CHAPTER FIVE
A PILOT STUDY TO MEASURE CORRELATES OF INDIVIDUAL DIFFERENCES IN NONDESTRUCTIVE INSPECTION PERFORMANCE

5.0 INTRODUCTION

An interim report examined previous research programs and studies in the area of nondestructive inspection (NDI) conducted by the Air Force, by the nuclear power industry, and by various academic and industry investigators (FAA/AAM & GSC, in press). A repeated finding and concern in these studies was the existence of substantial differences among inspectors' NDI proficiency. Those few studies that have attempted to examine correlates of NDI proficiency have been generally unsuccessful in establishing significant relationships with performance.

As noted in the interim report, however, there are a number of variables, measures of which appear potentially relevant to NDI selection and/or proficiency. This conclusion is based on the findings of studies of individual differences in the areas of inspection and vigilance, on the opinions of experts in the NDI field, and on the results of interviews with NDI inspectors and training supervisors. These variables can be roughly separated into the following categories:

- Boredom Susceptibility
- Concentration/Attentiveness/Distractibility
- Extraversion/Impulsivity
- Motivation/ Perseverance
- Decision Making/Judgement
- Mechanical/Electronics Aptitude
- Need for Autonomy.

Each of these categories and the measures used for them is discussed in greater detail under Section 5.1, Methodology.

The pilot study reported here represents the initial phase of an effort to investigate relationships of these variables to NDI performance. A second aspect of this effort is to investigate possible task-induced fatigue resulting from sustained performance and to examine possible interactions between performance changes and the above subject variables. The task employed in this study was a computer-simulated NDI eddy-current task developed by Dr. Colin G. Drury and his colleagues at the State University of New York (SUNY) at Buffalo. The task has been described in studies by Drury, Prabhu, Gramopadhye, and Latorella (1991) and Latorella, Gramopadhye, Prabhu, Drury, Smith, and Shanahan, (1992). In essence, the task utilized a SUN SPARC workstation and incorporated a standard keyboard and optical three-button mouse as input devices. As Latorella et al. (1992) have emphasized, the aim in developing this task was neither to develop a simulator for training on actual NDI tasks nor to develop a task to measure absolute values of the probability of detecting particular types and sizes of faults. Rather, their aim was to devise a task that closely approximated the characteristics and requirements of eddy-current inspection tasks to enable laboratory investigation of factors that may influence NDI performance.

Of the two previous studies that have used this task, neither has extensively evaluated possible predictor measures or possible fatigue effects resulting from sustained performance over successive task sessions. As noted above, the primary intent of this research program, of which the pilot study reported here is only the initial phase, is to investigate both the relationship of individual difference variables to NDI performance and the possible effects of sustained task performance on NDI efficiency .

The following sections more fully describe the criterion task, the tests and measures used, the procedures employed, and the results obtained. Preliminary observations and directions to be taken in the primary study are also discussed.

5.1 METHODOLOGY

5.1.1 Subjects

A total of 6 subjects, 3 males and 3 females, participated in the pilot study. All were right-handed, had normal near visual acuity (as determined from an Orthorater screening test), reported normal hearing, and were between 18 to 29 years of age. None had prior training or experience in aircraft maintenance or inspection. All had completed at least 12 years of school. Subjects were obtained through an existing Federal Aviation Administration (FAA) subject contract and were paid $10.00 an hour for their participation. Summary characteristics of the subjects are given in Table 5.1.

Table 5.1 Characteristics of Subjects

Subject Number	Age	Gender	Occupation	Education Level
1	21	male	roofer	12 years
2	21	female	medical aid	14 years
3	18	male	student	13 years
4	22	male	janitor	13 years
5	22	female	student	15 years
6	29	female	store manager	13 years

Table 5.1 Characteristics of Subjects

5.1.2 Apparatus

As indicated earlier, the basic apparatus consisted of a SUN SPARC Model 4/50GX-16-P43 workstation, a 19 inch color monitor, and a 3-button optical mouse. Since the nature of the task and its physical characteristics have been described in detail previously (Drury et al., 1991; Latorella et al., 1992), only characteristics relevant to the present study are reviewed here.

The display itself consisted of four basic task elements (windows). These are shown in Figure 5.1 and are described below.

