CR-172040

DEVELOPMENT OF A MENU OF PERFORMANCE TESTS
SELF-ADMINISTERED ON A PORTABLE MICROCOMPUTER

Robert L. Wilkes, Lois-Ann Kuntz, & Robert S. Kennedy

October 16, 1987

NASA Technical Report
Contract No. NASA 9-17326

Submitted by:
Essex Corporation
1040 Woodcock Road Suite 227
Orlando, FL 32803
(305) 894-5090

(NASA-CR-172040) DEVELOPMENT OF A MENU OF PERFORMANCE TESTS SELF-ADMINISTERED ON A PORTABLE MICROCOMPUTER (Essex Corp.) 40 p N88-16452
CSCL 09B Unclas
G3/61 0120668

This report was prepared for NASA, Lyndon B. Johnson Space Center, Space and Life Sciences Procurement Office, Houston, TX 77058, Contract No. NAS9-17326.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Eighteen cognitive, motor, and information processing performance subtests were screened for self-administration over 10 trials by 16 subjects. When altered presentation forms of the same test were collectively considered, the battery composition was reduced to 10 distinctly different measures. A fully automated microbased testing system was employed in presenting the battery of subtests. Successful self-administration of the battery provided for the field testing of the automated system and facilitated convenient data collection. Total test administration time was 47.2 minutes for each session. Results indicated that nine of the tests stabilized, but for a short battery of tests only five are recommended for use in repeated-measures research. The five recommended tests include: the Tapping series, Number Comparison, Short-term Memory, Grammatical Reasoning, and 4-Choice Reaction Time. These tests can be expected to reveal three factors: (1) cognition, (2) processing quickness, and (3) motor. All of the tests stabilized in 24 minutes, or approximately two 12-minute sessions.

INTRODUCTION

The primary purpose of the present study was to continue with the development of metrically sound human performance tests suitable for repeated-measures research. Eighteen microbased tests were examined in the process of fulfilling this study's purpose. A second, but equally important, purpose was to assess the viability of subject self-administration of the battery in nonlaboratory environments. The approach appears to have important implications for research using computers. Some researchers have used computers for self-monitoring as a method for intervention with children in classrooms (Tombari, Fitzpatrick, & Childress, 1985).

METHOD

SUBJECTS

Eighteen Casper College freshman and sophomore students were contacted regarding participation in the study. The individuals were solicited from a pool of subjects with previous experience in microbased human performance testing (NASA Contract No. 9-17326 and NSF award BNS 8460765). Subject motivation for participation was high with 100% of the individuals contacted volunteering. One subject was removed from the study for noncompliance with testing protocol and the data for a second subject were inadvertently destroyed during a data transfer process. Final analyses were based on data obtained from $\underline{N}$=16 subjects with nine women and seven men participating. Subject procurement and data collection procedures were carried out in accordance with APA principles for research with human subjects (American Psychological Association, 1982). The study was reviewed and approved by the Casper College Human Use Committee prior to subject solicitation (Appendix A). Subjects completing the study were paid for their efforts at the rate of $5.00 per session and most sessions were of approximately 45-minute duration.

PROCEDURE

In this experiment, all testing was accomplished with a fully automated microprocessor system. The microbased battery of eighteen subtests was programmed to be self-administered over 10 trials of testing. All testing was microbased and paper-and-pencil analogues of the automated tests were not administered. Self-administration of the battery provided the opportunity for field testing of the automated system as well as facilitating convenient data collection. Prior to initial testing, subjects were thoroughly introduced to the purpose and nature of the study. Pertinent biographical data were obtained and each subject was reviewed in the operation of the microbased testing system. Self-administration of the first battery was then completed in the experimenter's presence to ensure knowledge of system operation and to surface questions. Typically, the battery was self-administered twice per day or until a subject had fulfilled the 10 required replications.

The possibility for compromise of established testing protocol in nonstandardized testing sequences by subject-regulated data collection cannot be ignored. Therefore, special attention was given to experimental control. In order to handle this problem, training, orientation, and indoctrination were emphasized. As part of the effort to maintain the internal validity of the study, subjects were extensively trained and instructed during the laboratory data collection session. Care was taken with each subject in determining an adequate regimen for self-administration of the test battery. A testing regima was established relative to the subject's personal schedule and to general testing procedure. General procedure called for testing twice per day over a five-day period at times amenable to data collection. Departures were allowed within certain limitations; however, the prevailing criterion which was applied in such cases was consideration for maintenance of subject motivation. Special efforts were made to ensure that each subject understood the consequences to the study for engaging in activities likely to influence test performance in adverse and uncontrolled ways. The potential effects of drugs, alcohol, fatigue, emotional distress, illness, and other internal or environmental agents on behavior were reviewed and stressed. Subjects were admonished not to self-test if, for any reason, their performance could be compromised. It should be noted that repeated-measures studies are particularly susceptible to such problems. The microprocessor capability for monitoring test performance on a date/time basis was demonstrated and subjects were informed that test data would be checked and verified as a condition of final payment. As a further precaution, the microprocessors were "safed" to prevent memory access, thereby negating the possibility of subjects obtaining knowledge of results or altering test performance scores. Lastly, subjects were informed that the performance tests were the focus of the study as opposed to the individuals themselves, and handouts and reminders concerning the test system operation and testing protocol were provided.

Exit interviews were conducted individually at the conclusion of the study. During the interview, data were examined and questions were raised regarding performance. It appears that the subjects were well informed of the purpose and methods of the study and acted in accordance with study procedure and testing protocol. Furthermore, subjects were highly motivated to fulfill the research obligations with 94% of the volunteers completing the study. It is not certain that all data from all subjects met the desired standard, but the use of such measures is advocated for minimizing such risks.

APPARATUS

Microcomputer testing was accomplished with the Automated Performance Test System (APTS) implemented on the NEC PC8201A microprocessor. The NEC PC8201A is configured around an 80C85 microprocessor with 64K internal ROM containing Basic, TELCOM, and a TEXT EDITOR. RAM capacity may be expanded to 96K onboard, divided into three separate 32K banks. An RS-232 interface allows for hook-up to modem, to a CRT or flat-panel display, to a "smart" graphics module, to a printer, or to other computer systems. Visual displays are presented on a 8-line LCD with 40 characters per line. Memory may be transferred to 32K modules with independent power supplies for storage or mailing. The entire package is lightweight (3.8 lbs), compact (110W x 40H x 130D mm), and fully portable with rechargeable nickel cadmium batteries permitting up to four hours of continuous operation. Table 1 abstracts the technical features of the system which are more fully described in NEC (1983) and Essex (1985).

TABLE 1. NEC PC8201A TECHNICAL SPECIFICATIONS

FEATURES	SPECIFICATIONS
SIZE	30 CM (11 IN) X 22 CM (8.25 IN) X 6 CM (2.5 IN). 1.7 KG (3.8 LBS)
CPU	80C85 (CMOS VERSION OF 8085) WITH 2.4 MHZ CLOCK
ROM	32K (STANDARD) - 128K (OPTIONAL)
RAM	24K (STANDARD) - 96K (OPTIONAL)
KEYBOARD	67 STANDARD (10 FUNCTIONS, 4 CURSOR DIRECTIONAL AND 58 ADDITIONAL)
DISPLAY	19 CM (7.5 IN) X 5.0 CM (2.0 IN) WITH REVERSE VIDEO OPTION. MAY BE CONFIGURED AS EITHER A 240 X 62 ELEMENT MATRIX OR 40 CHARACTERS X 8 LINE DISPLAY
INTERFACES	1 PARALLEL (CENTRONICS COMPATIBLE) AND 3 SERIAL (RS232C AND 6 & 8 PIN BERG) JACKS
POWER SUPPLY	4 AA NONRECHARGEABLE BATTERIES, OR RECHARGEABLE NICKEL-CADMIUM PACK, OR AC ADAPTER 50/60 Hz @ 120 VAC, OR EXTERNAL BATTERY SYSTEMS (e.g., 8 AMP HR)

MATERIALS

The microbased test battery consisted of 18 individual performance subtests (Table 2). A number of the tests (i.e., Tapping, Reaction Time, Auditory Count, Visual Count) were presented in three forms. When altered presentation forms of the same test were collectively considered, the battery composition was reduced to 10 distinctly different measures. These tests were selected for inclusion into the test battery on the basis of one or more of the following criteria: (a) demonstrated conformity to the criteria for "good" performance tests (Bittner, Carter, Kennedy, Harbeson, & Krause, 1986); (b) indications representing factors associated with cognitive, perceptual, or motor skills; and (c) compatibility with the microbased testing mode. The tests in the order of their appearance in the test battery are discussed below.

