, will be a substantial improvement (in an ergonomic sense) than the workstation used in this experiment.

The research question at the heart of this engineering report concerns effects of time stress on the optimal format of a seven check-item VDT window proposed for the Space Station Information System. This student project includes proposal, execution, and analysis of a NASA/Johnson Space Center experiment modified by the student so as to allow evaluation of time stress effects. The student hopes to answer these questions: is time stress affecting human performance differently under different formats, and (if one format must be chosen) which format is optimum (i.e. which procedure steps should be in view)?

It is intended that this modification will enhance the cautious applicability of the results to a wider variety of manned aerospace systems, hopefully to include any manned system where cockpit and workstation display space is at a premium.

As previously mentioned, the original NASA/JSC experiment was part of project intended to evaluate the human factors impacts associated with the conversion from hard-copy to VDT display of on-orbit procedures. There are many issues that must be dealt with before NASA can commit to this conversion. This report, however, deals strictly with the configuration of a proposed VDT text window containing seven lines of checklist/procedure (currently used only in hard copy form). It is the student's hope that the results of this evaluation will helpful to control/display designers considering the use of small text screens for future manned systems. One promising application for these small text screens is in real-time (on-board) support of diagnostic analysis (and quick fixes) of in-flight hardware failures. This is exactly the scenario used for the experiment analyzed in this report.

To simplify the presentation of the student's proposed hypothesis and the statistical model used to test this hypothesis, a brief overview of the experimental procedure will be presented first. This is followed by descriptions of the model and hypothesis. Background material is then presented to support the student's hypothesis. The experimental data is then presented, and analyzed, and conclusions drawn. This engineering report concludes with a chapter of future research recommended.

Section I, Subpart A: EXPERIMENT OVERVIEW

The original NASA/JSC experiment is based on a scenario where an anomaly has been detected, presumably a hardware failure within a manned spacecraft. The test subjects were asked to work through a set of written procedures so as to arrive at a diagnosis for the failure. Seven lines of text from these written procedures were visible in the top right-hand corner of a VDT in front of the test subject. Twelve switches were displayed graphically in a single row across the center of the screen (see figure 1). The software (for IBM/PC) was written so that a Microsoft Mouse could move a cursor into each switch graphic and, at the push of the Mouse button, toggle the switch from on to off or vice-versa. The toggling of these switches by the test subject (in response to instructions in the text window) would affect the readings of twelve instrument displays (in two rows of six) across the bottom of the screen. Each subject must regularly check the status of these instrument displays to be able to answer conditional queries contained in the text window (such as: IF INSTRUMENT 6 READS HIGH AND INSTRUMENT 8 READS LOW, THEN TURN SWITCH 6 OFF, ELSE GO TO STEP 7). The subject must carefully follow these instructions in the correct order until the text window exibits a diagnosis step (such as: IF INSTRUMENT 4 IS LOW, THEN THE DIAGNOSIS IS C, ELSE THE DIAGNOSIS IS H). The subject then moves the cursor into a two-digit field beside the word DIAGNOSIS in the upper left-hand corner of the screen and keystrokes the correct diagnosis (hopefully).

Each subject will arrive at a diagnosis six times per set of procedures, for three sets.

In the original experiment design, the condition or independent variable applied to the task was the location (within the text window) of the current "open" check-item. This is the procedural line item (usually an IF...THEN statement) which the test subject is currently considering. The current line item (which was highlighted in reverse video) appeared either at the top, middle, or bottom (t,m,b) or bottom of the text window. Each set of six procedures (each of the six procedures ending in a diagnosis) was presented with the text window in a different configuration of t, m, or b. The order that each VDT window format (t,m,b) was presented was varied for each subject to cancel progressive errors (practice effect and fatique). The subjects were allowed to rest briefly between sets of six procedures, but otherwise had to work continually while within a set.

The data produced by this testing consisted of three sets of completion times (for each series of six procedures) paired with three error totals (integer values from 0 to 6 since each diagnosis was scored as right or wrong). The completion times (for correctly completed procedures) were divided by the associated number of completed procedure steps to yield completion times per correctly performed procedure step (henceforth called time per CPPS). Each data pair (time per CPPS and number of mistakes) was, of course, associated with one particular VDT window format of t, m, or b. Each subject (nine were used) was exposed to each format (one set of each). Thus, the data recorded for each subject might look as follows: Subject 1 -- format b: (3.46 seconds per CPPS, 2 mistaken diagnoses), format t: (2.48 seconds, 2 mistakes), format m: (2.42 seconds, 1 mistake).

The modification used for this engineering report involved the application of time stress as a second independent variable. As a test subject entered into a new set of six procedures, he or she encountered a new VDT window format (t,m, or b) **and** one of three stress level conditions (0,1, or 2). The stress levels were defined as follows:

0 -- Low stress with no time pressure applied. Subjects were asked to work efficiently and continuously, but with emphasis on error-free performance.

1 -- Moderate time pressure applied. Subjects were informed that a rocket propellant (oxidizer in this case) leak was suspected and that completion of all six procedures was required to identify and execute the "fix". Visual indications of declining oxidizer levels were provided (see Figure 2) and an auditory alarm sounded at three minutes into the current set of six (and kept on sounding).

2 -- High stress level. Layered on top of the above scenario was a second leak (fuel) with its own visual fluid level indicator and a second (distinctly different) alarm at four minutes into the set.

The test subjects were informed (for conditions 1 and 2) that failure to finish the entire set of six procedures before all propellant was lost would prevent successful return to earth and imminent death of the crew. The test subjects were, in reality, never informed of such total propellant loss, but were instead asked to continue to completion of all six procedures (with bells and alarms continuing to sound). Twenty-seven test subjects were used, with conditions blocked as follows:

Subjects	VDT Format	Stress Level
1,10,19	b,t,m	2,0,1
2,11,20	t,m,b	0,1,2
3,12,21	m,b,t	0,1,2
4,13,22	b,t,m	0,1,2
5,14,23	t,m,b	1,2,0
6,15,24	m,b,t	"
7,16,25	b,t,m	"
8,17,26	t,m,b	2,0,1
9,18,27	m,b,t	"

Section I, Subpart B: MODEL AND HYPOTHESIS

The design used for this experiment is one intended to facilitate a Two Factor Fixed Effects Analysis of Variance with each factor presented at three levels (3^2 ANOVA). The two factors are best described as **text window format** (presented as top, middle, and bottom of the window) and **scenario time stress** (presented as low, medium, and high stress).

The model used for this experiment is as follows:

$$Y_{ijk} = M + T_i + B_j + (TB)_{ij} + E_{ijk}$$

where Y_{ijk} = the response variable (either completion time per CPPS or number of mistakes), M = the overall mean effect, T_i = the effect of the i^th^ level of the row factor **text window format,** B_j = the effect of the j^th^ level of the column factor **scenario time stress,** $(TB)_{ij}$ = the interaction effect, E_{ijk} = the random error component, $i = 1, 2,$ or 3 (described herein as t, m, or b for top, middle, or bottom position/format for the open check-item within the text window), $j = 0, 1,$ or 2 (for low, medium, or high stress scenario), and $k = 1, 2, \ldots, 9$ (for replicate number).

The null hypothesis (implied by the original NASA experiment) was that time stress was not a factor which warranted laboratory study in the determination of optimum format (t, m, or b). The student's alternative hypothesis is that time stress will have significant interaction effects (interacting with format t, m, or

b) with either of the proposed human performance measures used as the response variable Y_{ijk} (number of mistakes or time per CPPS). Support for the student's contention is contained in Chapter III of this report. However, prior to presentation of the student's argument, background material will be provided to illustrate the types of issues involved in the transition from hard-copy to VDT display.

Chapter II

VDT TEXT DISPLAY VERSUS HARD-COPY

The proposed transition to electronic displays (VDT versus hard-copy) for space operations checklists is sure to generate some controversy. Many of the same concerns voiced by critics of the "electronic office" may have to be re-evaluated in light of Space Command and NASA-specific environments. Physiological or long-term health concerns, such as eye strain and radiation, may not be relevent since space system operators already face VDT's during much of their workday. There is very little general agreement concerning the severity of these physiological effects, and the mechanisms involved are poorly understood. NASA and Space Command will probably be more concerned with the visual quality of the VDT checklists versus the former hard-copy, and with operator performance effects resulting from the conversion. Using a controlled intermediate standard of visual quality, a study done by S. J. Starr (1984) involving 359 office workers indicated that visual quality (of the printed words) on VDT displays were preferred to that of the paper documents they replaced. The results of other studies, however, were mixed. Comments made to the student by potential users of on-screen checklists contradict Starr's results. Some of these same comments reflected concern for loss of the security a user derives from a procedure he can hold in his hands versus screen images which can vanish. Related to this is a concern over the inability to write comments on the checklist for later reference. If this is too difficult to do with on-screen checklists, then important comments by users may be suppressed.