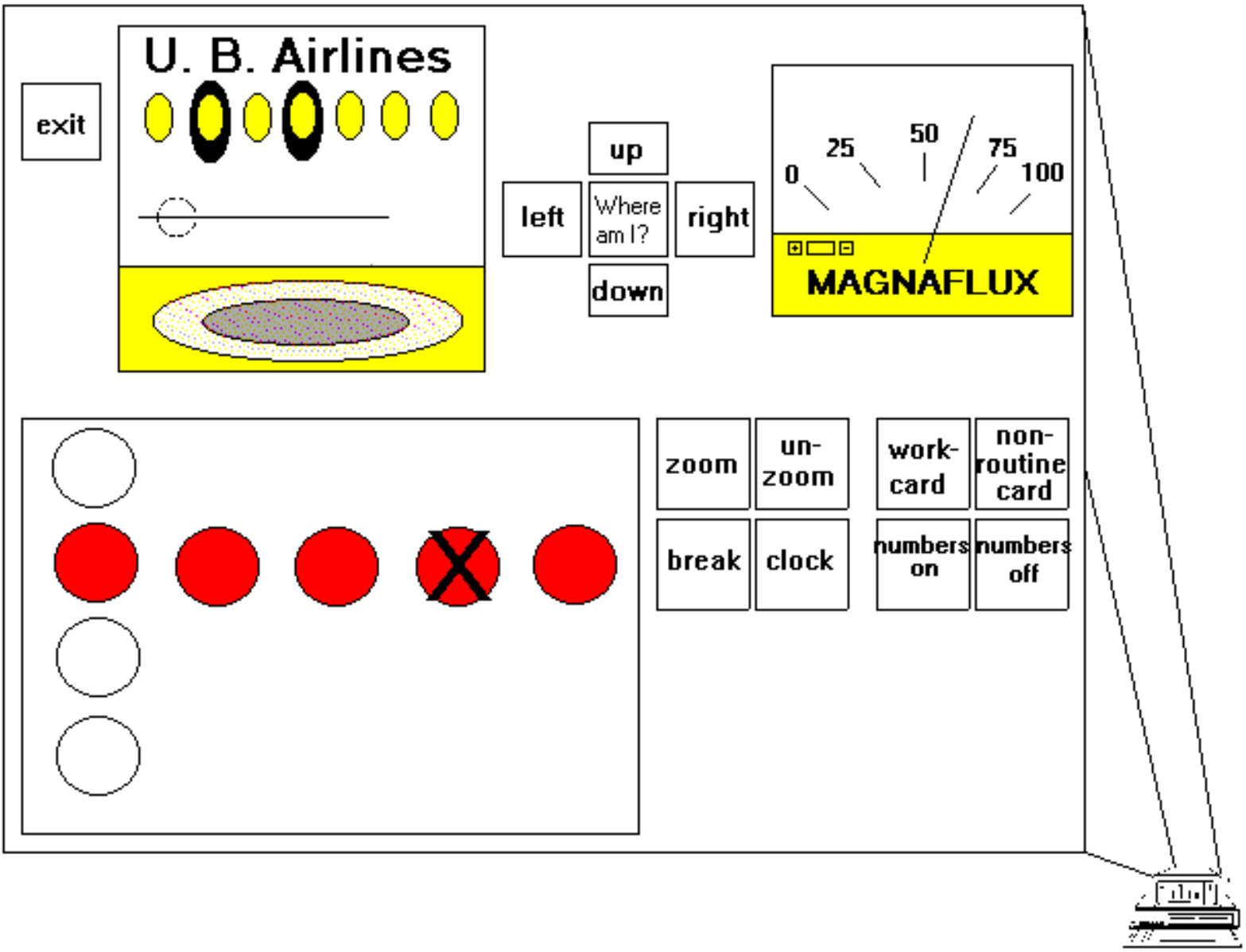

Figure 5.1 NDI Inspection Task Simulation (Drury et al., 1992)

5.1.2.1 Inspection Window

The lower left portion of the screen displayed the inspection window and contained the actual rivets to be inspected. Although it was possible to present more than one six-rivet row of rivets to the subject, only a single row was used in this study. An optical mouse was used to move a cursor around the circumference of each simulated rivet. The subject was free to examine the rivet until a decision was reached as to whether or not a crack was present. If it was decided that a rivet was defective, the right mouse button was pressed which caused a red cross to appear over the "defective" rivet and the words "rivet marked bad" appeared on the screen. If the rivet was judged to be nondefective, pressing the middle button caused the appearance of the words "rivet marked good." If a subject realized that an incorrect response had been made, it could be corrected by pressing the appropriate button.

When all six rivets had been inspected, the left mouse button was clicked on the directional block labeled "right." This caused a black marker ring to circle the last rivet inspected, and the next six rivets appeared in the inspection window.

5.1.2.2 Macro-View and Directionals

A macro-view in the upper left of the screen displayed a side view of the aircraft fuselage and the row of rivets being inspected. Since only a small portion of this row was being inspected at any given time during the task, the subject could move the cursor over the words "Where am I?" in this area and a momentary circle would appear over that portion of the rivet row currently being examined.

5.1.2.3 Eddy-Current Meter

The upper right segment of the display contained a simulated analog meter that served as the eddy-current output indicator. Deflections beyond a set point on the meter produced an audible alarm as well as a red flash on an indicator light. Meter deflections could be caused by:

- touching a rivet edge with the cursor or moving the cursor onto the rivet
- the cursor passing over a crack (All cracks were invisible and of varying length.)
- the cursor passing over or near simulated corrosion, scratches, or paint chips (These were simulated by 2 mm jagged lines at random locations adjacent to a rivet. Not all rivets contained such "noise," and no rivet contained more than one such noise spot.)

5.1.2.4 Lower Right Window

The subject could use the lower right area of the display to exercise a number of options (e.g., "zoom" to take a closer look at a rivet being inspected, stop task in order to take a break, display elapsed time). Only one feature was used in the present study; this feature caused a number to appear on each rivet and was only used by the experimenter during training feedback sessions to enable subjects to locate and re- check rivets incorrectly classified.

5.1.3 Predictors and/or Task Correlates

As noted earlier, the interim report (FAA/AAM & GSC, in press) identified a number of variables, measures of which appear potentially relevant to NDI selection and/or proficiency. These variables can be roughly separated into the following categories:

- Boredom Susceptibility
- Concentration/Attentiveness/Distractibility
- Extraversion/Impulsivity
- Motivation/Perseverance
- Decision Making/Judgement
- Mechanical/Electronics

Aptitude

- Need for Autonomy.

The following sections describe the tests and scales that were employed as measures of the above to identify possible correlates of performance on the NDI task.

5.1.3.1 Subjective Rating Scale (SRS)

The Subjective Rating Scale (SRS) is a simple self-rating scale that the author has used in several previous studies (e.g., Thackray, Bailey, & Touchstone, 1977; Thackray & Touchstone, 1991) to assess the current disposition of the subject, with measures generally taken before and after periods of task performance. The basic instrument consists of five 9-point scales measuring the dimensions of attentiveness, tiredness, strain, boredom, and annoyance. Two additional scales measuring perceived effort and perceived difficulty were used in the more recent study by Thackray and Touchstone (1991) and were included here as well. The SRS was studied most extensively in the early study by Thackray, Bailey, and Touchstone (1977). In that study, subjects falling at the extremes of rated boredom following performance of a simulated radar monitoring task were compared with respect to several performance and subjective variables. In general, those subjects who rated the task as quite boring showed the greatest decline in rated attentiveness and the largest performance decrement.