TABLE 2. MICROBASED BATTERY TASK ORDER AND TESTING TIME

Battery Task Order	Trials/ Battery	Practice Time	Trial Time	Total Task Time in a Battery	Total Task Time for 10 Battery Replications
Preferred Hand Tapping	2	10[a]	10	20	200
Reaction Time (1 Choice)	1	30	120	120	1200
Auditory Count (1 Stimulus)	1	0	300	300	3000
Short-Term Memory	1	30	120	120	1200
Auditory Count (2 Stimuli)	1	0	300	300	3000
Number Comparison	1	30	45	45	450
Auditory Count (3 Stimuli)	1	0	300	300	3000
Air Combat Maneuv.	1	0	120	120	1200
Reaction Time (2 Choice)	1	30	120	120	1200
Two-Hand Tapping	2	10	10	20	200
Pattern Comparison	1	30	120	120	1200
Visual Count (1 Stimulus)	1	0	300	300	30000
Associative Memory	1	0	90	90	900
Visual Count (2 Stimuli)	1	0	300	300	3000
Grammatical Reason.	1	30	120	120	1200
Reaction Time (4 Choice)	1	30	120	120	1200
Visual Count (3 Stimuli)	1	0	300	300	3000
Nonpreferred Hand Tapping	2	10	10	20	200
Totals		240	2805	2835	28350

[a] All time data are reported in seconds

TAPPING. The test is accomplished by alternately pressing keys on the microprocessor keyboard. The task was administered in three different forms: (a) Preferred-hand Tapping (PTAP); (b) Two-hand Tapping (THTAP); and (c) Nonpreferred-hand Tapping (NTAP). Performance is based on the number of alternate key presses made in the allotted time. In a recent study (Kennedy, Dunlap, Wilkes, & Lane, 1985), tapping was described as a psychomotor skill assessing factors common to both Aim and Spoke. Tapping has also been highly recommended for inclusion in a repeated-measures microcomputer battery (Kennedy, Dunlap, Jones, Lane, & Wilkes, 1985; Wilkes, Kennedy, Dunlap, & Lane, 1986).

REACTION TIME. The Visual Reaction Time Test (Donders, 1968) involves the presentation of a visual stimulus and measurement of a response latency to the stimulus. The subject's task is to respond as quickly as possible with a key press to a simple visual stimulus. On each trial the visual stimulus is prefaced by an auditory signal whose time preceding the visual stimulus is varied. The task was administered in three different forms: (a) 1-Choice (RT1), (b) 2-Choice (RT2), and (c) 4-Choice (RT4). Reaction time is measured from the onset of the visual stimulus to the key press. Simple reaction time has been described as a perceptual task responsive to environmental effects (Krause & Bittner, 1982) and has been recommended for repeated-measures research (Bittner et al., 1986; Kennedy, Dunlap, Jones, Lane, & Wilkes, 1985).

COUNTING (AUDITORY AND VISUAL). The Counting tests (Jerison, 1955; Kennedy & Bittner, 1980) are accomplished by the subject accurately monitoring the repeated occurrence of a particular stimulus. The subject must indicate when a stimulus has been presented four times in succession and then repeat the monitoring process until the end of the trial. The complexity of the task may be altered by presenting one, two, or three stimuli during the same trial and requiring the subject to monitor each. When multiple stimuli are employed the rate of presentation for each individual stimulus is varied at either 8, 6 or 5 presentations/minute. The subject indicates a perceived four count for a particular stimulus by making an appropriate key press. Performance is scored according to the number of correct four counts, the number of omissions, and the number of errors for each stimulus. In the auditory test mode, the stimuli were varied by presenting "beeps" of three different frequencies: low (ACTL), medium (ACTM), and high (ACTH). In the visual task mode, the stimuli were varied by presenting lighted boxes at different locations on the screen: right (VCTR), middle (VCTM), and left (VCTL). The Counting tests are best presented with automated testing and are described as coding and short-term, memory-type tasks. Previous repeated-measures research have not been conducted with the Counting tests in their visual modes.

SHORT-TERM MEMORY (STM). The Short-Term Memory Task (Sternberg, 1966) involves the presentation of a set of four digits for one second (positive set), followed by a series of single digits presented for two seconds (probe digits). The subject's task is to determine if the probe digits accurately represent the positive set and respond with the appropriate key press. Performance is based on the number of probes correctly identified. Short-Term Memory is described as a cognitive-type task which reflects short-term memory scanning rate (Bittner et al., 1986). Previous research with the task (Carter, Kennedy, Bittner, & Krause, 1980; Kennedy, Dunlap, Jones, Lane, & Wilkes, 1985; Wilkes et al., 1986) has indicated that Short-Term Memory is acceptable for use in repeated-measures research.

NUMBER COMPARISON (NC). The Number Comparison task (Ekstrom, French, Harman, & Dermen, 1976) involves the presentation and comparison of two sets of numbers. The subject's task is to compare the first and second set and decide if they are the same or different. Numbers ranged from 3 to 7 digits in length with the sets always equal in length. Number sets that differed, did so on the basis of only one digit. Number comparison has been described as a perceptual task with perceptual speed, an important factor to performance. Previous research with Number Comparison has indicated that the task is acceptable for repeated-measures research and highly correlated with longer and more complex tests of arithmetic computation (Bittner, Carter, Krause, Kennedy, & Harbeson, 1983; Carter & Sbisa, 1982).

AIR COMBAT MANEUVERING (ACM). The Air Combat Maneuvering test emulates a combat-type video game. The subject's task is to "shoot" a randomly moving stimulus target. The subject laterally positions and fires a projectile through activation of appropriate microprocessor keys. Direct hits result in a more rapid accumulation of points than peripheral hits. The subject is provided with visual and auditory feedback for scoring hits and a continuous update of accumulated points is displayed. Air Combat Maneuvering can only be presented in the microbased testing mode and has been described as a pursuit tracking-type task (Kennedy, Bittner, & Jones, 1981). Previous research (Kennedy, Bittner, Harbeson, & Jones, 1981) has indicated that a related task was acceptable for use in repeated-measures research.

PATTERN COMPARISON (PC). The Pattern Comparison task (Klein & Armitage, 1979) is accomplished by the subject examining a pair of dot patterns and determining whether they are similar or different. Patterns are randomly generated with similar and different pairs presented in random order. Performance is scored according to the number of pairs correctly identified as similar or different. Pattern Comparison has been described as a spatial ability important to perceptual performance. According to Bittner et al. (1986), Pattern Comparison "assesses an integrative spatial function neuropsychologically associated with the right hemisphere" (p. 699). A review of Pattern Comparison studies (Bittner et al., 1986) indicated that the task is acceptable for use in repeated-measures research. Recent field testing with a microcomputer adaptation of the task (Kennedy, Dunlap, Jones, Lane, & Wilkes, 1985; Kennedy, Wilkes, Lane, & Homick, 1985; Wilkes et al., 1986) resulted in strong recommendations for inclusion of Pattern Comparison in repeated-measures microcomputer test batteries.

ASSOCIATIVE MEMORY (AM). This is a memory test (Underwood, Boruch, & Malmi, 1977) that requires the participant to view five sets of three letters that are numbered 1 to 5 and then to memorize this list. After an interval, successive trigrams are displayed and the participant is required to press the key of the number corresponding to that letter set. In previous research (Krause & Kennedy, 1980) this associative memory task was recommended for inclusion in a performance testing battery for environmental factors using the percent correct score.

GRAMMATICAL REASONING (GR). The Grammatical Reasoning Test (Baddeley, 1968) involves five grammatical transformations on statements about the relationship between two letters A and B. The five transformations are: (1)

active versus passive construction, (2) true versus false statements, (3) affirmative versus negative phrasing, (4) use of the verb "precedes" versus the verb "follows," and (5) A versus B mentioned first. There are 32 possible items arranged in random order. The subject's task is to respond "true" or "false," depending on the verity of each statement. Performance is scored according to the number of transformations correctly identified. Grammatical Reasoning is described as measuring "higher mental processes" with reasoning, logic, and verbal ability, important factors in test performance (Carter, Kennedy, & Bittner, 1981). According to Bittner et al. (1986), Grammatical Reasoning "assesses an analytic cognitive neuropsychological function associated with the left hemisphere" (p. 699). Previous studies with Grammatical Reasoning, identified in Bittner et al. (1986), have indicated that the task is acceptable for use in repeated-measures research. Recent field testing with a microcomputer version of the task (Kennedy, Dunlap, Jones, Lane, & Wilkes, 1985; Kennedy, Wilkes, Lane, & Homick, 1985; Wilkes et al., 1986) have resulted in strong recommendations for inclusion of Grammatical Reasoning in repeated-measures microcomputer test batteries.