A discussion of all the relative merits of VDT versus hard-copy text manipulation is certainly beyond the scope of this report. However, one area which is relevent to this experiment is that of **reading speed decrements**. Fortunately, there is general agreement on the effects of VDT's on reading speed. Several researchers have demonstrated that reading speeds are 20% to 30% slower using VDT text displays versus hard copy (Muter, et al, 1982; Gould, 1984; Kruk, 1984). After replicating these results, R. S. Kruk (1984) conducted further experimentation to investigate possible causes. Varying the distance between the reader and the screen had no effect on reading speed, and neither did varying the contrast ratio of the video image. Similarly, the time used to fill the screen with text did not account for the difference. VDT reading speed was improved by increasing the number of characters per line (from 39 to 60) and by use of double spacing. These changes still cannot account for the 20% to 30% decrement in VDT reading speed.

Askwall (1985) conducted an experiment in which subjects were asked to search and integrate information found in short blocks of text (22 sentences long). Askwall found search and data retrieval performance to be unaffected by medium (VDT or paper) for sixteen test subjects. However, Askwall did find significant differences in search methods used by her subjects. When reading from paper, test subjects searched almost twice as much text as when searching VDT text. On the other hand, VDT search methods took the subjects twice as long. It is conceivable that search efficiency can be optimized by using VDT versus hard-copy, but with displays formatted so as to focus the operator's attention on pertinent text. Hence, in a properly formatted text window, the slower VDT reading speeds may not necessarily yield a performance decrement in text-line search.

The next section briefly describes stress effects on human performance, primarily dealing with time stress and information transmission to the human subject.

Chapter III

STRESS EFFECTS ON HUMAN PERFORMANCE

As previously stated, the purpose of this experiment is to determine the interaction effect of stress on the optimal location of the current open check-item within a seven-line checklist VDT window. The purpose of the following section of this engineering report is to provide background data supporting the methods used to apply this stress, and to offer support for the student's hypothesis.

Selge (1950) defined stress as an internal response within an organism resulting from external demands made upon the organism. Selge preferred to label the external conditions which made these demands **stressors**. Applying this internal form of stress to man, Chapman (1959) described two types of internal response:

failure stress -- resulting from our inability to perform at the level to which we aspire,

discomfort stress -- resulting from the disparency between what can be done with the remaining time versus what needs to be done in the time remaining. According to Chapman, failure stress can improve our efficiency while discomfort stress leads to slopiness and careless shortcuts.

Siegel (1980) describes the "build-up" of stress involved in the man-machine interface. Stress can build-up in the man-machine environment due to the operator's perception of his own inability to keep up with task demands and from concern over errors he may have recently made. Siegel also suggests that stress results from the need to routinely wait for machine reactions to operator inputs.

Fitts (1967) defines stress not as a feeling or response within the organism, but as a specification of the demands on the operator's time and abilities. Stress, so defined, is therefore an independent variable. Much of the research involving stress as a variable draws a distinction between load stress and speed stress (Conrad, 1951; Wickens, 1984). Load stress refers to the number of channels along which (or sources from which) stimuli may appear. A performance decrement is naturally expected if increasing the number of channels produces an increase in the number of signals an operator must handle per unit time. But Goldstein and Dorfman (1978) have demonstrated that performance decreases even if information transmission rate remains constant. That is, even though an operator can handle one channel that carries Y stimuli per minute, that same operator may not *be able to handle* N channels each carrying Y/N stimuli per minute (Wickens, 1984). Because of the multiple inputs used in this experiment to apply levels of urgency, both load stress and speed stress may be involved. A brief discussion of one type of speed stress (referred to herein as time stress) is included in the next section.

The student's hypothesis to be tested in this experiment is that increasing the level of stress will affect the optimum configuration of the VDT checklist-window. The primary cause of this effect will probably be related to a phenomenon which Sheridan (1981) calls **cognitive tunnel vision** under stress. Research on performance under stress has produced mixed results, including performance enhancements, no change, and decrements. One consistent result has been the observance of **perceptual narrowing** at increasing levels of stress (Easterbrook, 1959; Kahneman, 1973; Hockey, 1970). Coffer and Appley (1964) related this effect to the performance versus stress extrapolation of the Yerkes-Dobson relationship. This mature theory (usually depicted graphically as an inverted "U") implies the existence of an optimal level of stress for various types of tasks. Coffer and Appley suggested that high levels of "stress may interfere with the concentration or flexibility required to deal with the solution of complex problems" (Fitts, 1967). Coffer and Appley added that high stress levels that result in high motivation will improve the performance of subjects engaged in simple tasks.

The stress applied to subjects in this VDT-window experiment is essentially that of **time pressure**, though the mechanisms used (visual and auditory warnings) to simulate this time pressure somewhat muddy the waters. This would be more of a concern if the experiments intent were the study of various types of stress. The purpose here is to re-run the VDT checklist-window experiment under three different levels of stress, each providing a greater sense of urgency. The methods used to apply this stress were chosen for their fidelity with spacecraft operations scenarios (based on the limitations of a non-specialized human factors laboratory). Seperate sections of this report will deal with the specific effects of **time pressure** (applied by the decreasing fluid levels depicted on the visual aids) and **noise** (produced by the auditory warning tones and buzzers).

Section III, Subpart A: TIME PRESSURE

The primary component of the stress applied to the subjects of this experiment is that of time pressure. A sense of urgency is simulated within the experiment by description of an emergency scenario, the progression of which is made evident (to the subject) by visual aides (see figure 2). At pre-determined points in time, auditory warnings will sound in the testing room, providing an additional time cue (for noise/performance effects, see next section). Since the task time allowed is relatively brief (in the test blocks where stress is applied), and the mental activity level for the subject is fast-paced, it is expected that the subjects own time perception will increase the stress level. Ornstein (1969) has demonstrated a phenomenon termed **filled duration illusion** whereby subjects tend to over-estimate the passage of time during periods filled with activity. Ward (1975) demonstrated that subjects tend to over-estimate the duration of brief-intervals of time. It is expected, therefore, that the subjects of this experiment will perceive that they have "used up" more time than has actually passed, and their sense of urgency will be aggravated.

A quantitative measure of time stress has been proposed by Siegel (1980) referred to as the **time stress index** (s_{ij} for task i and operator j). This rating for time stress is defined to be "the ratio of time required to perform the remaining tasks in a series to the time available to do so". Using this definition, a time stress index of 1.0 would indicate that the operator is fully occupied. Research indicates that, depending on consequences of failure and the likelihood of success (or task complexity), subjects will perform more efficiently as they become aroused by increasing s_{ij}, but only to a point. Once this optimal point is reached, performance suffers as s_{ij} increases. Siegel and Wolf (1969) found that a performance decrement often begins at a perceived time stress level of 2.0 to 2.3. Siegel investigated one measure of performance, reaction time (RT), as it related to his time stress index. As s_{ij} increased, RT decreased linearly until a breaking point or stress threshold was reached. At this threshold, RT jumped substantially.

The other component of performance relevent to this VDT experiment is **accuracy**. There has been a great deal of research in the area of speed-accuracy trade-off. This research supports the notion that "speed and accuracy are somewhat incompatible goals" and that subjects "do have some control over the particular mix that they generate through their performance on speeded tasks" (Howell, 1982). The "mix" that Howell refers to represents the degree to which subjects allow themselves to incur more errors while they speed-up their performance. Kreidler and Howell (1964) found that "speed instructions" (i.e. "please emphasize speed on this trial", "now emphasize accuracy") had a large effect on both reaction time and error rate for both simple and complex decision-making tasks. Unfortunately, most time pressure experiments tend to produce a narrow range of error data, which Howell asserts is due to the sample population's common stereotypes related to what constitutes acceptable performance. The "number of mistakes" data for this VDT experiment is probably so affected. Where specific attempts have been made to induce subjects into exhibiting a trade-off of speed for accuracy, the subjects typically make fewer mistakes while working at a normal (subject-specific) self-pace (Howell,1982). Wickens (1984) maintains that the research in this area does not necessarily support such a conclusion. However, lack of time pressure will normally improve accuracy when there is a varying complexity to the required decision response, and where the stimulus remains in view at the subject's whim (Drury and Coury, 1981). Such are the conditions of this VDT window experiment.

In this VDT experiment, the range (between trials) of data for number of mistakes may be suppressed due to other factors as well as those mentioned above. According to Yellott's (1971) Fast-Guess Model, test subjects will select their responses to task items in two different ways: 1) as a result of processed input (given adequate time), and 2) randomly by fast-guessing (with very little input processing). In this model, the numbers of mistakes "are governed largely by the proportion of responses generated by fast-guessing" (Howell, 1982). In the VDT checklist-window experiment, subjects will be observing the current status of one, two, or three instruments and switches and then acting on pursuant instructions in the VDT window. Fast-guessing and avoidance of input-processing should be minimally observed behaviors in this experiment. Compression of variance for the number of mistakes data may result. However, task completion times should still vary significantly, if not for the different VDT window formats, then atleast between the different time pressure conditions.