5.1.3.2 Bennett Mechanical Comprehension Test

One of the recommendations of a Southwest Research Institute study of ways to improve NDI technician proficiency was to select individuals who scored high on mechanical/electronics aptitude (Schroeder, Dunavant, and Godwin, 1988). This recommendation was also echoed in interviews with NDI instructors who expressed the belief that individuals who were above average in mechanical aptitude made better inspectors (Galaxy Scientific Corporation, in press). For these reasons, the Bennett Mechanical Comprehension Test was included in the test battery. This test appears to be one of the better measures of mechanical aptitude, and its norms include those for aviation mechanical jobs and electrical inspector trainee jobs. Unfortunately, the norms for aviation mechanical jobs, although updated in 1980, still do not include separate normative data for women.

5.1.3.3 Typical Experiences Inventory (TEI)

The ability to resist distraction, if it can be measured, would appear to have at least face validity in selecting inspectors (Wiener, 1975). A scale developed for use in several previous studies (Pearson and Thackray, 1970; Thackray, Jones, and Touchstone, 1973) and titled "Typical Experiences Inventory" consists of a series of statements designed to measure ability to work under conditions of (a) time stress, (b) threat of failure, (c) distraction, (d) social stress, and (e) physical stress. In the study by Thackray et al. (1973), subjects were selected who scored high, as well as low, on the distractibility subscale of this inventory. The high scorers showed significantly greater lapses of attention during performance of a repetitive task than did those scoring low. Because of these findings, it was decided to examine the relationship of scores on this subscale to possible performance decrement on the NDI task.

5.1.3.4 Arithmetic, Digit Span, and Digit Symbol Tests of the Wechsler Adult Intelligence Scale (WAIS)

Scores on these three subtests of the WAIS have been shown in numerous factor analytic studies to measure a factor that has been variously named "Freedom from Distractibility," "Attention-Concentration," or "Concentration-Speed" (e.g., Goodenough and Karp, 1961; Karp, 1963). Some or all of these WAIS tests have been found to relate significantly to inspection performance (Gallwey, 1982; Wang and Drury, 1989). Consequently, these tests were included as another measure of attention/concentration or, conversely, distractibility.

5.1.3.5 Eysenck Personality Inventory (EPI)

The Eysenck Personality Inventory is a short inventory that measures extraversion and neuroticism. The extraversion dimension has been studied extensively in the context of vigilance research because of the hypothesis, originally formulated by Eysenck (1967), that extraverts should have more frequent lapses of attention and hence more omission errors than would introverts. Reviews of the use of this personality dimension in vigilance research (Berch and Kantor, 1984; Wiener, 1975) have lent some support to the belief that extraverts generally do not perform as well on vigilance tasks as do introverts. Much less research has been conducted on personality variables in the area of inspection, and no studies of extraversion and inspection performance had been conducted at the time of Wiener's 1975 review. Since then, the author is aware of only one inspection study that has incorporated a measure of extraversion. Using a visual search task, Gallwey (1982) found that introverts, as measured by the EPI scale, had fewer search errors.

A factor analysis of the EPI scale has shown that it is comprised of two factors: sociability and impulsivity (Eysenck & Eysenck, 1963). Several studies have shown that the impulsivity dimension of extraversion is most clearly related to performance changes (Revelle, Humphreys, Simon, & Gilliland, 1980; Thackray, Jones, & Touchstone, 1974). In the latter study, extraverts showed a greater increase in lapses of attention during performance of a monotonous, repetitive task than did introverts. Further, it was determined that it was impulsivity, rather than the sociability dimension of extraversion, that was responsible for the obtained decrement.

Although a recent meta-analysis of vigilance studies over a 30-year period has concluded that evidence for the superiority of introverts is considerably less than previously believed (Koelega, 1992), Koelega feels that there is enough consistency in the findings to warrant continued research. Because of this, it was decided to include extraversion as measured by the EPI and the subscales measuring impulsivity and sociability in the present study.

5.1.3.6 Boredom Proneness Scale (Life Experiences Scale)

NDI inspection is typically a repetitive task and is frequently considered boring and monotonous (Schroeder, Dunavant, & Godwin, 1988). While the evidence relating experienced boredom to poor performance is tenuous at best, at least one study demonstrated a significant relationship of reported boredom and monotony to vigilance performance. As noted earlier, subjects falling at the extremes of rated boredom following performance of the simulated radar monitoring task showed the greatest decline in rated attentiveness and the largest performance decrement (Thackray et al., 1977).

Boredom in the above study was measured following task performance and thus can be considered a "state" assessment of boredom. The only scale specifically developed to assess the general construct of boredom proneness (i.e. a "trait" measure of boredom susceptibility) was developed by Farmer and Sundberg (1986). To the author's knowledge, this scale has not been used in studies of inspection performance. For this reason, it was included in the present study.

In order to somewhat disguise the intent of the scale, it was relabeled "Life Experiences Scale." Besides yielding a measure of boredom proneness, a recent factor analysis of the scale revealed that it was comprised of two factors, "apathy" and "inattentiveness" (Ahmed, 1990). Scores on these two subscales were also obtained. Ahmed's analysis revealed that the last item of the Farmer and Sundberg scale (item # 28) contributed to unreliability of the total test score, so this item was removed from the scale administered in this study.

5.1.3.7 Matching Familiar Figures Test (MFFT)

The MFFT is a test developed by Kagan and his associates (Kagan, Rosman, Day, Albert and Phillips, 1964) and consists of a series of 12 "stimulus" pictures each associated with 8 "response" pictures. Except for one picture in each set of the 8 response pictures, all differ from the stimulus in some minute detail. Subjects point to the picture they believe to be the correct one in each set and continue to point until the correct one is identified. Both the time to first response and the number of errors are scored. According to the above authors, the test measures a cognitive style known as reflection-impulsivity. Those who make quick inaccurate decisions on the test are said to have an impulsive cognitive style; those who make slow accurate decisions are said to have a reflective cognitive style.