ANALYSES

The advantages associated with automated microbased testing have often been noted in the literature (Wilkes et al., 1986; Baker, Letz, & Fidler, 1985; Baker, Letz, Fidler, Shalat, Plantamura, & Lyndon, 1985; Thompson, & Wilson, 1982; Fletcher, 1978), but are not without some problems -- notably decreased reliability over more traditional presentation (Smith, Krause, Kennedy, Bittner, & Harbeson, 1983). However, the rich variety of potential response measures and the facilitation of repeated automated battery applications are two of the most obvious benefits. Moreover, the sheer quantity of data resulting from these advantages creates special problems and obligations. Self-administration could provide additional opportunities for unknown sources of variance. Therefore, data inspection and review procedures must be completed prior to statistical treatments. The primary purpose of these procedures is to surface data anomalies (e.g., reaction times which are too short; percent correct scores of 50%), and facilitate the selection of appropriate and representative scores for analyses. Summaries of each of the metric properties examined follow.

ANALYTIC APPROACH FOR A REPEATED-MEASURES PERFORMANCE BATTERY

Six psychometric theory criteria have been followed in the test battery work under this contract. Initially stability and reliability (criteria 1 and 2) were emphasized. As the nucleus of a battery became available, the concern became centered around the economy of time and so trials/time to stability as well as over all reliability efficiency (criteria 3 and 4) were added. Task ceiling and factor structures (criteria 5 and 6) have become a later focus and should be addressed in this and the next reports. These are listed below in greater detail.

1. STABILITY. Repeated-measures studies of environmental influences on performance require stable measures if changes in the treatment (i.e., the environment) are to be meaningfully related to changes in performance (Jones, 1970a). Of particular concern is the fact that a subject's scores may differ significantly over time due to measure instability. For example, the Jones

two-process theory of skill acquisition (Jones, 1970a, 1970b) maintains that the advancement of a skill involves an acquisition phase in which persons improve at different rates, and a terminal phase, in which persons reach or approximate their individual limits. The theory further implies that when the terminal phase is reached, scores will cease to deviate, despite additional practice. Unless tests have been practiced to this point of differential stability, the determination of changes in scores due to practice or some other variable would be impossible. Therefore, a stable test implies that the same thing is being consistently measured and an unstable test implies the converse. For example, in a study of the effects of a toxic substance, if scores on a performance test remained the same before or after exposure, and if the test were not differentially stable, it would not be possible to determine whether a decline in performance was masked by practice effects or whether there was no treatment effect. Only after differential stability is clearly and consistently established between subjects can the investigator place confidence in the adequacy of his measures.

2. TASK DEFINITION. Task definition is the average reliability of the stabilized task (Jones, 1980). Task Definition is obtained by averaging stable intertrial correlations. Higher average reliability improves power in repeated-measures studies when variances are constant. The lower the error within a measure the greater the likelihood that mean differences will be detected, provided variances are also well behaved. Therefore, tasks with low task definition are insensitive to such differences and are to be avoided. Because different tasks stabilize at different levels, task definition becomes an important criterion to task selection. Task definitions for different tests, however, cannot be directly compared without first standardizing tests for test length (i.e., reliability efficiency).

3. RELIABILITY EFFICIENCY. Test reliability is known to be influenced by test length (Guilford, 1954). Tests with longer administration times and/or more items maintain a reliability advantage over shorter test times. Test length must be equalized before meaningful comparisons can be made. A useful tool for making relative judgments is the reliability-efficiency, or standardized reliability, of the test (Kennedy, Carter, & Bittner, 1980). Reliability-efficiencies are computed by correcting the reliabilities of different tests to a common test length by use of the Spearman-Brown prophecy formula (Guilford, 1954, p. 354). Reliability-efficiency not only facilitates judgments concerning different tests, but also provides a means for comparing the sensitivity of one test with the sensitivity of another test.

4. STABILIZATION TIME. The evaluation of highly transitory changes in performance may be necessary when studying the effects of various treatments, drugs, or environmental stress. Good performance measures should quickly stabilize following short periods of practice without sacrificing metric qualities, and good performance measures should always be economical in terms of time. A task under consideration for environmental research must be represented in terms of the number of trials and/or the total amount of time necessary to establish stability. Stabilization time must be determined for the group means, standard deviations, and intertrial correlations (differential stability).

5. TASK CEILING. If all subjects asymptote at the maximum level of performance, then the task is said to have a ceiling (Jones, 1980). Ceilings are undesirable because they limit discrimination between subjects. When subjects perform equally well, except for random error, between-trial correlations fall to zero.

6. FACTOR RICHNESS. Following stability analyses and ratification, stable tests within a battery are subjected to factor analysis. Where sample size permits, factor structure is determined based on the principal factors method with squared multiple correlations as initial communality estimates, followed by varimax rotation. Factor extraction is terminated when eigenvalues dropped below unity.

In the present study data anomalies were surfaced by graphing performances for clusters of 3 to 5 subjects for all 10 trials of each test. As a result of these comparisons the following problems and corrections were identified: (a) a programming error in the Grammatical Reasoning test required that the number correct score be discarded. The decision to drop the number correct score therefore impacted the derived percent correct score; (b) a second programming error resulted in the nonadministration of the Nonpreferred-hand Tapping task to two left-handed subjects. As a result of the omission, no data on that test for the two subjects were entered; (c) atypical scores were observed for each subject on the first trial of the Number Comparison test which has subsequently been traced to a software error. Those scores were not analyzed. There were no other obvious anomalies.

The inspection and review process also aided in the selection of representative scores for analyses. Because several types of scores were recorded for each test (i.e., number correct, percent correct, number wrong, number omitted, response latency), all the scores were examined in an attempt to establish their ability to accurately describe performance. Many of these scores are derivatives of each other and therefore redundant, but their availability can be useful for diagnostic purposes in certain experimental paradigms.

Lastly, the Complex Counting tests (auditory and visual) were each represented by three different levels of complexity. Inspection of the data indicated that the low complexity auditory and visual tests suffered from ceiling effects, particularly as practice ensued, and too few data points per session for meaningful analyses. For this reason, only scores from the most complex (level three) of the Counting tests were included in the analyses.

RESULTS

Following the data inspection and review described above, stability analyses (Jones, 1980; Jones, Kennedy, & Bittner, 1981) were conducted. The group means, standard deviations and intertrial correlational matrices were calculated for each subtest. Group means and standard deviations were examined for evidence of test stabilization and intertrial correlations were assessed for evidence of correlational stability (i.e., differential stability, task definition, reliability efficiency).

MEAN AND STANDARD DEVIATION STABILITY. The means and standard deviations for the number correct, percent correct, and response latency scores appear in Table 3. Inspection of Table 3 indicates that the means for all measures stabilized between trials 2 to 5 and that the corresponding standard deviations achieve stability between trials 2 to 6. Percent correct scores permit easy comparison across tests and were greater than 92% on all occasions except for associative memory (53% -75%), a task which the subjects found difficult. Mean and Standard Deviation trial stability estimates for each subtest, across each measure (where appropriate) are summarized in Table 4. Inspection of Table 4 reveals that the different scores (i.e., number correct, percent correct, response latency) stabilize quickly.

TABLE 3. MEANS AND STANDARD DEVIATIONS

Subtests	Trials 1	2	3	4	5	6	7	8	9	10
PTAP(N)*	39 (10**)	42 (10)	43 (8)	43 (9)	44 (10)	43 (9)	43 (9)	45 (10)	44 (8)	45 (10)
RT1(RL)	338 (60)	301 (43)	285 (42)	290 (44)	281 (49)	285 (61)	279 (49)	284 (59)	298 (82)	294 (77)
ACTL(NC)	5 (2)	6 (1)	6 (1)	6 (2)	6 (2)	6 (2)	6 (2)	6 (1)	5 (2)	7 (1)
STM(PC)	98 (2)	97 (3)	97 (3)	97 (3)	97 (3)	96 (4)	96 (3)	96 (4)	96 (3)	96 (4)
STM(RL)	853 (181)	834 (175)	793 (196)	773 (182)	766 (177)	758 (187)	759 (174)	768 (244)	731 (153)	721 (144)
ACTM(NC)	4 (2)	4 (2)	4 (1)	5 (1)	5 (2)	4 (2)	5 (2)	5 (1)	4 (2)	5 (1)
NCP(NC)	*** ***	43 (7)	43 (7)	44 (8)	45 (8)	45 (9)	45 (9)	46 (9)	47 (8)	44 (13)
NCP(PC)	*** ***	95 (4)	92 (5)	93 (8)	93 (6)	92 (5)	92 (6)	93 (5)	95 (4)	93 (5)
NCP(RL)	607 (130)	582 (136)	556 (120)	554 (115)	527 (118)	521 (112)	526 (99)	521 (101)	522 (100)	516 (108)
ACTH(NC)	3 (2)	3 (1)	3 (1)	3 (1)	4 (1)	3 (2)	3 (2)	4 (1)	4 (2)	4 (1)
ACM(N)	78 (16)	90 (23)	95 (24)	99 (22)	104 (20)	103 (15)	103 (18)	109 (17)	111 (20)	109 (17)
RT2(RL)	441 (222)	372 (63)	351 (61)	348 (85)	346 (79)	351 (66)	331 (49)	327 (47)	337 (55)	335 (74)