Time pressure will not effect all elements of the task equally, but will particularly effect those involving greater uncertainty and those depending on short-term memory (McCormick, 1982). The placement of the current open check-item within the VDT text window is expected to effect the level of uncertainty associated with checking-off the current item. Short-term memory comes into play when the subject must recall the action-sequence necessary to end one task and rapidly begin the next. Short-term memory may also come into play when matching up paired "IF...THEN...ELSE" statements when both statements are not simultaneously in view. The student therefore hypothesized that VDT window format and time pressure levels would have significant interaction effects on task (procedure step) completion times, and also, on total mistakes (numbers of incorrect fault diagnoses).

Section III, Subpart B: NOISE STRESS

A great deal of research has been completed related to the effects of noise on human performance. Studies have indicated that task performance requiring serial reactions (where response to one stimulus brings on the next tasking stimulus in an unpredictable fashion) loud noise can increase the incidence of errors or can result in lengthy pauses between responses (Jones, 1983). Further research confirmed these findings, and found these same performance decrements (longer pauses and increased errors) to be independent of test duration and pre-test fatigue level (Hartley, 1973).

Recent research has been directed at the measurement of effects due to moderate noise (less that 85 dB) on tasks which make heavy demands on memory. Findings from this research indicate that moderate noise levels may cause us to make subtle adjustments in information processing strategies. The changes often manifest themselves in one aspect of task performance, but not in others (Jones, 1983). Some results indicate that noise (both intermittent and continuous) can actually improve short-term recall and may lead to a more focused performance. Unfortunately, these contradictory findings make it difficult to predict what effects the warning tones (used as time cues in this student's experiment) had on our subjects. This is especially true since noise effects on performance have also been found to be dependent on the subject's attitude about the noise (Glass and Singer, 1972).

It was the students intention that the warning tones would uniformly increase the amount of time pressure or anxiety felt by the subject during the higher stress scenarios. Unfortunately, the effects of noise have been shown to continue after noise offset, so this desired uniform increase in stress level (if it occurred at all) may not have been limited to those portions of the experiment for which it was intended (Hartley, 1973). With noise levels from the alarms measured at the site of the subjects' ears at 65 dB (A scale) continuous, (overlayed with total combined peaks at 68 dB), noise stress, in and of itself, may not have been a factor in my experiment.

Chapter IV

EXPERIMENT RESULTS

The modified experiment was conducted during the period of 30 October to 20 November 1986. The NASA software which produced the graphics, scrolling text window, and mouse logic, also included sub-routines which generated a very detailed printout of each test subjects performance (see Figure 3). The software produced twenty-one such pages for each subject (eighteen procedures plus three practice procedures). These printouts included many details not used by this experimenter, such as event times for the beginning and end of each movement of the mouse/cursor, as well as the (x,y) coordinates for the endpoints of such movements. This detailed data output capability was intended to support other research efforts (see Chapter VII).

For purposes of this experiment, the event times were used to measure task completion times. Since the screen reconfiguration subroutine (applied between each procedure) artificially created exaggerated time intervals at the beginning (and sometimes the end) of each page of printout, the first and last recorded times were ignored. For each combination of conditions and for each subject, the completion time per correctly performed procedure step (time per CPPS) was calculated by subtracting the second time hack from the second to the last time hack, and summing these time intervals divided by the number of correctly performed procedure steps included. The time hacks (endpoint event times) were first converted to decimal minutes to support these calculations, then back to seconds. This procedure was then repeated for each of the four hundred and eighty-six procedures which were completed (twenty-seven subjects times eighteen procedures each). The results of these calculations are shown in Table 1.

The printouts also provided the diagnoses offered by each test subject at the end of each of the eighteen procedures (three sets of six). These answers were compared to the NASA-provided answer sheet for grading. Three scores were thus produced for each subject (one for each set of six procedures where each set of six was performed under a different combination of conditions; e.g. top format plus moderate stress, or bottom format plus high stress). Each of these scores represented the number of incorrect diagnoses per set of six and were, therefore, integer values between zero and six. Since each subject, in a sense, performed (and yielded scores) in three different experiment runs (each under different conditions), there were actually eighty-one scores (and times) or nine data pairs (time, # of mistakes) within each of the nine cells of this experiment design. The results of the scoring calculations (number of mistakes) are shown in Table 2.

Stress Levels

Low			Medium			High			
2.48	4.60	3.12	3.80	2.86	2.81	1.96	3.40	2.93	TOP
3.83	5.33	3.05	4.40	3.39	3.25	2.63	3.10	2.55	
3.25	4.44	3.55	3.73	6.16	3.53	2.47	3.83	3.61	
3.13	4.25	3.22	2.42	3.50	4.33	3.40	4.45	5.94	MID
5.92	2.94	4.50	3.14	4.10	3.77	3.03	2.25	3.17	
7.83	5.17	5.00	2.37	3.23	3.33	2.90	6.44	3.50	
4.66	3.40	2.76	2.38	5.00	2.92	3.46	3.08	3.24	BOTTOM
5.44	2.73	4.25	4.20	3.83	2.94	4.25	3.03	3.21	
4.72	5.28	7.17	3.83	4.17	3.17	3.83	3.40	2.73	

Table 1.

TIME PER CORRECTLY PERFORMED PROCEDURE STEP

(in seconds)

	NUMBER OF MISTAKES PER CELL				NUMBER OF MISTAKES PER PERSON PER SET OF 6		
FORMAT	Low	Medium	High	STRESS	Low	Medium	High
Top	***** ****	***** ***	***** *****		2 2 0 3 1 0 1 0 0	1 3 0 0 0 0 1 3 0	0 3 0 1 0 1 0 3 2
Mid	*****	***** ****	***** ***** ***** ****		0 0 1 1 0 1 0 2 0	1 2 1 0 0 1 1 0 3	0 4 2 0 0 4 1 4 4
Bott	***** ***** ***** **	***** ***** ****	***** **		0 2 1 1 1 3 0 5 4	3 2 0 0 2 0 1 4 2	2 2 0 0 1 1 1 0 0