This test has been considered to measure the tendency of subjects performing inspections tasks to opt for speed or accuracy in their speed/accuracy tradeoff (Drury, Gramopadhye, Latorella, Patel, Prabhu, and Reynolds, 1992). Presumably, impulsive subjects would tend to opt for speed at the expense of accuracy; conversely, reflective subjects would opt for the opposite. In a recent study, scores on the MFFT were found to be significantly related to several measures of inspection performance (Latorella et al., 1992). Since the task used in this latter study was the NDI simulation developed by Drury et al. and used in the present study, it seemed desirable to investigate further the relationship of MFFT scores to performance on this task.

5.1.3.8 Internal-External Locus of Control Scale

Rotter's (1966) Internal-External (I-E) Locus of Control Scale was developed to measure differences between individuals in the extent to which they believe either that rewards and reinforcements are contingent on or independent of their own behavior. The internal person believes that rewards are contingent on his own effort, attributes, or capacities; the external person believes that rewards result largely from luck, chance, fate, or forces outside of his control.

In a study of vigilance performance, Sanders, Halcomb, Fray, and Owen (1976) hypothesized that "internals," by constantly striving for mastery of a situation and by exhibiting a belief in their own ability to determine the outcome of their efforts, would perform better on a vigilance task than would an "external" person. The results supported this hypothesis in that internals, relative to externals, missed significantly fewer signals. Also, the internal subjects continued to progress in the monitoring task with a very small decline in performance, while the externals showed a consistently increasing performance decrement.

Because the Rotter scale has apparently not been used previously in inspection research, it seemed important to determine whether relationships similar to those found in vigilance would apply to inspection performance.

5.1.3.9 Jackson Personality Research Form (PRF)

The PRF is a widely used test designed to yield a set of scores for personality traits broadly relevant to the functioning of individuals in a wide variety of situations. It is a personality test that focuses primarily upon areas of normal functioning, rather than psychopathology.

The PRF used in this study consists of sixteen scales, of which nine were actually employed. The included scales were (a) Achievement, (b) Endurance, (c) Understanding, (d) Cognitive Structure, (e) Autonomy, (f) Change, (g) Impulsivity, (h) Infrequency, and (i) Desirability. A brief description of each and the reason(s) for its inclusion are as follows:

- *Achievement.* A measure of the willingness to put forth considerable effort to accomplish difficult tasks. This was included as a possible measure of intrinsic motivation or perseverance in task performance and was mentioned in the interim report as a desirable quality of NDI technicians (Galaxy Scientific Corporation, in press).
- *Endurance.* A measure of the willingness to work long hours and to be patient and unrelenting in work habits. This trait appears somewhat related to the above measure, and, in fact, loads on the same factor in a factor analysis of the test. It was included for the same reasons as the Achievement trait.
- *Understanding.* A measure of intellectual curiosity and the desire to understand many areas of knowledge. This was included because it was felt that it might correlate negatively with performance on a task as constrained and repetitive as eddy-current testing.
- *Cognitive Structure.* A measure of the need to make decisions based upon definite knowledge with a dislike of ambiguity and uncertainty. It was felt that this trait might be positively related to decision time, i.e. the time spent searching each rivet for possible faults.

- *Autonomy*. This is a measure of the need to be independent and not to be tied down, restrained, confined, or restricted in any way. This was mentioned in the interim report as a trait characterizing the most proficient inspectors (FAA/AAM & GSC, in press). This trait was also mentioned by some of the NDI instructors interviewed by the present author.
- *Change*. A like of new and different experiences, with a dislike and avoidance of routine activities. Inclusion of this trait is self-evident, since NDI tasks are so often referred to as boring and monotonous.
- *Impulsivity*. A measure of the tendency to act on the "spur of the moment" and without deliberation. This was included as an additional measure of impulsivity to compare with similar measures obtained from the EPI and from the MFFT described previously.
- *Infrequency and Desirability*. These were two scales included as measures of carelessness and response bias in taking the test.

The various tests and measures described in the preceding sections were included because it was felt that each might serve to measure some aspect of the categories mentioned under Section 5.1.3 as predictors and/ or correlates of NDI performance. It should be apparent that a number of these tests and measures bear similarities to one another and may indeed be measures of the same trait, aptitude, or ability. However, one cannot always tell from test titles and descriptors whether or not they measure similar things, and some were included to determine empirically the extent of their interrelationships, or lack thereof.

5.1.4 Procedure

Each subject was tested over two successive days. The morning of the first day was devoted to administration of the various tests and measures; during the afternoon, subjects practiced using the mouse and were required to read and be tested on a document describing eddy-current testing. They then practiced the NDI simulation task. The afternoon training procedures were essentially the same as those used and reported by Drury et al. (1991). Subjective rating scales were administered at various times during the course of both days.

Training in the use of the mouse was provided by a display program consisting of an enlarged picture of a rivet head with a training circle surrounding it. The subject practiced using the mouse and cursor to circle the rivet while staying within the circle. After each pre-selected block of training trials, feedback was provided consisting of average times required to circle the rivet, and averages of the number of times the cursor head touched the rivet or went outside the circle. Training continued until the subject reached a consistent level of performance. This usually required approximately 30 minutes of practice.

Training on the task began with a short (20-rivet) demonstration session in which the basic elements of the NDI task were explained. This was followed by three training sessions each 60 rivets long. Thirty percent of the rivets in each of the three training sessions contained faults (cracks). In addition, the second and third sessions also contained small, but visible (2 mm) "noise" spots at various locations at or near a rivet. Frequency of "noisy rivets" was also thirty percent. Locations of faults and noise were randomly assigned for each task session (both training and subsequent test tasks). Performance feedback was automatically provided after each block of 10 rivets. In the first and second sessions, training circles were provided around each rivet to assist the subject in keeping the cursor in the appropriate region while circling the rivets; no training circles were used in the third session.