* Codes: (N)=Number, (NC)=Number Correct, (PC)=Percent Correct
(RL)=Response Latency

** Standard Deviations in Parentheses

*** Trial 1 for Number Comparison not analyzed

TABLE 3. (continued)

Subtests	Trials 1	2	3	4	5	6	7	8	9	10
THTAP(N)	44 (10)	46 (11)	46 (11)	45 (9)	46 (10)	46 (10)	47 (10)	46 (10)	46 (12)	47 (11)
PC(NC)	86 (12)	89 (12)	92 (13)	92 (12)	96 (13)	95 (14)	96 (13)	97 (15)	97 (13)	97 (11)
PC(PC)	96 (3)	95 (4)	96 (3)	94 (3)	95 (3)	95 (4)	95 (3)	96 (3)	95 (3)	94 (4)
PC(RL)	1030 (257)	963 (211)	939 (207)	905 (173)	861 (180)	871 (195)	869 (162)	866 (217)	854 (179)	835 (169)
VCTR(NC)	6 (1)	6 (1)	6 (1)	6 (1)	6 (2)	6 (1)	7 (1)	7 (1)	6 (1)	6 (1)
AM(NC)	11 (4)	12 (3)	12 (5)	12 (5)	13 (4)	15 (4)	14 (3)	14 (5)	15 (4)	15 (4)
AM(PC)	53 (18)	60 (16)	60 (23)	62 (23)	67 (22)	74 (21)	71 (16)	69 (24)	73 (22)	75 (18)
AM(RL)	460 (106)	440 (118)	442 (113)	381 (121)	390 (92)	406 (87)	380 (91)	404 (109)	370 (100)	367 (96)
VCTM(NC)	5 (2)	5 (1)	5 (1)	5 (1)	4 (2)	5 (1)	6 (1)	5 (1)	5 (1)	5 (1)
GR(RL)	3175 (945)	3041 (932)	2817 (864)	2694 (739)	2679 (893)	2731 (870)	2587 (751)	2609 (683)	2715 (922)	2629 (774)
RT4(RL)	495 (83)	458 (102)	454 (86)	436 (82)	431 (114)	429 (92)	416 (90)	427 (74)	407 (79)	403 (81)
VCTL(NC)	4 (2)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)
NTAP(N)	34 (8)	37 (8)	37 (9)	38 (9)	38 (8)	38 (8)	39 (9)	38 (8)	39 (8)	39 (8)

* Codes: (N)=Number, (NC)=Number Correct, (PC)=Percent Correct
(RL)=Response Latency

** Standard Deviations in Parentheses

TABLE 4. STABILITY OF MEANS AND INTERTRIAL CORRELATIONS

Variable	Trial Means Stabilize	SD's	Trial of Differential Stability
Preferred Hand Tapping (PTAP)	2	2	1
Average Reaction Time 1 (RT1)	3	2	3
Auditory Counting Low NC (ACTL)	4	3	4
Short Term Memory NC (STMNC)	3	3	1
Short Term Memory PC (STMPC)	2	3	-*
Short Term Memory RL (STMRL)	3	4	1
Auditory Counting Medium NC(ACTM)	4	3	-
Number Comparison NC (NCNC)	2	4	3
Number Comparison PC (NCPC)	3	5	3
Number Comparison RL (NCRL)	5	3	3
Auditory Counting High NC (ACTH)	5	3	-
Air Combat Maneuvering (ACM)	5	4	3
Average Reaction Time 2 (RT2)	3	2	2
Two Hand Tapping (THTAP)	2	2	2
Pattern Comparison NC (PCNC)	5	3	3
Pattern Comparison PC (PCPC)	2	2	3
Pattern Comparison RL (PCRL)	4	5	2
Visual Counting Right NC (VCTR)	3	6	-
Associative Memory NC (AMNC)	5	3	5
Associative Memory PC (AMPC)	6	3	5
Associative Memory RL (AMRL)	4	5	5
Visual Counting Middle NC (VCTM)	3	3	-
Grammatical Reasoning RL (GRRL)	3	3	3
Average Reaction Time 4 (RT4)	2	3	2
Visual Counting Left NC (VCTL)	3	3	-
Nonpreferred Tapping (NTAP)	2	2	1

* These tests did not reach stability in the 10 trials

CORRELATIONAL STABILITY. The intertrial correlations for all subtests and appropriate scores may be examined in Appendix B. Correlational stability estimates for each subtest are summarized in Table 5. In general, both the number correct and response latency scores demonstrate differential stability by trial 2 to 6 (see Table 4). Two execeptions were noted, however, and include Reaction Time 1 (response latency) and Associative Memory (number correct), which, if they stabilized, did so after trial 5. These tests did not demonstrate intertrial correlational stability over the 10 trials. Comparison of the Task Definitions and Reliabilty Efficiencies for the number correct and response latency scores are also presented in Table 5. Task Definitions range from .61 (Air Combat Maneuvering) to .85 (Pattern Comparison) for number correct, and from .85 (Number Comparison and Sternberg) to .99 (Preferred and Nonpreferred Hand Tapping) for response latency with corresponding Reliability Efficiencies ranging from .70 to .89 and .89 to 1.00. These indicators demonstrate that, for stable subtests, the number correct and response latency scores across trial reliabilites are high and above criteria (i.e., $\underline{r} \geq .70$). In general, the reliabilities of the percent correct scores were not as impressive and were always lower than number correct or latency. Associative Memory, Pattern Comparison, Number Comparison and Short Term Memory did not give indications of Differential Stability for the percent correct scores. Furthermore, the Task Definitions and Reliability Efficiencies were correspondingly low. These results indicate that number correct and response latency scores should be considered as the "score of choice" for most of the evaluated subtests. These results are also consistent with the findings of Carter and Woldstad (1985) in which the Manikin subtest stability was assessed with both accuracy and log latency scores. Although in most cases the number correct and response latency scores proved to be purer and more viable measures, the percent correct score should not be dismissed from consideration. Percent correct scores are helpful in determining the legitimacy of a subject's test taking strategy. For example, a subject that randomly and rapidly presses true/false response keys could conceivably generate a higher number correct score than serious respondents. Simply guessing, however, would be reflected by a percent correct score not significantly different than $\underline{p}$ = .50. It is highly recommended that in cases where subject motivation and test taking strategy are questionable, the percent score should be closely examined.

TABLE 5. TASK DEFINITION ACTUAL (OBTAINED) AND PREDICTED FOR A THREE-MINUTE TEST FROM SPEARMAN-BROWN'S ADJUSTMENT

Variable Minutes	Obtained Task Definition	Predicted Three-Minute Reliability
Preferred Hand Tapping	.99	1.00
Average Reaction Time 1	.58	.67
Auditory Counting Low NC	.44	.32
Short-Term Memory NC	.80	.86
Short-Term Memory PC	- *	-
Short-Term Memory RL	.85	.89
Auditory Counting Medium NC	-	-
Number Comparison NC	.71	.91
Number Comparison PC	.54	.82
Number Comparison RL	.85	.96
Auditory Counting High NC	-	-
Air Combat Maneuvering	.61	.70
Average Reaction Time 2	.76	.83
Two Hand Tapping	.92	1.00
Pattern Comparison NC	.85	.89
Pattern Comparison PC	.50	.60
Pattern Comparison RL	.89	.92
Visual Counting Right NC	-	-
Associative Memory NC	.37	.54
Associative Memory PC	.37	.54
Associative Memory RL	.65	.79
Visual Counting Middle NC	-	-
Grammatical Reasoning (RL)	.86	.90
Average Reaction Time 4	.83	.88
Visual Counting Left NC	-	-
Nonpreferred Tapping	.99	1.00

*Tests that do not become stable cannot have task definition and relatedly no 3-minute reliability can be obtained.

Although clearly too small a sample in which to form firm conclusions, a factor analysis was conducted on the data to guide future battery development. In this analysis, four factors were surfaced. Factor 1 included reaction time and memory tasks (Short Term Memory, Number Comparison, Pattern Comparison). Factor 2 was made up of the Visual Counting and the Gramatical Reasoning task. The third factor was of the motor tests: the Tapping tests and Air Combat Manuevering. Factor 4 was the Auditory Counting, one of the tapping tests, one of the reaction time tests, and Associative Memory. The results were consistent with the two previous attempts (also with small samples) (Kennedy, Dunlap, Jones, Lane, & Wilkes, 1985; and Kennedy, Wilkes, Lane, & Homick, 1985) and with two larger studies (McCombs, Doll, Baltzley, & Kennedy, 1986; and Jones, 1987). In Table 6 the average correlations are shown. These are the correlations of all the tests.