Table 2

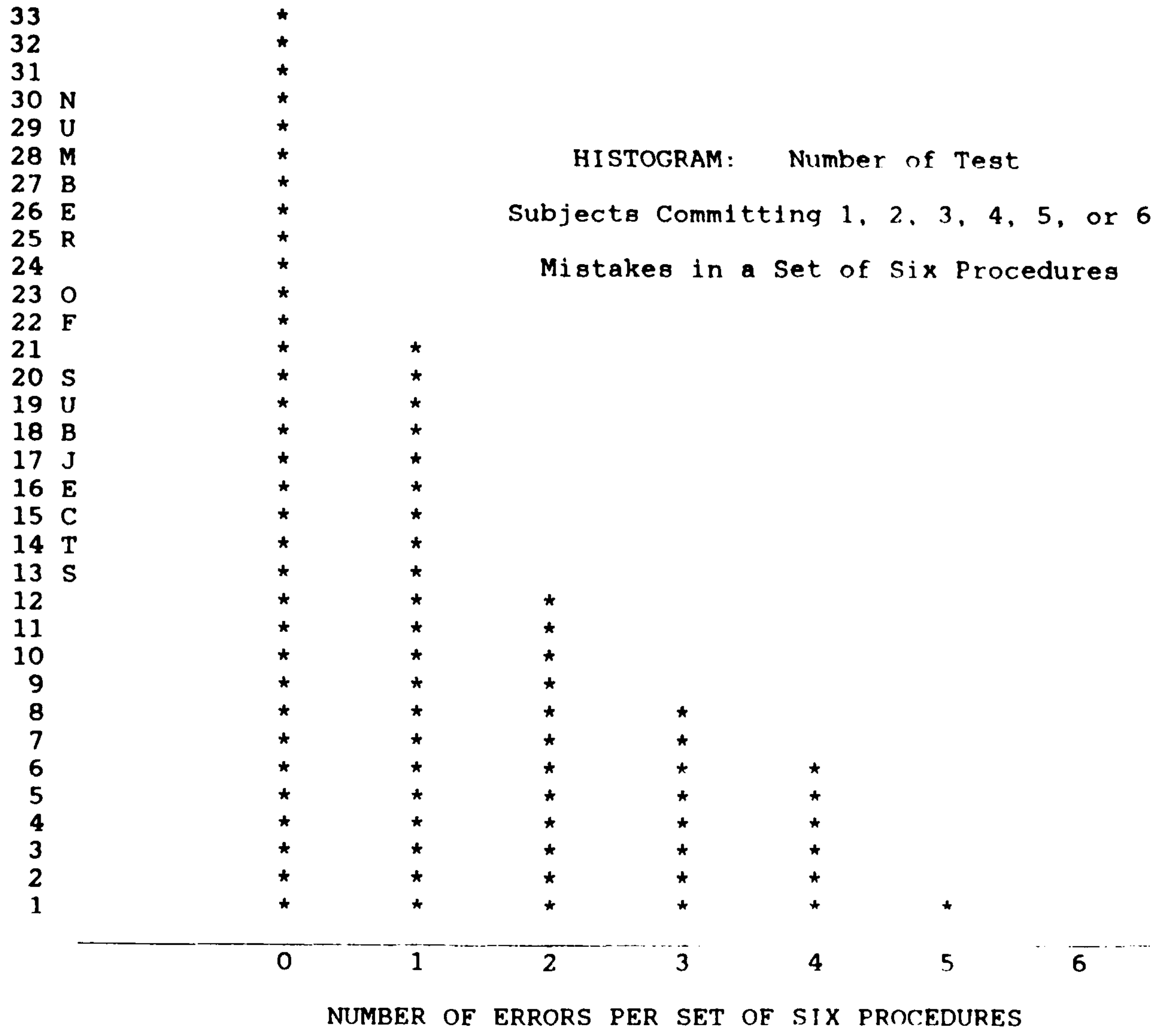
RAW DATA FOR MISTAKES

Chapter V

DATA ANALYSIS

As described in Section 1, Subpart B, the experiment was designed to support a Two Factor (each at three levels) Fixed Effects Analysis of Variance. This 3^2 ANOVA procedure was applied first using number of subject errors (= number of incorrect diagnoses per set of six) as the response variable. The 3^2 ANOVA procedure was then repeated using time per CPPS as the dependent variable.

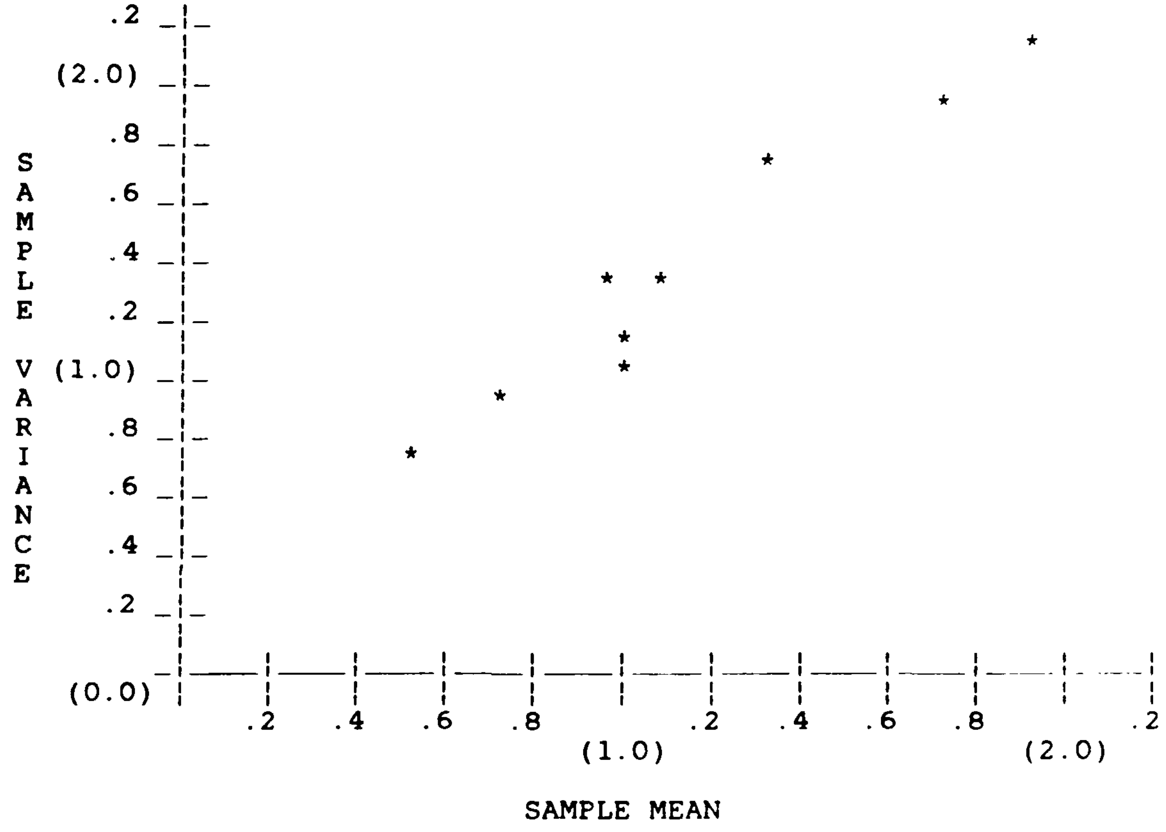
Figure 3 shows a histogram plot of the error (or incorrect diagnosis) data. Figure 4 is a plot of the variance of each of the nine cells (where each cell has nine replicates) versus the cell means. Linear regression values were calculated to evaluate the relationship between mean and variance, as suggested by Bartlett (1947). Since the square root transformation is appropriate when the mean and variance are linearly correlated and the data is of the discrete counting type (Montgomery, 1984), the square root transformation was applied. This technique was supported by the resulting coefficient of determination (r^2) which revealed that 91% of the variation in the value of the cell variances was due to changes in the cell means. Since the original data was composed of small integer values with frequent zeros, it was suggested (Bartlett, 1947) that adding a 1 to each data element before taking the square-root would control the effect of the zeros on the data analysis.



HISTOGRAM: Number of Test Subjects Committing 1, 2, 3, 4, 5, or 6 Mistakes in a Set of Six Procedures

Figure 3

Graph of Sample Mean vs. Sample Variance



Linear Regression Components:

$b = .39 \quad m = .72$

$y = (.72)x + .39$

$r = .96$

$r^2 = .91$

FIGURE 4

SAS 3^2 ANOVA mainframe software routines were then run against the transformed incorrect diagnosis counts (see Table 4). Residuals were also plotted across fitted values (see Figure 5) and across treatment levels for both treatment factors (see Figures 6 and 7). These plots were produced to support conclusions based on equality of variance.

The above procedures were repeated on the task completion time data (time per CPPS). The results are shown in Figures 8, 9, 10 and 11 and Tables 5 and 6. The relationship between mean and variance for the time data is less defined, though still worrisome. Also of concern is the funnel effect in Figure 9 where residuals are plotted vs. fitted values. Application of a transform intended for integer counting data may be questionable, but should be considered. Equality of variance is certainly in doubt for the completion time per CPPS data.

Conclusions to be drawn from the results of this data analysis will be presented in the next two sections.

Table of Sample Mean and Sample Variance for Original Data Cells

(x, s^2)	(1,1.12)	(.89,1.27)	(1.11,1.27)
	(.56,.73)	(1,1)	(2.11,1.9)
	(1.89,1.76)	(1.56,1.43)	(.78,.83)

Table 3

MEAN NUMBER OF MISTAKES BY CELL (with associated variance)

SOURCE	DF	SS	MS	Fo	PR > F
model	8	1.7305	.2163	1.19	.3178
format	2	.2336	.1168	.64	.5292
stress	2	.0396	.0198	.11	.8971
inter.	4	1.4573	.3643	2.00	.1032
error	72	13.0990	.1819		
total	80	14.8295			