On the morning of the second day, subjects performed a short (20-rivet) "refresher" version of the NDI task and then performed two lengthy (180-rivet) test sessions. Since these sessions were self-paced, test durations for each subject varied from 60 to 90 minutes. There was a fixed 10-15 minute rest break between each session, although subjects were told that they could take short (10-20 second "stretch" breaks as needed during any session. Following a 60-minute lunch break, this same procedure (two 180-rivet sessions), minus the short practice session, was followed in the afternoon. No feedback was provided following test sessions and the frequency of both faults and noise was held at 30 percent each. At the end of the second day, subjects were debriefed and questioned about various attitudes about and approaches to the NDI task.

5.2 RESULTS

5.2.1 Criterion Performance

Figure 5.2 shows mean values for misses and false alarms across the three training sessions. Visual examination of this figure reveals an apparent increase in false alarms during the second training session. However, subjects typically made no errors (misses or false alarms) during these training sessions and the increase noted above was caused by two subjects making two false alarms during this second session. The only important information to be gained from this figure is that miss rate remains level during the last two training sessions and false alarm rate returns to its initial level.

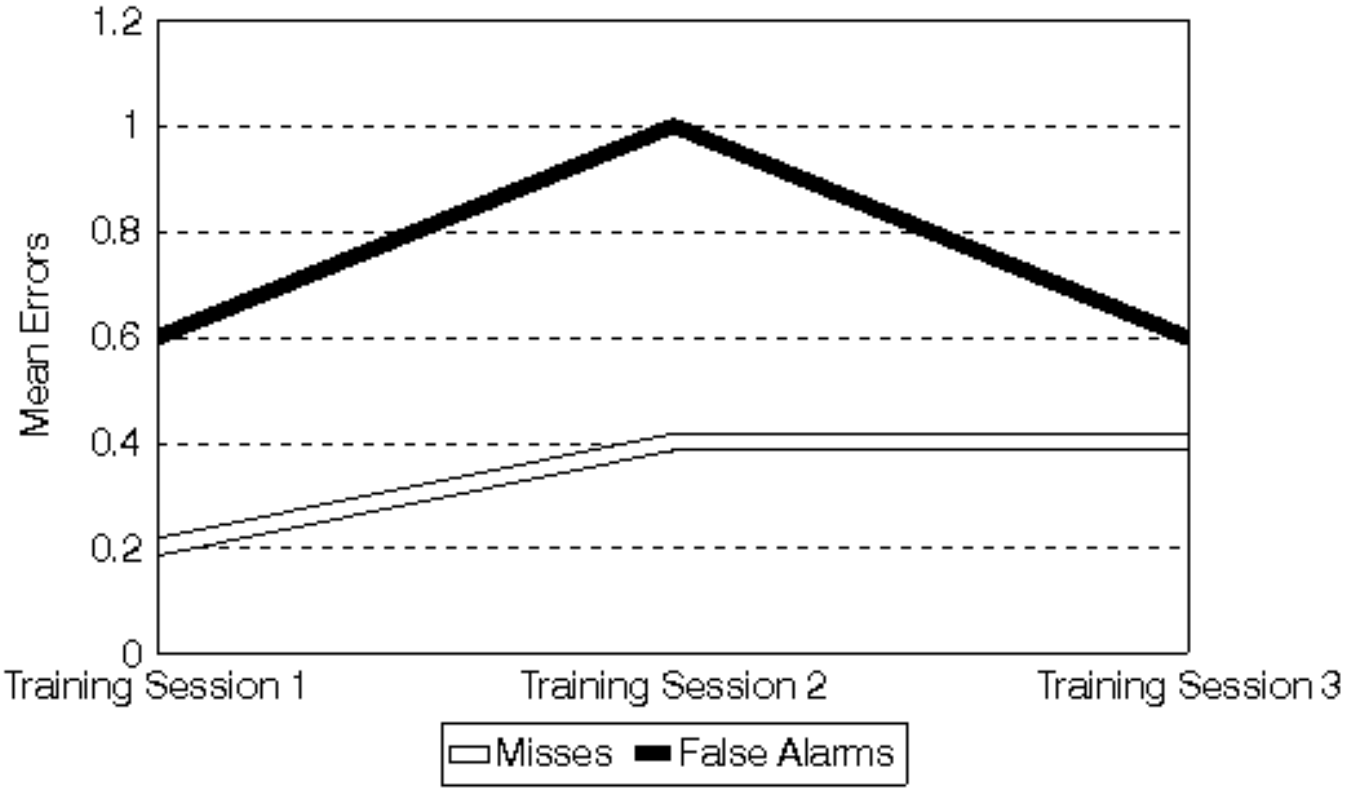

Figure 5.2 Mean Misses and False Alarms Across Training Sessions

Raw data for misses and false alarms during the four test sessions are given in Tables 5.2 and 5.3respectively. Mean values obtained from these two tables are plotted as bar graphs in Figure 5.3. In addition, this figure also plots miss and false alarm data for the first and second half of each test session shown in Tables 5.2 and 5.3. Notice the general increase in both misses and false alarms across sessions and the general increase in both types of errors within sessions. Analyses of variance revealed the differences between sessions to be significant for both misses ($F(3/15)=4.90$, $p=.014$) and false alarms ($F(3/15)=3.72$, $p=.035$), while differences between first and second halves of the sessions approached significance at the .05 level for both misses ($F(1/5)=5.79$, $p=.061$) and false alarms ($F(1/5)=4.65$, $p=.084$). The Interaction of Session Halves with Sessions was not significant for either misses or false alarms.