TABLE 6. AVERAGED CORRELATIONS WITH ESTIMATED 3-MINUTE RELIABILITY

Percent Correct or Most Appropriate Score*

	STMPC	NCPC	PCPC	AMPC	PTAP	THTAP	NTAP	RT1	ACTL	ACTM	ACTH	ACM	RT2	RT4	VCTR	VCTM	VCTL
STMPC	--	34	58	21	-15	-16	-24	-11	22	16	3	-20	-6	-4	18	-1	17
NCPC		54	43	26	-23	-3	-16	-1	26	29	13	-18	6	1	35	26	20
PCPC			50	24	-10	-10	-21	-18	17	16	-2	-12	-4	-7	17	9	6
AMPC				37	8	25	5	-17	30	33	15	8	-15	-23	21	24	26
PTAP					99	50	86	-43	34	27	26	60	-40	-38	29	40	36
THTAP						92	50	-48	38	37	47	47	-36	-43	23	41	28
NTAP							99	-30	27	21	19	62	-28	-24	30	42	38
RT1								58	-45	-35	-31	-41	51	56	-14	-18	-17
ACTL									44	55	42	18	-20	-35	40	44	36
ACTM										--	58	14	-26	-30	24	45	33
ACTH											--	21	-35	-35	17	36	30
ACM												61	-35	-42	11	22	15
RT2													76	71	-13	-24	-24
RT4														83	-14	-24	-21
VCTR															--	65	55
VCTM																--	60
VCTL																	--

Response Latency or Most Appropriate Score

	PTAP	THTAP	NTAP	RT1	STMRL	NCRL	ACTL	ACTM	ACTH	ACM	RT2	PCRL	AMRL	GRRL	RT4	VCTR	VCTM	VCTL
PTAP	99	50	86	-43	-48	-19	34	27	26	60	-40	-47	-34	-35	-38	29	40	36
THTAP		92	50	-48	-42	-10	38	37	47	47	-36	-54	-34	-29	-43	23	41	28
NTAP			99	-30	-44	-21	27	21	19	62	-28	-42	-29	-32	-24	30	42	38
RT1				58	35	-2	-45	-35	-31	-41	51	30	16	18	56	-14	-18	-17
STMRL					85	63	-45	-36	-32	-47	48	77	32	53	64	-33	-49	-44
NCRL						85	-23	-9	-5	-29	24	69	32	39	45	-11	-24	-11
ACTL							44	55	42	18	-20	-29	1	-42	-35	40	44	36
ACTM								--	58	14	-26	-25	-1	-17	-30	24	45	33
ACTH									--	21	-35	-33	-5	-19	-35	17	36	30
ACM										--	-35	-54	-47	-20	-42	11	22	15
RT2											76	55	28	28	71	-13	-24	-24
PCRL												89	50	44	64	-13	-35	-24
AMRL													65	30	34	1	0	-4
GRRL														86	36	-50	-46	-48
RT4															83	-14	-24	-21
VCTR																--	65	55
VCTM																	--	60
VCTL																		--

Number Correct or Most Appropriate Score

	PTAP	THTAP	NTAP	RT1	STMNC	NCNC	ACTL	ACTM	ACTH	ACM	RT2	PCNC	AMNC	RT4	VCTR	VCTM	VCTL
PTAP	99	50	86	-43	38	7	34	27	26	60	-40	39	11	-38	29	40	36
THTAP		92	50	-48	31	3	38	37	47	47	-36	51	26	-43	23	41	28
NTAP			99	-30	33	12	27	21	19	62	-28	31	8	-24	30	42	38
RT1				58	-35	-14	-45	-35	-31	-41	51	-35	-18	56	-14	-18	-17
STMNC					80	45	49	38	31	38	-47	68	25	-61	36	46	44
NCNC						71	28	10	9	28	-28	46	27	-39	13	21	17
ACTL							44	55	42	18	-20	35	30	-35	40	44	36
ACTM								--	58	14	-26	26	33	-30	24	45	33
ACTH									--	21	-35	31	16	-35	17	36	30
ACM										61	-35	49	10	-42	11	22	15
RT2											76	-51	-16	71	-13	-24	-24
PCNC												85	33	-65	19	38	24
AMNC													37	-26	21	24	27
RT4														83	-14	-24	-21
VCTR															--	65	55
VCTM																--	60
VCTL																	--

*For some tasks (e.g., reaction time) only one score is meaningful.

DISCUSSION

The usual paradigm followed in studies of environmental stress and toxic agents exposes one or more subjects to the intervention, then compares the individual's score under the treated and nontreated conditions. However, implicit in such a design is that over and above the name of the test being the same, the behavioral element or construct being tapped must also be the same on each testing. It is well known that learning a task can entail skills and abilities which are different from those required to perform the task after it is well practiced (Ackerman & Schneider, 1984) even to the extent that different structures in the brain appear necessary for these two functions. Therefore, a chief requirement for any test employed to reveal change due to treatment is that it be stable when no treatments are applied. Such a requirement permits "attribution of effect" when changes are found. The two-process theory advanced by Jones (1970a) states that early in practice individuals may improve at different rates and only after these differences have disappeared does a task approach a level of stability which will permit its utility in a repeated-measures context. Provocative evaluations of stability shall be conducted not only for means and variance -- but for between session correlations, as well (Bittner et al., 1986; Jones, 1980). Only when a test demonstrates symmetry of the variance co-variance matrix (Campbell, & Stanley, 1963) is there assurance that neither the task nor the subject taking the test is changing (Alvares & Hulin, 1972). Another major criterion for test selection was that, if the test revealed individual differences, the retest reliability should be high (tests with these differences are acceptable, but virtually unknown). High reliability is desired because 1) low reliablity suggests insensitivity, and 2) experiments for sensitivity imply small numbers of subjects used repeatedly.

The field testing of the automated system indicates that the battery can be successfully self-administered over repeated applications, outside a research laboratory environment. The research director need only initially instruct the subjects in the use of the battery, establish testing protocol and properly motivate the individuals involved in the study. The importance of opening data collection to research free from environments cannot be ignored. However, while the results of this study appear "clean," the reliability of the scores on the last day are among the lowest of all the stable days. Additionally, three different types of scores have been compared including number correct, percent correct, and response latency. Comparative analyses indicates that number correct and response latency are the "purer" scores and the recommended scores of choice. There are, however, instances where the use of the percent correct score is recommended, such as when motivation may be low. Of the tests given from the original battery the longest to stabilize were Pattern Comparison, and Reaction Time 4, but these tests stabilized by trial 5. The tests that had not been administered previously were the auditory and visual complex counting tests, as well as Associative Memory and Air Combat Maneuvering which had been given in paper and pencil form (Kennedy, Wilkes, Lane, & Homick, 1985). More data should be collected on these tests in an experimental situation that utilizes more subjects as well as more trials.

Summaries of the results for each test given follows:

THE TAPPING SERIES. These tasks stabilized quickly and had high reliabilities for each of the three tests. The test itself taps motor ability and does not overlap much with the other tests. These tests are highly recommended for a battery.

THE REACTION TIME TESTS. These tests exhibited stability but lower reliabilities for 1-Choice Reaction Time (.58). Only the 4-Choice Reaction Time is recommended as it does have higher reliability (.83), and is very similar to the 1-Choice and 2-Choice Reaction Times.

GRAMMATICAL REASONING. Only the response latency was available for analysis, but Grammatical Reasoning did show high reliability and fairly rapid stability. This test is recommended for a battery.

ASSOCIATIVE MEMORY. Because this task required five trials to stabilize and failed to meet the required (.70) reliability criterion it is not recommended.

PATTERN COMPARISON. This test was also slower to stabilize, it required five trials for Number Correct to become stable, but exhibited high reliability in Number Correct (.85) and Response Latency (.89). Therefore, this test is tentatively recommended but further study is needed.

AIR COMBAT MANEUVERING. This test was slow to stabilize (trial 5) and reliability that when corrected for attenuation is .70. The test itself does seem to be a "motivating" task for subjects. Air Combat Maneuvering is recommended for further study.

NUMBER COMPARISON. Number Comparison stabilizes within three trials and exhibits acceptable reliability. Additionally, this test may bolster the mathematical factors of the test battery that was found weak when compared to WAIS (Kennedy, Wilkes, Lane, & Homick, 1985). This task is highly recommended.

SHORT-TERM MEMORY. This test stabilizes quickly and has high reliability for number correct and response latency. It is also highly recommended for use in a test battery.

AUDITORY AND VISUAL COUNTING. Both of these versions of the counting tests showed ceiling effects in the first and second levels. The third level appeared to stablize but did not show differential stability. The tests were therefore not reliable. These tests appear to tap a separate factor from the factor analysis and the test seems to warrant further study, but is not recommended at this time.

In conclusion, nine of the tests stabilize and there is a menu providing test stabilization (in seconds) for each test. Total test time is 47.2 minutes, but the factor analysis would recommend for a short performance battery that only five of the tests be used -- the Tapping series, Number Comparison, Short Term Memory, Grammatical Reasoning, and 4-Choice Reaction Time. These should reveal three factors: (1) cognition, (2) processing quickness, and (3) motor. All of these would stabilize in 24 minutes, or approximately two 12-minute sessions.