Table 4

ANOVA RESULTS FROM SAS

FOR NUMBER OF MISTAKEN DIAGNOSES

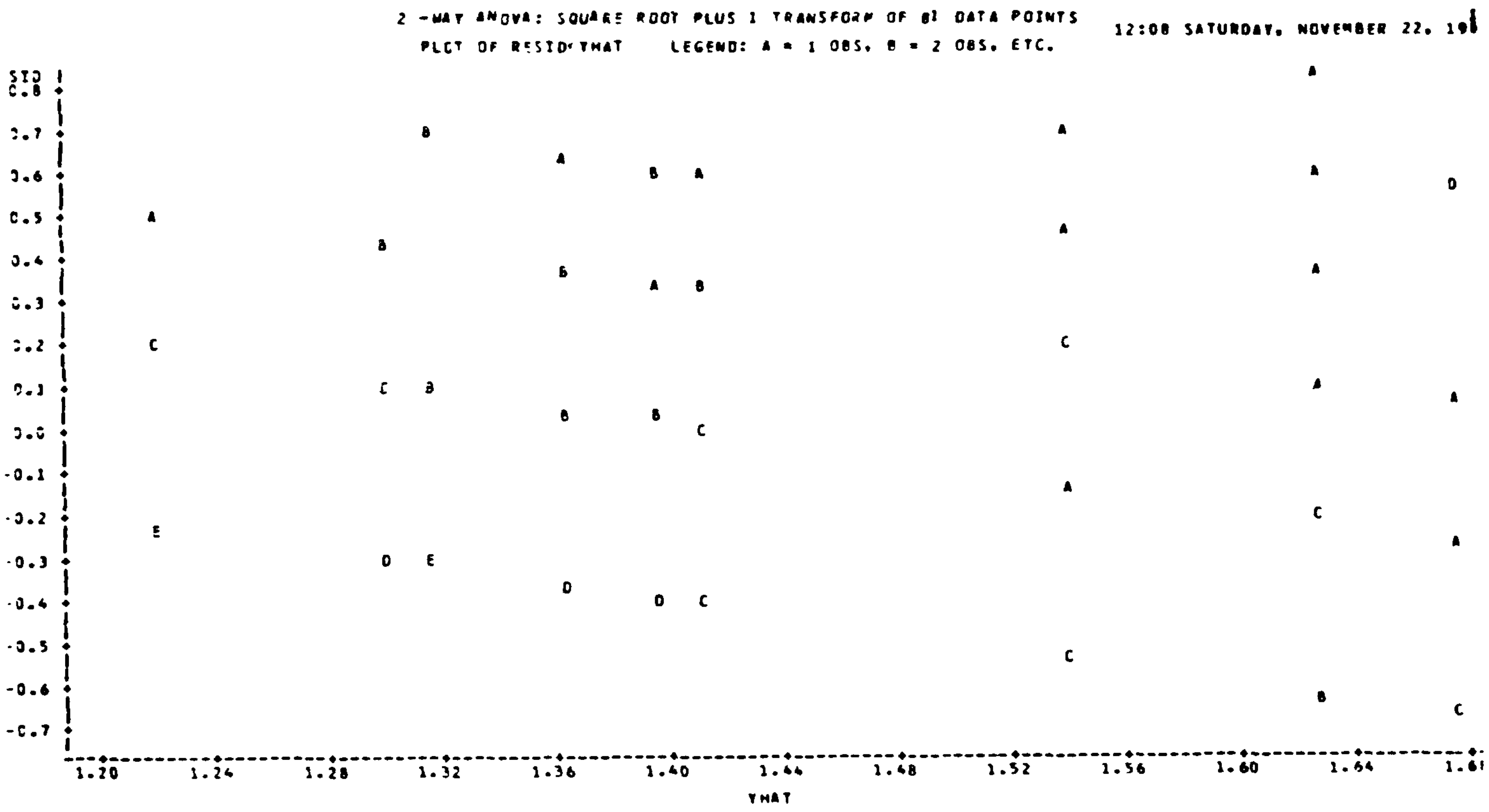
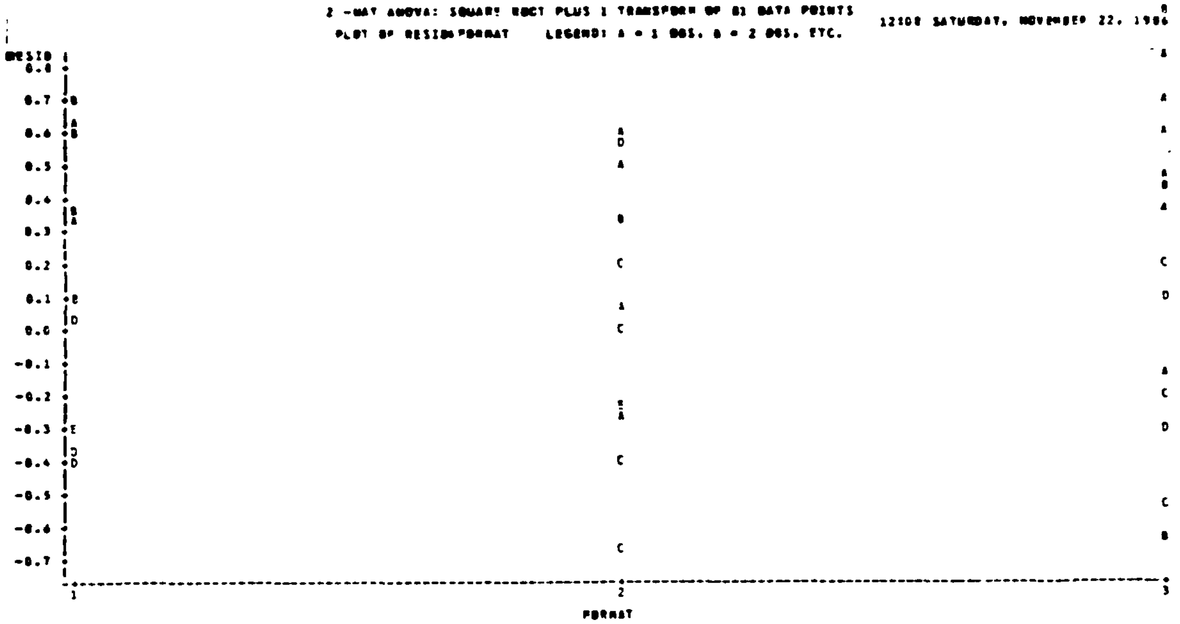


Figure 5 RESIDUALS VERSUS FITTED VALUES

(response variable is number of mistakes)

Figure 6 RESIDUALS VS FORMATS (where 1,2,3 = t,m,b)

(response variable is number of mistakes)



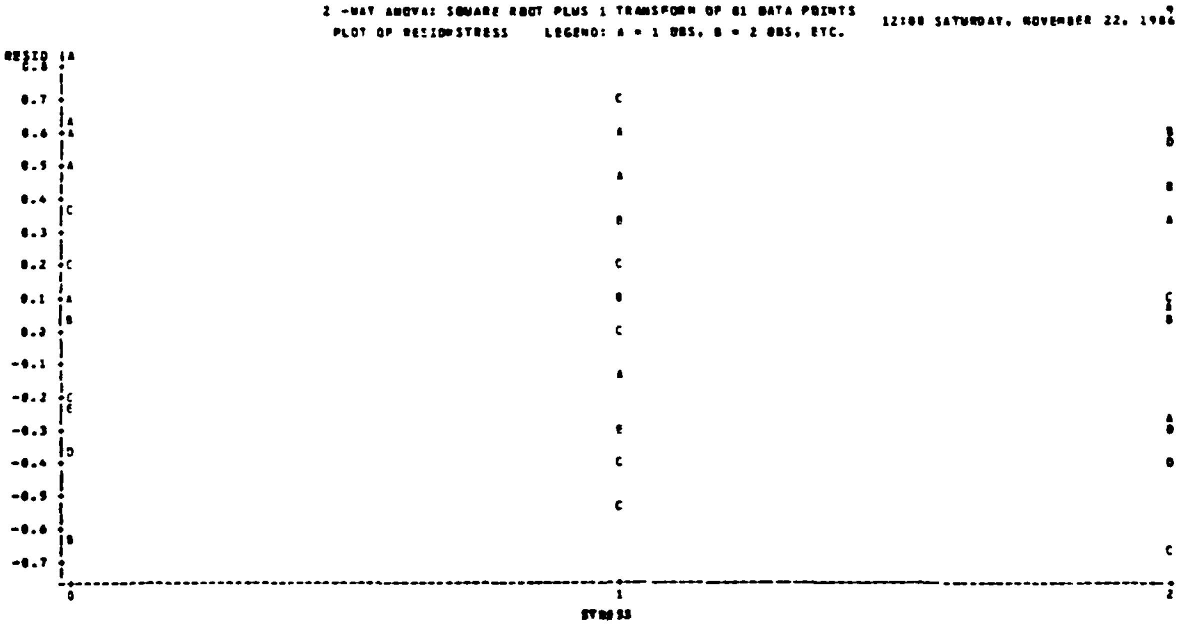


Figure 7 RESIDUALS VS STRESS LEVELS (where 0,1,2 = low, moderate, high)

Stress Levels						Formats
Low		Medium		High		
3.74	.811	3.77	1.042	2.94	.363	T
4.66	2.447	3.35	.451	3.23	2.23	M
4.49	2.016	3.60	.665	3.36	.206	B

Table 5

MEAN AND VARIANCE OF TIME PER CPPS DATA
(with nine replicates per cell)

(mean times per CPPS are in seconds throughout this report)

SOURCE	DF	SS	MS	Fo	PR > F
model	8	21.4624	2.6827	2.40	0.0233
format	2	3.3584	1.6792	1.50	0.2290
stress	2	12.2041	6.1021	5.47	0.0062
inter.	4	5.8997	1.4749	1.32	0.2701
error	72	80.3484	1.1159		
total	80	101.8108			