Table 5.2 Frequency of Missed Faults Over the Four Test Sessions

Misses		Session 1		Session 2		Session 3		Session 4		
Subject Number	Gender	1st 1/2	2nd 1/2	1st 1/2	2nd1/ 2	1st 1/2	2nd 1/2	1st 1/2	2nd 1/2	Total
01	Male	1	5	3	2	2	2	5	10	30
02	Female	0	1	5	2	3	1	1	0	13
03	Male	5	11	5	16	19	13	10	11	90
04	Male	3	6	6	7	5	5	6	8	46
05	Female	2	5	6	13	12	9	2	13	62
06	Female	3	4	1	1	4	4	7	18	42

Table 5.2 Frequency of Missed Faults Over the Four Test Sessions

Table 5.3 Frequency of False Alarms Over the Four Test Sessions

False Alarms		Session 1		Session 2		Session 3		Session 4		
Subject Number	Gender	1st 1/2	2nd 1/2	1st 1/2	2nd 1/2	1st 1/2	2nd 1/2	1st 1/2	2nd 1/2	Total
01	Male	0	0	0	1	4	8	6	7	26
02	Female	0	0	0	1	0	0	0	2	03
03	Male	3	4	6	14	6	16	19	15	83
04	Male	0	1	1	1	0	3	12	14	32
05	Female	0	0	0	1	1	0	0	1	03
06	Female	2	0	0	2	0	0	0	1	05

Table 5.3 Frequency of False Alarms Over the Four Test Sessions

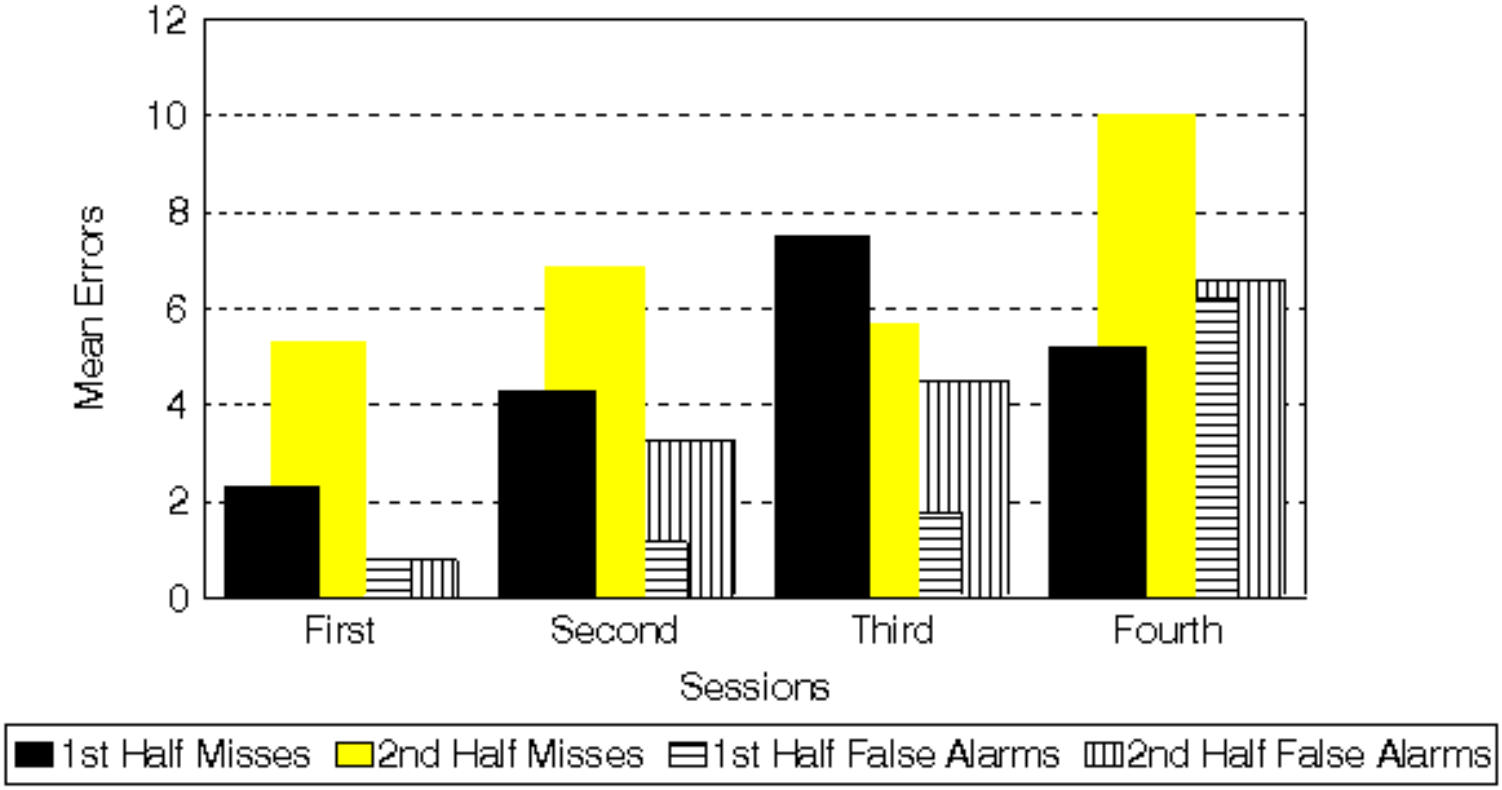

Figure 5.3 Mean Misses and False Alarms Across Test Sessions

Total task time and mean time per rivet are shown in Table 5.4. As can be seen in this table, there was a general decrease in both total task time and mean time per rivet during training as well as during the test sessions. Analysis of variance revealed the decline in total task time to be nonsignificant both during training and test sessions (p>.10). Mean time per rivet, however, decreased significantly both during training (F(8/2)=4.27, p=.05) and during test sessions (F(3/15)=8.22, p=.002). The increase in mean time per rivet from the last training session to the first test session can probably be attributed to the fact that the larger number of rivets in the test sessions caused the meter in the simulation to be slightly more sluggish in its response. This aspect was explained to subjects before the first test session, and subjects characteristically began the session by circling the rivets with greater care and deliberation.