NASA sponsored research has contributed greatly to the expanding body of knowledge concerning human performance testing. Identification of factor structure, the importance of different scores associated with each subtest and the number of metrically acceptable tests available for research purposes represent a few of the more important advances. NEC computer capabilities no longer suit the commensurate data collection needs associated with the derived benefits. In an attempt to keep pace, the decision was made to upgrade hardward capabilities with the purchase of new Zenith portable computers. This timely change in hardware has facilitated the incorporation of previous findings into a highly advanced and sophisticated microbased testing system.

REFERENCES

Ackerman, P. L., & Schneider, W. (1984, August). Individual differences in automatic and controlled information processing (Rep. No. HARL-ONR-8401). Champaign, IL: Human Attention Research Laboratory.

Alvares, K. M., & Hulin, C. L. (1972). Two explanations of temporal changes in ability-skill relationships: A literature review and theoretical analysis. Human Factors, 12, 295-308.

American Psychological Association. (1982). Ethical principles to the conduct of research with human participants. Washington, DC: Author.

Baddeley, A. D. (1968). A three-minute reasoning test based on grammatical transformation. Psychonomic Science, 10, 341-342.

Baker, E. L., Letz, R. E., & Fidler, A. T. (1985). A neurobehavioral evaluation system for occupational and environmental epidemiology: Rationale, methodology, and pilot study results. Journal of Occupational Medicine, 25, 125-130.

Baker, E. L., Letz, R. E., Fidler, A. T., Shalat, S., Plantamura, D., & Lyndon, M. (1985). A computer-based neurobehavioral evaluation system for occupational and environmental epidemiology methodology and validation studies. Neurobehavioral Toxicology & Teratology, 7, 369-377.

Bittner, A. C., Jr., Carter, R. C., Kennedy, R. S., Harbeson, M. M., & Krause, M. (1986). Performance Evaluation Tests for Environmental Research (PETER): Evaluation of 114 measures. Perceptual and Motor Skills, 63, 683-708.

Bittner, A. C., Jr., Carter, R. C., Krause, M., Kennedy, R. S., & Harbeson, M. M. (1983, October). Performance tests for repeated measures: Moran and computer batteries. Proceedings of the 26th Annual Meeting of the Human Factors Society (pp. 747-751). Seattle, WA: Human Factors Society.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Carter, R. C., Kennedy, R. S., & Bittner, A. C., Jr. (1981). Gramatical reasoning: A stable performance yardstick. Human Factors, 23, 587-591.

Carter, R. C., Krause, M., & Harbeson, M. M. (1986). Beware the reliability of slope scores for individuals. Human Factors, 28, 673-683.

Carter, R. C., & Sbisa, H. E. (1982). Human performance tests for repeated measurements; Alternate forms of eight tests by computer (Research Report No. NBDL-82R003). New Orleans: Naval Biodynamics Lab. (NTIS No. AD A115021)

Carter, R. C., & Wolstad, J. C. (1985). Repeated measurements of spatial ability with the Manikin test. Human Factors, 27(2), 209-219.

Donders, F. C. (1968). On the speed of mental processes (W. G. Koster, Trans.). Acta Psychologica, 30, 412-431.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976, August). Manual for kit of factor-referenced cognitive tests (Office of Naval Research No. N00014-71-C-0117). Princeton, NJ: Educational Testing Service.

Essex Corporation. (1985). Automated performance test system. Orlando, FL: Brochure.

Fletcher, B. (1978). Automated psychological testing: An alternative to Denner. (1977). Bulletin of the British Psychological Society, 31, 223.

Guilford, J. P. (1954). Psychometric methods (2nd ed.). New York: McGraw-Hill.

Jerison, H. J. (1955, December). Effect of a combination of noise and fatigue on a complex counting task (WADC TR-55-360). Wright-Patterson Air Force Base, OH: Wright Air Development Center, Air Research and Development Command, United States Air Force.

Jones, M. B. (1970a). A two-process theory of individual differences in motor learning. Psychological Review, 77(4), 353-360.

Jones, M. B. (1970b). Rate and terminal process in skill acquisition. American Journal of Psychology, 83(2), 222-236.

Jones, M. B. (1980). Stabilization and task definition in a performance test battery (Final Rep., Contract N00203-79-M-5089). New Orleans, LA: U.S. Naval Aerospace Medical Research Laboratory. (AD A099987)

Jones, M. B. (1987). Factor and regression analyses of selected tests from the APTS and PAB batteries. Unpublished manuscript. Orlando, FL: Essex Corporation.

Jones, M. B., Kennedy, R. S., & Bittner, A. C., Jr. (1980). Video games and convergence or divergence with practice. Proceedings of the 7th Psychology in the DoD Symposium (pp. 465-470). Colorado, Springs, CO: USAF Academy.

Jones, M. B., Kennedy, R. S., & Bittner, A. C., Jr. (1981). A video game for performance testing. American Journal of Psychology, 94, 143-152.

Kennedy, R. S., & Bittner, A. C., Jr. (1980). Performance Evaluation Tests for Environmental Research (PETER): Complex counting. Aviation, Space, and Environmental Medicine, 51, 142-144.

Kennedy, R. S., Bittner, A. C., Jr., Harbeson, M. M., & Jones, M. B. (1981). Perspectives in performance evaluation tests for environmental research: Collected papers (NBDL-80R004). New Orleans, LA: Naval Biodynamic Laboratory. (AD A111180)

Kennedy, R. S., Bittner, A. C., Jr., Harbeson, M. M., & Jones, M. B. (1982). Television computer games: A new look in performance testing. Aviation Space, and Environmental Medicine, 53, 49-53.

Kennedy, R. S., Bittner, A. C., & Jones, M. B. (1981). Video-game and conventional tracking. Perceptual and Motor Skills, 53, 310.

Kennedy, R. S., Carter, R. C., & Bittner, A. C., Jr. (1980). A catalogue of performance evaluation tests for environmental research. Proceedings of the Human Factors Society (pp. 334-348). Los Angeles, CA: Human Factors Society.

Kennedy, R. S., Dunlap, W. P., Jones, M. B., Lane, N. E., & Wilkes, R. L. (1985). Portable human assessment battery: Stability, reliability, factor structure, and correlation with tests of intelligence (Tech. Rep. No. EOTR-86-4). Orlando, FL: Essex Corporation.

Kennedy, R. S., Dunlap, W. P., Wilkes, R. L., & Lane, N. E. (1985, October). Development of a protable computerized performance test system. Paper presented at the 27th Annual Meeting of the Military Testing Association. San Diego, CA.

Kennedy, R. S., Wilkes, R. L., Lane, N. E., & Homick, J. L. (1985). Preliminary evaluation of microbased repeated-measures testing system (Report No. EOTR-85-1). Orlando, FL: Essex Corporation.

Klein, R., & Armitage, R. (1979). Rhythms in human performanc: 1 1/2-hour oscillations in cognitive style. Science, 204, 1326-1328.

Krause, M., & Bittner, A. C., Jr. (1982). Repeated measures on a choice reation time task (Res. Rep. No. NBDL-82R006). New Orleans: Naval Biodynamics Laboratory. (AD A121904)

Krause, M., & Kennedy, R. S. (1980). Performance Evaluation Tests for Environmental Research (PETER): Interference susceptibility test. Proceedings of the 7th Psychology in the DoD Symposium (pp. 459-464). Colorado Springs, CO: USAF Academy.

McCombs, B. L., Doll, R. E., Baltzley, D. R., & Kennedy, R. S. (1986, December). Predictive validities of primary motivation scales for reenlistment decision-making (Contract No. MDA 903-86-C-0114). Orlando, FL: Essex Corporation.

NEC Home Electronics (USA), Inc. (1983). NEC PC-8201A users guide. Tokyo, Japan: Nippon Electric Co. Ltd.

Smith, M. G., Krause, M., Kennedy, R. S., Bittner, A. C., Jr., & Harbeson, M. M. (1983). Performance testing with microprocessors: Mechanization is not implementation. Proceedings of the 27th Annual Meeting of the Human Factors Society (pp. 674-678). Norfolk, VA: Human Factors Society.

Sternberg, S. (1966). High-speed scanning in human memory. Science, 153, 652-654.

Thompson, J. A., & Wilson, S. L. (1982). Automated psychological testing. Journal of Man-Machine Studies, 17, 279-289.

Tombari, M. L., Fitzpatrick, S. J., & Childress, W. (1985). Using computers as contingency managers in self-monitoring interventions: A case study. Computers in Human Behavior, 1, 75-82.