Table 6

ANOVA OUTPUT FOR TIME PER CPPS

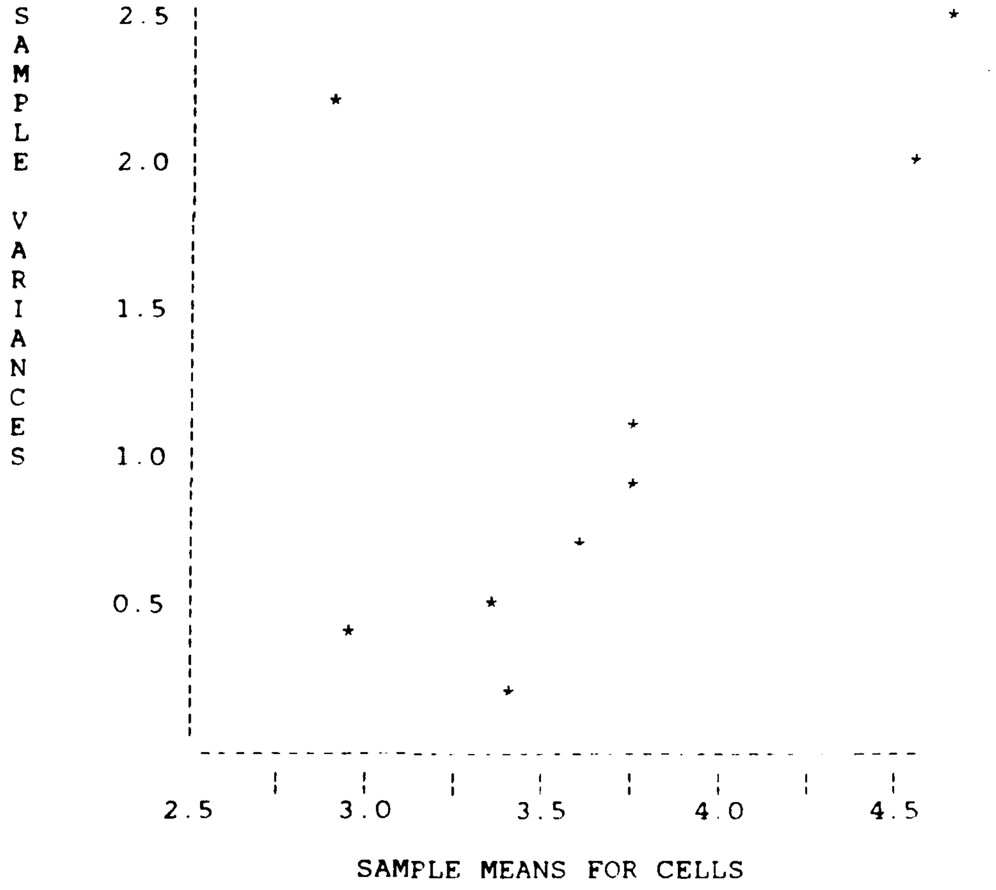


Figure 8

GRAPH OF CELL MEANS VS. CELL VARIANCES FOR TIME PER CPPS DATA

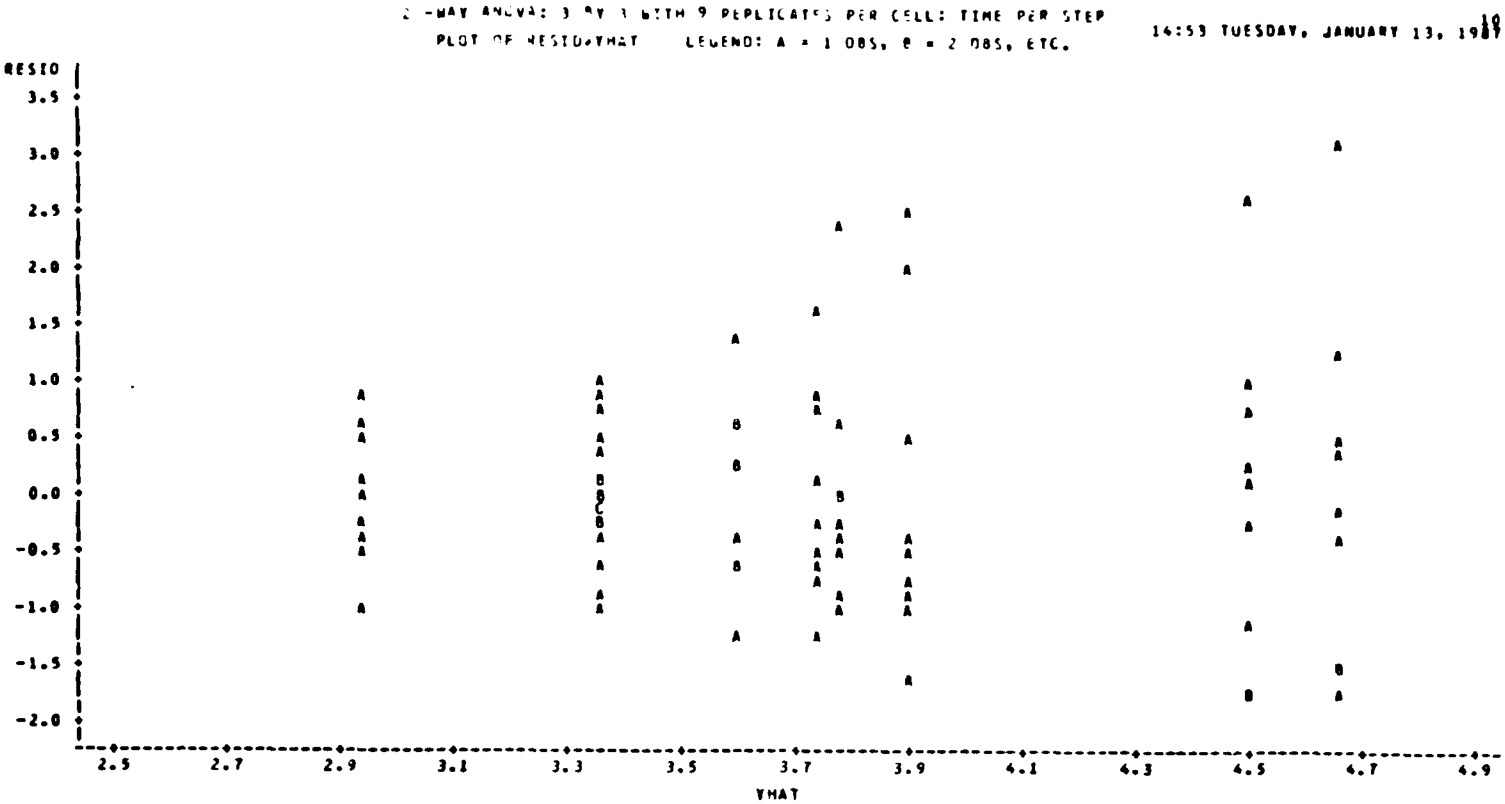
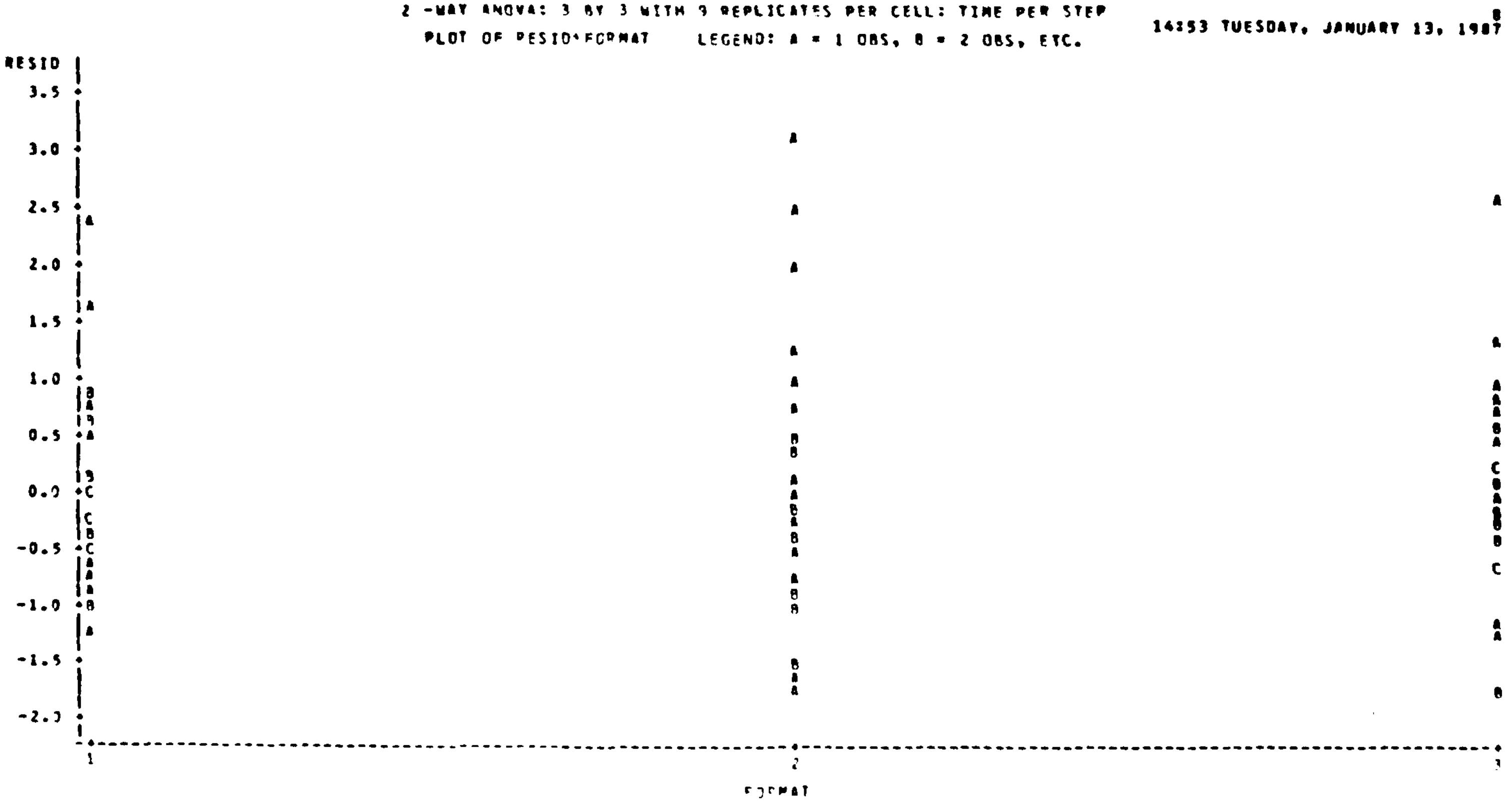