Table 5.4 Mean Total Task Time and Mean Time per Rivet

Task	Mn Total Time (min)	Mn Time Per Rivet (sec)
Training Session 1	29.7	23.4
Training Session 2	27.9	23.6
Training Session 3	26.3	20.3
Test Session 1	92.7	28.8
Test Session 2	82.8	24.1
Test Session 3	93.1	24.9
Test Session 4	78.6	19.4

Table 5.4 Mean Total Task Time and Mean Time per Rivet

Estimates of task reliability were obtained by comparing the combined data of test sessions 1 and 2 with the combined data of sessions 3 and 4. The correlation between combined sessions was r=.82 ($p < .05$) for false alarms and r = .81 ($p < .05$) for missed events. Total false alarms across sessions was also compared with total misses across sessions. The obtained correlation was r=.73 (p=.10) which, although failing to reach the .05 level, suggests that the two measures might be significantly related, given a larger sample size.

5.2.2 Subjective Measures

As indicated earlier, subjective ratings of attentiveness, tiredness, strain, boredom and annoyance were obtained at various times during the two-day period. Measurements of each were obtained at the beginning of each morning and afternoon of the two-day periods each subject was tested. These same measures were obtained at the end of the first day's training session and at the end of the morning and afternoon test sessions of the second day. In addition, items relating to perceived task difficulty and to the amount of effort required to maintain alertness were also administered at the end of the above three sessions. Data for the subjective ratings are given in Table 5.5.

Table 5.5 Mean Pre- and Post-Session Subjective Ratings

Variable	Mn Pre-Session Ratings	Mn Post-Session Ratings
Attentiveness	7.0	5.7
Tiredness	4.3	5.3
Strain	3.5	4.3
Boredom	3.2	3.9
Annoyance	1.2	1.8
Effort	4.4	5.1
Difficulty	2.7	3.9

Table 5.5 Mean Pre- and Post-Session Subjective Ratings

Comparing the pre- and post-session data of Table 5.5, it is apparent that attentiveness decreased, while tiredness, strain, boredom, and annoyance all increased. Analyses of variance, however, revealed that only those changes associated with tiredness and strain were significant ($F(1/4)=7.90$, $p=.048$ and $F(1/4)=11.0$, $p=.029$ respectively). The increases in perceived effort and perceived difficulty were nonsignificant ($p>.10$).

5.2.3 Predictor Measures

Means and standard deviations for the various predictor tests and measures used are shown in Table 5.6. Because of the small sample size used in the pilot study, with the attendant risk of obtaining spurious correlations, no attempts were made to correlate or compare any of these measures with performance variables. Where normative data were available, most measures appeared to fall within one standard deviation of the reported mean values. Several deviations are worth mentioning, however: The sample as a whole scored relatively low on impulsivity as measured by the PRF and EPI scales, on boredom susceptibility as measured by the Boredom Proneness Scale, and relatively high on endurance as measured by the PRF scale.

Table 5.6 Means and Standard Deviations of the Predictor Measures

Variable	Mean	Standard Deviation
Typical Exper Inv	22.00	6.90
Bennett Mech Comp	40.67	2.88
LES Boredom Prone	5.33	4.03
LES Inattentiveness	2.33	2.58
LES Apathy	3.83	2.32
MFFT Mn Error	1.23	0.37
MFFT Mn Time	59.80	23.40
EPI Extraversion	12.50	1.52
EPI Impulsivity	3.83	1.47
EPI Sociability	7.33	1.37
WAIS Digit Symbol	56.17	14.18
WAIS Digit Span	13.00	2.83
WAIS Arithmetic	7.83	2.93
PRF Achievement	12.50	2.59
PRF Autonomy	6.83	2.48
PRF Change	8.50	1.64
PRF Cog Structure	10.17	2.71
PRF Endurance	13.33	1.97
PRF Impulsivity	3.50	3.39
PRF Understanding	7.00	4.34
I-E Scale	8.33	3.61

Table 5.6 Means and Standard Deviations of the Predictor Measures

5.3 DISCUSSION

Discussion of the findings of this pilot study is not extensive for two reasons. First, the essential purpose of this initial study was simply to test and evaluate the various tests, scales, procedures and task used in order to determine whether changes were necessary for the major study of which this was only the initial phase. Second, the small size of the sample used increases the risk of obtaining spurious results and makes any conclusions drawn from the data risky at best. However, some comment on the findings is necessary and appropriate.

The significant increase in misses (faults) obtained over the four test sessions suggests a boredom and/or fatigue effect resulting from the repeated testing, perhaps causing the subjects to become less careful or attentive in examining the rivets. This decline in performance efficiency was accompanied by an increase in ratings of both boredom and fatigue. However, only the increase in tiredness was found to be significant. There was also a significant increase in strain, suggesting that continued task performance over the four successive sessions was not only tiring, but emotionally demanding as well.

While an increase in fatigue (tiredness) across sessions seems a plausible explanation for the increase in missed faults, it is somewhat puzzling to see how increasing tiredness could also result in an increase in false alarms as Figure 5.3 shows. The possible increase in carelessness and inattentiveness with fatigue could again be a contributing factor. Recall that Table 5.4 showed a significant decrease during the test sessions in mean time spent examining rivets. While this could merely be an indication of increased skill in the use of the mouse, it seems more likely that it indicates less careful examination of the individual rivets by at least some of the subjects, especially during the last session. Less careful examination would likely increase the number of times a rivet edge was touched or "noisy" area was crossed. The resulting meter deflections might then be erroneously interpreted as faults.

One should also not overlook the possibility that the increase in false alarms might simply be an artifact of the particular sample used. Examination of Table 5.3 reveals that half of the subjects (subjects 2, 5, and 6) showed essentially zero false alarms across all four sessions. The increase in false alarms was essentially the result of subjects 1, 3, and 4. Thus, whether the larger, more representative sample to be used in the primary study to follow this one will show an increase or no increase in false alarms across sessions cannot really be predicted from these data. Likewise, one should not attribute much significance to the fact that the three subjects of Table 5.3 showing no increase in this type of error were all females. Whether or not a gender difference exists in this respect again must await the findings of the subsequent study.