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1977, May). The composition of episodic memory (ONR No. N00014-76-C-0270). Evanston, IL: Northwestern University. (AD A040696)

Wilkes, R. L., Kennedy, R. S., Dunlap, W. P., & Lane, N. E. (1986). Stability, reliability, and cross-mode correlations of tests in a recommended 8-minute performance assessment battery (Tech. Rep. No. EOTR-86-4 for NASA Contract No. NAS9-17326). Orlando, FL: Essex Corporation.

data shall not be emphasized and scores shall be presented in a positive manner. Furthermore, all subjects shall be informed that the performance and IQ test are the focus of the study as opposed to the subjects themselves. Upon completion of the study summarized results shall be made available to all participants. A more detailed description of the general procedures may be found in the publication <u>Development of a Portable Computerized Test System</u> (Wilkes, Kennedy, Dunlap & Lane, 1985).

<u>Submitted for Review</u>
R. Wilkes
January, 1987

<u>Reviewing Committee</u>	<u>Signatures</u>
Dr. Thomas J. Clifford (Biology)	
Dr. Gerald E. Nelson (Geology)	
Dr. James W. O'Neill (History)	James W. O'Neill
Mrs. Jeanine Jones (President, Casper College Association)	

<u>Funding Agency</u>
Department of Defense, Spring 1987.

Repectfully submitted,

Robert L. Wilkes

Robert L. Wilkes

jlr

data shall not be emphasized and scores shall be presented in a positive manner. Furthermore, all subjects shall be informed that the performance and IQ test are the focus of the study as opposed to the subjects themselves. Upon completion of the study summarized results shall be made available to all participants. A more detailed description of the general procedures may be found in the publication Development of a Portable Computerized Test System (Wilkes, Kennedy, Dunlap & Lane, 1985).

Submitted for Review
R. Wilkes
January, 1987

Reviewing Committee	Signatures
Dr. Thomas J. Clifford (Biology)	
Dr. Gerald E. Nelson (Geology)	
Dr. James W. O'Neill (History)	
Mrs. Jeanine Jones (President, Casper College Association)	

Funding Agency
Department of Defense, Spring 1987.

Repectfully submitted,

Robert L. Wilkes

jlr

ORIGINAL PAGE IS
OF POOR QUALITY

data shall not be emphasized and scores shall be presented in a positive manner. Furthermore, all subjects shall be informed that the performance and IQ test are the focus of the study as opposed to the subjects themselves. Upon completion of the study summarized results shall be made available to all participants. A more detailed description of the general procedures may be found in the publication <u>Development of a Portable Computerized Test System</u> (Wilkes, Kennedy, Dunlap & Lane, 1985).

<u>Submitted for Review</u>
R. Wilkes
January, 1987

<u>Reviewing Committee</u>	<u>Signatures</u>
Dr. Thomas J. Clifford (Biology)	Thomas J. Clifford
Dr. Gerald E. Nelson (Geology)	
Dr. James W. O'Neill (History)	
Mrs. Jeanine Jones (President, Casper College Association)	

<u>Funding Agency</u>
Department of Defense, Spring 1987.

Repectfully submitted,

Robert L. Wilkes

Robert L. Wilkes

jlr

APPENDIX B

Intertrial Correlations

Trial Number

Preferred Hand Tapping

	1	2	3	4	5	6	7	8	9	10
1	1.0	.99	.99	.99	.98	.99	.99	.98	.98	.98
2		1.0	.99	.99	.99	.99	.99	.99	.99	.99
3			1.0	.99	.99	.99	.99	.99	.99	.99
4				1.0	.99	.99	.99	.99	.99	.99
5					1.0	.99	.99	.99	.99	.99
6						1.0	.99	.99	.99	.99
7							1.0	.99	.99	.99
8								1.0	.99	.99
9									1.0	.99
10										1.0

Reaction Time 1

	1	2	3	4	5	6	7	8	9	10
1	1.0	.00	-.04	.31	.42	.28	.29	.25	.16	.25
2		1.0	.57	.57	.65	.52	.38	.55	.66	.44
3			1.0	.83	.65	.80	.79	.68	.26	.76
4				1.0	.79	.93	.69	.81	.32	.77
5					1.0	.81	.56	.80	.44	.61
6						1.0	.75	.93	.42	.88
7							1.0	.70	.44	.87
8								1.0	.55	.83
9									1.0	.43
10										1.0

Auditory Count 1 Tone

	1	2	3	4	5	6	7	8	9	10
1	1.0	-.17	.08	.31	.63	.52	.32	.44	-.18	.05
2		1.0	.44	.56	.26	.09	.54	.48	.44	.74
3			1.0	.57	.45	.36	.32	.52	.10	.72
4				1.0	.72	.60	.56	.86	.13	.81
5					1.0	.84	.40	.84	.07	.52
6						1.0	.44	.68	.10	.43
7							1.0	.47	.28	.65
8								1.0	.16	.66
9									1.0	.16
10										1.0

Trial Number

Short-Term Memory NC

	1	2	3	4	5	6	7	8	9	10
1	1.0	.89	.81	.81	.84	.87	.81	.74	.85	.66
2		1.0	.85	.88	.91	.87	.85	.64	.90	.69
3			1.0	.93	.93	.89	.81	.76	.82	.86
4				1.0	.89	.89	.83	.76	.81	.75
5					1.0	.87	.77	.72	.83	.75
6						1.0	.86	.82	.75	.79
7							1.0	.79	.76	.76
8								1.0	.55	.73
9									1.0	.56
10										1.0

Short-Term Memory PC

	1	2	3	4	5	6	7	8	9	10
1	1.0	.48	.42	.71	.59	.59	.28	.70	.34	.59
2		1.0	.65	.50	.62	.39	.55	.55	.34	.62
3			1.0	.62	.55	.22	.44	.39	.15	.74
4				1.0	.56	.34	.53	.65	.31	.74
5					1.0	.32	.55	.60	.08	.71
6						1.0	.54	.69	.10	.26
7							1.0	.55	.06	.41
8								1.0	.53	.72
9									1.0	.40
10										1.0

Short-Term Memory RL

	1	2	3	4	5	6	7	8	9	10
1	1.0	.88	.89	.86	.88	.83	.87	.72	.89	.87
2		1.0	.91	.89	.91	.92	.85	.52	.93	.78
3			1.0	.97	.98	.97	.91	.73	.92	.87
4				1.0	.97	.96	.94	.74	.88	.82
5					1.0	.94	.91	.73	.92	.83
6						1.0	.87	.64	.90	.84
7							1.0	.78	.87	.84
8								1.0	.53	.77
9									1.0	.80
10										1.0

Trial Number

Auditory Count 2 Tones

	1	2	3	4	5	6	7	8	9	10
1	1.0	.03	.31	.49	.30	.19	.36	.12	.42	.46
2		1.0	.41	.33	.04	.67	.26	.50	.17	.52
3			1.0	.75	.59	.68	.76	.31	.00	.39
4				1.0	.59	.42	.72	.29	.28	.52
5					1.0	.44	.76	.37	.21	-.01
6						1.0	.63	.42	.11	.44
7							1.0	.32	.27	.35
8								1.0	.02	.27
9									1.0	.52
10										1.0

Number Comparison NC

	1	2	3	4	5	6	7	8	9	10
1	1.0	*	*	*	*	*	*	*	*	*
2		1.0	.76	.62	.73	.75	.68	.65	.60	.48
3			1.0	.77	.84	.86	.74	.80	.85	.57
4				1.0	.78	.79	.87	.83	.81	.80
5					1.0	.81	.86	.86	.84	.59
6						1.0	.85	.83	.91	.71
7							1.0	.89	.88	.76
8								1.0	.89	.76
9									1.0	.72
10										1.0

Number Comparison PC

	1	2	3	4	5	6	7	8	9	10
1	1.0	*	*	*	*	*	*	*	*	*
2		1.0	.56	.56	.75	.35	.35	.38	.03	.47
3			1.0	.70	.73	.50	.46	.42	.22	.49
4				1.0	.68	.66	.79	.60	.51	.76
5					1.0	.56	.58	.47	.41	.62
6						1.0	.59	.65	.52	.55
7							1.0	.63	.68	.51
8								1.0	.66	.56
9									1.0	.45
10										1.0

* Trial 1 of Number Comparison deleted due to software error.