Figure 9

RESIDUALS VERSUS FITTED VALUES

(response variable is time per CPPS in sec)



```
2 -WAY ANOVA: 3 BY 3 WITH 9 REPLICATES PER CELL: TIME PER STEP          8
                                                   14:53 TUESDAY, JANUARY 13, 1987
PLOT OF RESID*FORMAT    LEGEND: A = 1 OBS, B = 2 OBS, ETC.
```

Figure 10   RESIDUALS VERSUS FORMATS (where 1,2,3 = t,m,b)

```
2 -WAY ANOVA: 3 BY 3 WITH 9 REPLICATES PER CELL: TIME PER STEP                                9
  PLOT OF RESID*STRESS     LEGEND: A = 1 OBS, B = 2 OBS, ETC.          14:53 TUESDAY, JANUARY 13, 1987
```

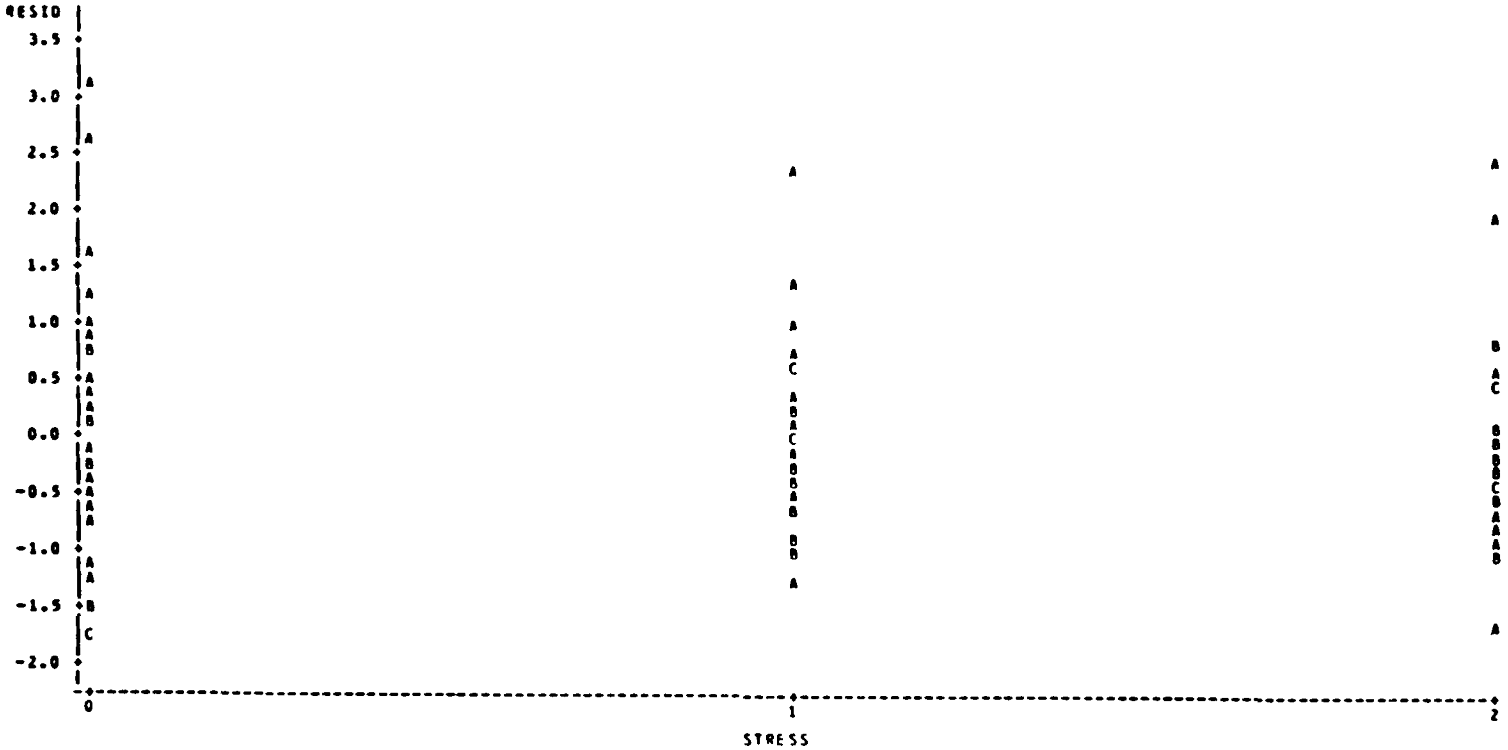


Figure 11

RESIDUALS VERSUS STRESS LEVELS (where 0,1,2 - low, moderate, high)

Chapter VI

PAIRWISE COMPARISONS FOR OPTIMUM FORMAT

The alternative hypothesis at the heart of the student's original research proposal (precipitating this engineering report) was that format and stress levels would have significant interaction effects on human subject performance (mistakes and completion times). The validity (or lack thereof) as regards this student hypothesis can be determined via the analysis of variance already described. However, selecting the optimum format (t, m, or b) is the purpose of the NASA/JSC Human Factors Lab effort (which this report is intended to support). Therefore, pairwise comparisons across formats (t, m, or b) will now be made for both completion times and mistakes.

SAS statistical software was used to perform Duncan's Multiple Range tests (as well as the ANOVAs). No significant differences were found between the average completion times (time per CPPS) or average number of mistakes associated with each format (these tests were performed at alpha = 0.1, the highest value possible with SAS).

These same pairwise comparisons were repeated using the Least Significant Difference (LSD) method (at alpha = 0.2). This method compares the difference between each treatment mean (the average performance values for each format t, m, or b) with the interval LSD, where:

$$\mathrm{LSD} = t_{\mathrm{alpha}/2, N-a}[2MS_{\mathrm{error}}/n]^{1/2}$$

and N = 81 (sample size), a = 3 (treatments or t, m, and b), and n = 27 (number of performance scores -- mistakes or CPPS times --associated with each treatment factor top, middle, or bottom). For the analysis of number of mistakes (incorrect diagnoses), LSD = 0.150. For the analysis of completion time per CPPS, LSD = 0.371.

Regarding the average number of mistakes, the LSD method also shows no significant differences due to format. However for completion time per CPPS, having the current open check-item at the top of the window (format = t of t, m, or b) results in a significant improvement. The relationship between the format treatments and their associated average times per CPPS are as follows:

Middle	Bottom	Top
3.972 seconds	3.818 seconds	3.484 seconds
************	************	
	*************	*************

where differences spanned by the same bar of asterisks are not statistically significant.

Chapter VII

CONCLUSIONS

There are several conclusions that can be cautiously drawn from the aforementioned results. First off, although one would normally expect a trade-off between completion time and number of mistakes, one format (the top format) produced both the least number of mistakes and the fastest completion times. However, the only statistical significance that can be associated with the tests for number of mistakes is the interaction term (found to be a significant cause of variability with a P-value equal to 0.103). Therefore, the null hypothesis is rejected in favor of the student's alternative hypothesis as regards interaction effects between format and stress and number of mistakes. Interaction was not statistically significant where completion time per correctly performed procedure step is the response variable, and the student's second alternative hypothesis remains unproven. Format (t, m, or b) alone was minimally significant as a factor affecting completion times per correctly performed procedure step (P-value equals 0.229). Format alone was insignificant as a factor affecting number of mistaken diagnoses (P-value equals 0.5292).

Curiously, the stress levels were negligible in their own right as a factor affecting number of mistaken diagnoses (P-value equals 0.897). However, these same stress levels were very significant in their effect on completion times (P-value equals 0.0062). The medium stress level caused faster task performance than the low stress level, and the high stress level completion times were quicker still.