Of the remaining data, there is little more that can be discussed. Reliability of the NDI task, in terms of both misses and false alarms, was shown to be acceptable, and one would hope that these reliability estimates will approximate those to be obtained in the primary study. As noted earlier, the particular sample used scored somewhat below average on measures of impulsivity, on measured boredom susceptibility, and above average on the PRF measure of task endurance. Also, the sample appeared to show a gender difference in frequency of false alarms on the NDI task. Whether any or all of the above characteristics of this sample will be reflected in the larger sample of the primary study remains to be determined. The experiment protocol of the pilot study proved to be workable with testing and training occurring on the first day and with the second day devoted to successive test trials. No problems were encountered in administration of the various psychometric tests and measures and all will be used in the primary study.

As indicated in the Introduction, the pilot study reported here represents the initial phase of a larger effort (the primary study) to investigate relationships of selected subject variables to NDI performance. A second aspect of this effort is to investigate possible task-induced fatigue resulting from sustained performance on the NDI task and to examine possible interactions between performance changes and the various subject variables examined in the pilot study. Approximately 25-30 subjects will be tested in the primary study, but the basic procedures and approach will remain essentially those incorporated in this pilot study.

5.4 REFERENCES

Ahmed, S. M .S. (1990). Psychometric properties of the boredom proneness scale. *Perceptual and Motor Skills, 71*, 963-966.

Berch, D. B. & Kanter, D. R. (1984). Individual differences. In J. S. Warm (Ed.). *Sustained attention in human performance*. New York: Wiley.

Drury, C. G., Gramopadhye, A., Latorella, K., Patel, S., Prabhu, P., & Reynolds, K. (1992). *Human reliability in aircraft inspection*. Phase II Report on FAA Contract to Galaxy Scientific Corporation, Atlanta, Georgia.

Drury, C. G., Prabhu, P., Gramopadhye, A., & Latorella, K. (1991). *Nondestructive testing in aircraft inspection*. Report of a pilot study prepared under subcontract 89-1014-SC-3 to Galaxy Scientific Corporation, Mays Landing, New Jersey.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, Illinois: Thomas.

Eysenck, S. B. G. & Eysenck, H. J. (1963). On the dual nature of extraversion. *British Journal of Psychology, 2*, 46-55.

Farmer, R. & Sundberg, N. D. (1986). Boredom proneness - The development and correlates of a new scale. *Journal of Personality Assessment, 50*, 4-17.

Gallwey, T. J. (1982). Selection of tests for visual inspection on a multiple fault type task. *Ergonomics, 25*, 1077-1092.

Goodenough, D. R. & Karp, S. A. (1961). Field dependence and intellectual functioning. *Journal of Abnormal and Social Psychology, 63*, 241-246.

Kagan, J., Rosman, B., Day, D., Albert, J., & Phillips, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs, 78*, (1, Whole No. 578).

Karp, S. A. (1963). Field dependence and overcoming embeddedness. *Journal of Consulting Psychology, 27*, 294-302.

Koelega, H. S. (1992). Extraversion and vigilance performance: 30 years of inconsistencies. *Psychological Bulletin, 112*, 239-258.

Latorella, K. A., Gramopadhye, A. K., Prabhu, P. V., Drury, C. C., Smith, M. A., & Shanahan, D. E. (1992, October). *Computer-simulated aircraft tasks for off-line experimentation*. Paper presented at the Annual Meeting of the Human Factors Society, Atlanta, Georgia.

Pearson, D. W. & Thackray, R. I. (1970). Consistency of performance change and autonomic response as a function of expressed attitude toward a specific stress situation. *Psychophysiology, 6*, 561-568.

Revelle, W., Humphreys, M. S., Simon, L., & Gilliland, K. (1980). The interactive effect of personality, time of day, and caffeine: A test of the arousal model. *Journal of Experimental Psychology: General, 109*, 1-31.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*, (1, Whole No. 609).

Sanders, M. G., Halcomb, C. G., Fray, J. M., & Owens, J. M. (1976). Internal-external locus of control and performance on a vigilance task. *Perceptual and Motor Skills, 42*, 939-943.

Schroeder, J. E., Dunavant, D. W., & Godwin, J. G. (1988). *Recommendations for improving Air Force nondestructive inspection technician proficiency*. SwRI Project No. 17-7958-845, San Antonio Air Logistics Center, Air Force Logistics Command, Kelly Air Force Base, Texas.

Thackray, R. I., Bailey, J. P., & Touchstone, R. M. (1977). Physiological, subjective, and performance correlates of reported boredom and monotony while performing a simulated radar control task. In R. R. Mackie (Ed.). *Vigilance: Theory, Operational Performance, and Physiological Correlates*. New York: Plenum.

Thackray, R. I., Jones, K. N., & Touchstone, R. M. (1973). Self-estimates of distractibility as related to performance decrement on a task requiring sustained attention. *Ergonomics, 16*, 141-152.

Thackray, R. I., Jones, K. N., & Touchstone, R. M. (1974). Personality and physiological correlates of performance decrement on a monotonous task requiring sustained attention. *British Journal of Psychology, 65*, 351-358.

Thackray, R. I. & Touchstone, R. M. (1991). Effects of monitoring under high and low taskload on detection of flashing and coloured radar targets. *Ergonomics, 34*, 1065-1081.

Wang, M. J. & Drury, C. G. (1989). A method of evaluating inspector's performance differences and job requirements. *Applied Ergonomics, 20.3*, 181-190.

Wiener, E. L. (1975). Individual and group differences in inspection. In C. G. Drury & J. G. Fox (Eds.). *Human reliability and quality control*. New York: Taylor & Francis.