Trial Number

Number Comparison RL

	1	2	3	4	5	6	7	8	9	10
1	1.0	.59	.48	.50	.52	.54	.54	.51	.49	.48
2		1.0	.89	.71	.89	.85	.79	.63	.65	.86
3			1.0	.85	.89	.92	.86	.75	.83	.89
4				1.0	.82	.82	.91	.81	.80	.86
5					1.0	.85	.89	.76	.77	.93
6						1.0	.85	.79	.87	.86
7							1.0	.91	.85	.89
8								1.0	.91	.77
9									1.0	.77
10										1.0

Auditory Count 3 Tones

	1	2	3	4	5	6	7	8	9	10
1	1.0	.36	.54	.27	.64	.46	.74	.21	.40	.27
2		1.0	.41	.53	.49	.56	.24	.32	.38	.04
3			1.0	.62	.60	.55	.46	.00	.35	.05
4				1.0	.63	.58	.37	.30	.51	.18
5					1.0	.61	.61	.31	.72	.24
6						1.0	.51	.46	.44	-.12
7							1.0	.28	.67	.44
8								1.0	.32	.39
9									1.0	.28
1.00										1.0

Air Combat Manuvering

	1	2	3	4	5	6	7	8	9	10
1	1.0	.67	.76	.34	.60	.31	.60	.37	.31	.41
2		1.0	.77	.49	.68	.44	.46	.29	.33	.35
3			1.0	.65	.77	.77	.69	.58	.49	.60
4				1.0	.76	.60	.60	.72	.64	.57
5					1.0	.61	.85	.76	.76	.69
6						1.0	.52	.73	.68	.71
7							1.0	.69	.73	.75
8								1.0	.89	.84
9									1.0	.84
10										1.0

Trial Number

	1	2	3	4	5	6	7	8	9	10
Reaction Time 2										
1	1.0	.58	.67	.35	.62	.43	.48	.73	.69	.34
2		1.0	.89	.83	.84	.77	.73	.73	.78	.71
3			1.0	.87	.91	.81	.83	.75	.86	.78
4				1.0	.90	.80	.75	.63	.82	.94
5					1.0	.86	.85	.77	.88	.89
6						1.0	.90	.61	.77	.82
7							1.0	.74	.80	.79
8								1.0	.85	.65
9									1.0	.85
10										1.0

	1	2	3	4	5	6	7	8	9	10
Two-Hand Tapping										
1	1.0	.87	.97	.95	.93	.94	.94	.91	.95	.95
2		1.0	.87	.90	.90	.96	.85	.88	.88	.81
3			1.0	.97	.94	.92	.96	.91	.95	.94
4				1.0	.96	.94	.95	.91	.94	.92
5					1.0	.92	.96	.88	.95	.89
6						1.0	.90	.93	.93	.90
7							1.0	.87	.94	.92
8								1.0	.91	.90
9									1.0	.91
10										1.0

	1	2	3	4	5	6	7	8	9	10
Pattern Comparison NC										
1	1.0	.79	.87	.81	.82	.76	.79	.68	.81	.76
2		1.0	.78	.85	.86	.87	.70	.82	.78	.81
3			1.0	.89	.86	.86	.88	.84	.91	.91
4				1.0	.92	.92	.84	.85	.87	.88
5					1.0	.90	.81	.82	.84	.87
6						1.0	.82	.92	.88	.96
7							1.0	.79	.86	.84
8								1.0	.93	.96
9									1.0	.94
10										1.0

Trial Number

Pattern Comparison PC

	1	2	3	4	5	6	7	8	9	10
1	1.0	.71	.51	.54	.67	.30	.11	.44	.48	.51
2		1.0	.47	.62	.47	.47	.27	.64	.47	.39
3			1.0	.53	.48	.67	.47	.29	.77	.65
4				1.0	.30	.58	.32	.56	.55	.50
5					1.0	.33	.31	.44	.58	.31
6						1.0	.74	.72	.75	.76
7							1.0	.56	.55	.69
8								1.0	.47	.53
9									1.0	.73
10										1.0

Pattern Comparison RL

	1	2	3	4	5	6	7	8	9	10
1	1.0	.86	.94	.88	.88	.90	.83	.78	.91	.93
2		1.0	.86	.89	.91	.96	.77	.89	.85	.91
3			1.0	.89	.91	.90	.82	.83	.93	.94
4				1.0	.95	.92	.82	.83	.86	.91
5					1.0	.94	.83	.83	.88	.95
6						1.0	.84	.92	.90	.97
7							1.0	.80	.88	.87
8								1.0	.91	.91
9									1.0	.94
10										1.0

Visual Count 1 Right Screen Cue

	1	2	3	4	5	6	7	8	9	10
1	1.0	.54	.67	.83	.19	.74	.63	.65	.56	.66
2		1.0	.76	.59	.48	.37	.65	.57	.61	.52
3			1.0	.59	.45	.43	.64	.51	.45	.65
4				1.0	.02	.49	.81	.37	.51	.44
5					1.0	.23	.29	.55	.61	.27
6						1.0	.27	.66	.61	.27
7							1.0	.25	.71	.24
8								1.0	.68	.83
9									1.0	.41
10										1.0

Trial Number

Associative Memory NC

	1	2	3	4	5	6	7	8	9	10
1	1.0	.73	.57	.15	-.02	-.23	.03	.30	-.09	-.24
2		1.0	.55	.45	.22	.03	.17	.46	-.03	-.05
3			1.0	.66	.48	.20	.45	.69	.35	.23
4				1.0	.50	.47	.33	.53	.35	.25
5					1.0	.57	.79	.70	.53	.56
6						1.0	.56	.56	.63	.70
7							1.0	.66	.51	.62
8								1.0	.38	.52
9									1.0	.73
10										1.0

Associative Memory PC

	1	2	3	4	5	6	7	8	9	10
1	1.0	.72	.57	.16	.00	-.23	.04	.32	-.09	-.24
2		1.0	.52	.47	.22	.06	.10	.44	-.05	-.07
3			1.0	.68	.50	.21	.42	.70	.35	.23
4				1.0	.52	.51	.35	.56	.36	.26
5					1.0	.56	.78	.70	.54	.57
6						1.0	.53	.57	.63	.70
7							1.0	.65	.51	.62
8								1.0	.39	.52
9									1.0	.73
10										1.0

Associative Memory RL

	1	2	3	4	5	6	7	8	9	10
1	1.0	.76	.65	.37	.67	.53	.56	.75	.69	.62
2		1.0	.57	.44	.67	.71	.74	.75	.70	.83
3			1.0	.24	.84	.68	.74	.55	.90	.64
4				1.0	.50	.19	.20	.19	.36	.38
5					1.0	.79	.84	.56	.88	.76
6						1.0	.91	.56	.77	.81
7							1.0	.61	.80	.81
8								1.0	.66	.51
9									1.0	.71
10										1.0

Trial Number

Visual Count 2 Middle Screen Cue

	1	2	3	4	5	6	7	8	9	10
1	1.0	.38	.75	.70	.44	.76	.77	.10	.65	.60
2		1.0	.16	.22	.29	.26	.41	.32	.18	.12
3			1.0	.72	.51	.64	.51	.18	.62	.39
4				1.0	.63	.74	.63	.20	.71	.49
5					1.0	.54	.58	.06	.49	.55
6						1.0	.80	−.13	.60	.45
7							1.0	−.18	.56	.58
8								1.0	.30	.00
9									1.0	.35
10										1.0

Grammatical Reasoning Response Latency*

	1	2	3	4	5	6	7	8	9	10
1	1.0	.91	.79	.86	.80	.94	.86	.90	.93	.74
2		1.0	.77	.79	.75	.93	.84	.88	.91	.65
3			1.0	.96	.96	.79	.98	.87	.84	.84
4				1.0	.96	.88	.96	.87	.91	.84
5					1.0	.79	.96	.85	.85	.87
6						1.0	.84	.87	.96	.70
7							1.0	.89	.90	.82
8								1.0	.90	.90
9									1.0	.75
10										1.0

Reaction Time 4

	1	2	3	4	5	6	7	8	9	10
1	1.0	.72	.86	.68	.71	.60	.65	.77	.66	.55
2		1.0	.89	.89	.93	.91	.80	.89	.84	.81
3			1.0	.85	.88	.78	.78	.86	.86	.74
4				1.0	.90	.96	.94	.83	.90	.95
5					1.0	.89	.90	.84	.88	.84
6						1.0	.89	.83	.84	.95
7							1.0	.77	.83	.92
8								1.0	.86	.79
9									1.0	.90
10										1.0

*Only the reponse latency available.

Trial Number

	1	2	3	4	5	6	7	8	9	10

Visual Count 3 Left Screen Cues

	1	2	3	4	5	6	7	8	9	10
1	.0	.31	.48	.76	.50	.26	.47	.73	.40	.05
2		1.0	.34	.38	.02	.22	.19	.25	.41	.38
3			1.0	.59	.57	.03	.47	.30	.62	.03
4				1.0	.49	.52	.67	.62	.70	.26
5					1.0	-.08	.67	.23	.43	.11
6						1.0	.15	.45	.26	.03
7							1.0	.26	.68	.50
8								1.0	.30	-.18
9									1.0	.41
10										1.0

Nonpreferred Tapping

	1	2	3	4	5	6	7	8	9	10
1	1.0	.99	.99	.99	.99	.99	.99	.99	.99	.99
2		1.0	.99	.99	.99	.99	.99	.99	.99	.99
3			1.0	.99	.99	.99	.99	.99	.99	.99
4				1.0	.99	.99	.99	.99	.99	.99
5					1.0	.99	.99	.99	.99	.99
6						1.0	.99	.99	.99	.99
7							1.0	.99	.99	.99
8								1.0	.99	.99
9									1.0	.99
10										1.0