The purpose of the NASA/JSC effort (from which this student effort was derived) was the selection of an optimum VDT text window format. With some reservation (due to poor ANOVA and LSD significance levels), the top format is recommended. This is despite the fact that twenty-one of the twenty-seven subjects said they preferred the middle format (with three preferences each for the other two formats). Having the current open check-item at the top of the window produced consistently (and reasonably) low error rates across all three stress levels. Furthermore, completion times were somewhat better in the top format.

Chapter VIII

FUTURE RESEARCH RECOMMENDED

Future research possibilities related to this engineering report fall into two catagories: first, those which put to further use the existing data or the existing experiment software without modification, and secondly, experiments which are logical extensions of this effort but may involve substantial research costs and time-consuming data collection.

In the first catagory, the existing 567 pages of printout can be used to study human response times and cursor control capabilities and the way these capabilities are affected by target locations on a computer screen. For instance, how is human response time and speed of cursor travel affected if the path between current cursor location is vertical, horizontal, or angular. What if the path intersects graphic images or text? Since the printouts provide (x,y) coordinates and event times for each movement of the mouse, such a study may be feasible with the existing data.

Another possibility is to place the experiment station on a force platform and repeat a smaller number of trials measuring for fidgetiness (as measured in a previous ASU force platform study by H. H. Young). The intent here would be to correlate stress level indicated by fidgetiness to applied stress level in the modified VDT text window experiment. Such a study might allow for differentiating between time stress, noise stress, and just the low level stress associated with being under observation while performing tasks.

A third research area is made possible by the further use of subject questionnaire data (already collected to support a joint class effort described in this report's Acknowledgement section). These questionnaires collected data from all 27 subjects concerning relevant factors in their personal make-up or background that might affect their performance in this experiment. These personal factors were not of particular interest to the experimenter, but might be expected to add to the experimental error. This student intends to pursue the study of using such questionnaire data to produce a stepwise multiple regression model for predicting subject performance. The model will use parameters which are of no interest to the experimenter and which cannot be controlled by the experimenter, but are all the same expected to affect subject performance (personal factors such as hours of sleep last night, hours since last meal, hours of mouse experience, gender, age, etc). The model will predict for each subject three data pairs (mistakes and time per CPPS for each of 3 sets of procedures) which is then compared to the grand mean for the entire sample population. This difference is then used to adjust the subject's **real** data, and an entirely new data set is produced (hopefully, with variability due to personal factors reduced).

In the other catagory of proposed research (experiments which may require substantial modifications or costs), three possibilities will be very briefly presented.

If it is not economically feasible to repeat this experiment on-orbit (without which the translation of the data herein is certainly questionable), then the experiment should be modified to test affects due to unusual angles. In weightlessness, the viewers orientation to the VDT screen will often be very different than that experienced in this study. The mouse/cursor movement data from such an experiment may be of even more interest. Such a modification should include the sort of left forearm strap-on palette (with caged mouse) designed for spaceflight.

This student's experiment was limited to the top, middle, and bottom locations in the text window. The software could certainly by modified to allow testing of other locations (second to the top, third from the bottom, etc). Such a modification should be considered.

In further support of the NASA effort to understand the trade-offs between hard-copy paper procedures and on-screen procedures, this experiment should be repeated using a blank text window with the same procedures instead contained in a three-ring binder, with all other conditions identical.

REFERENCES

1. Adiga, S. and Young, H. H., "The Human Force Platform for Measuring Mental Stress in Seated Operations", IIE Proceedings, 1985, pp 204-211.

2. Askwall, S., "Computer supported reading vs reading text on paper: a comparison of two reading situations", International Journal of Man-Machine Studies, Vol. 22, 1985, pp 425-439.

3. Bartlett, M., "The Use of Transformations", Biometrics, Vol. 3, 1947, pp 39-52.

4. Beaumont, J. G., "Speed of response using keyboard and screen-based microcomputer response media", International Journal of Man-Machine Studies, Vol. 23, 1985, pp 61-70.

5. Beringer, D., "Underlying behavioral parameters of the operation of touch-input devices: biases, models, and feedback", Human Factors, Vol. 27, 1985, pp 445-458.

6. DeVore, J., Probability and Statistics for Engineering and the Sciences, Brooks/Cole Publishing, 1982, 639 pp.

7. Drury, C., "The effect of speed of working on industrial inspection accuracy", Applied Ergonomics, Vol. 4.1, 1973, pp 2-7.

8. Fitts, P. and Posner, M., Human Performance, Brooks/Cole Publishing, 1967, 162 pp.

9. Gould, J. D. and Grischkowsky, N., "Doing the same work with hard copy and with CRT computer terminals", Human Factors, Vol. 26, 1984, pp 323-337.

10. Howell, W. and Fleishman, E., Human Performance and Productivity, Vol. 2 "Information Processing and Decision Making", Lawrence Erlbaum Associates, Publishers, 1982, 193 pp.

11. Hockey, R., Stress and Fatigue in Human Performance, John Wiley and Sons, 1983, 396 pp.

12. Huchingson, R. D., New Horizons for Human Factors in Design, McGraw-Hill, 1981.

13. Kruk, R. and Muter, P., "Reading of continuous text on video screens", Human Factors, 1984, pp 339-345.

14. McCormick, E. and Sanders, M., Human Factors in Engineering and Design, McGraw-Hill, 1982, 653 pp.

15. Montgomery, D., Design and Analysis of Experiment, Wiley and Sons, 1984, 538 pp.

16. Rouse, S. and Rouse, W., "Computer-based manuals for procedural information", IEEE Transactions on Systems, Man, and Cybernetics, Vol. smc10, 1980, pp. 506-510.

17. Rouse, S. et al, "Design and evaluation of an onboard computer-based information system for aircraft", IEEE Transactions on Systems, Man, and Cybernetics, Vol. smc12, 1982, pp. 451-463.

18. Schwartz, D. and Howell, W., "Optional stopping performance under graphic and numeric CRT formatting", Human Factors, Vol. 27, 1985, pp. 433-444.

19. Starr, S., "Effects of video display terminals in a business office", Human Factors, Vol. 26, pp. 347-356.

20. Wickens, C. D., Engineering Psychology and Human Performance, Merrill Publishing Company, 1984, 513 pp.

21. Young, H. H., "Human performance effectiveness modeling and enhancement using computer-analyzed survey data", IIE Annual Conference Proceedings, 1982, pp. 348-358.

ATTACHMENT

```
Run Number  3

  Time      Delta         Mouse   Switch ONON  Kb Exp  Scroll  Got

Starting procedure 99. It has  15 lines.

36:41.13   0: 3.35     ( 8, 3, 2)                           0
37:17.28   0:36.15   * (11,20, 2)                           0
37:17.66   0: 0.38     (11,20, 2)  3 0101                 U 2
37:22.22   0: 4.56   * (11,43, 2)                           2
37:22.60   0: 0.38     (11,43, 2)  7 0101                 U 3
37:29.80   0: 7.20   * ( 7,31, 2)                           3
37:29.91   0: 0.11     ( 7,31, 2)                         U 4
37:41.33   0:11.42   * ( 7,31, 2)                           4
37:41.44   0: 0.11     ( 7,31, 2)                         U 5
37:46.22   0: 4.78   * ( 7,31, 2)                           5
37:46.33   0: 0.11     ( 7,31, 2)                         U 6
37:47.48   0: 1.15   * ( 7,31, 2)                           6
37:47.65   0: 0.17     ( 7,31, 2)                         U 7
37:48. 3   0: 0.38   * ( 7,31, 2)                           7
37:48.20   0: 0.17     ( 7,31, 2)                         U 8
37:48.64   0: 0.44   * ( 7,31, 2)                           8
37:48.80   0: 0.16     ( 7,31, 2)                         U 9
37:49.24   0: 0.44   * ( 7,31, 2)                           9
37:49.35   0: 0.11     ( 7,31, 2)                         U 10
37:53.86   0: 4.51   * (11,13, 2)                          10
37:54.29   0: 0.43     (11,13, 2)  2 0112                 U 11
37:59.46   0: 5.17   * (11,48, 2)                          11
37:59.90   0: 0.44     (11,48, 2)  8 1001                 U 12
38: 4.51   0: 4.61   * (11,25, 2)                          12
38: 4.95   0: 0.44     (11,25, 2)  4 1020                 U 13
38:10.72   0: 5.77   * (11,48, 2)                          13
38:11.10   0: 0.38     (11,48, 2)  8 0112                 U 14
38:17.75   0: 6.65   * ( 7,47, 2)                          14
38:17.97   0: 0.22     ( 7,47, 2)                         U 15
38:24.17   0: 6.20   * ( 6,12, 2)                          15
38:24.34   0: 0.17     ( 6,12, 2)                          15
38:25. 5   0: 0.71     ( 6,12, 2)              b   *       15
38:26.81   0: 1.76     ( 6,12, 2)              CR  *       15
38:26.81   0: 0. 0     ( 6,12, 2)                          15
38:29.34   0: 2.53   * ( 8, 2, 2)                          15

Procedure 99 terminates at line 15 with diagnosis of b
```

LMED

8