CR-172039

STABILITY, RELIABILITY, AND CROSS-MODE CORRELATIONS OF TESTS IN A RECOMMENDED 8-MINUTE PERFORMANCE ASSESSMENT BATTERY

ESSEX ORLANDO TECHNICAL REPORT EOTR 86-4

R. L. Wilkes, R. S. Kennedy,
W. P. Dunlap, and N. E. Lane

1985-86 WORK EFFORT SUMMARY
NASA CONTRACT NO. NAS 9-17326

Submitted by:

Essex Corporation
1040 Woodcock Road, Suite 227
Orlando, Florida 32803

April 1986

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

ABSTRACT

Introduction. A need exists for an automated performance test system to study drugs, agents, treatments, and stresses of interest to the aviation, space, and environmental medical community. The ethics and pragmatics of such assessment demand that repeated-measures in small groups of subjects become the customary research paradigm. In such cases, test stability, reliability-efficiency, and the underlying structure of a test battery take on extreme significance; in a previously conducted program of study, 80% of 150 tests studied failed to meet minimum metric requirements. The purpose of the present study is to evaluate tests for inclusion in the NASA-sponsored Automated Performance Test System (APTS).

Methods. Twenty-one subjects were tested over 10 replications with tests previously identified as "good" candidates for repeated-measures research. The tests were concurrently administered in paper-and-pencil (marker battery) and microcomputer modes. Performance scores for the two modes were compared.

Results. Data from trials 1-10 were examined for indications of test stability and reliability. Nine of the 10 APT System tests achieved stability. Reliabilities were generally high ($r \geq .707$). Cross-correlations of microbased tests with traditional paper-and-pencil versions revealed similarity of content within tests in the different modes, and implied at least three cognitive and two motor factors.

Conclusions. This portable, inexpensive, rugged, computerized battery of tests is recommended for use in repeated-measures studies of environmental and drug effects on performance. Identification of other tests compatible with microcomputer testing and potentially capable of tapping previously unidentified factors is recommended. Documentation of APTS sensitivity to environmental agents is available from more than a dozen facilities and is reported briefly. Continuation of such validation remains critical in establishing the efficacy of APTS tests.

INTRODUCTION

Performance and Exotic Environments

Many work environments entail exposure to unusual and atypical stressors; military and space environments often combine these agents. In manned space flight the success of the mission is contingent upon the efforts of a limited number of critical individuals. These highly trained and skilled workers are called upon to continuously perform complex and demanding tasks. The nature of the work and the setting in which it occurs further demand that all tasks be carried to completion, virtually error free. Ability to quickly and accurately process information, generate correct decisions, and perform complex tasks forms the basis for success. Of obvious importance to the overall manned space effort are the identification of factors that degrade performance and the systematic quantification of the deleterious effects associated with these factors. Conditions which could be encountered in space flight which are known to adversely affect performance in other settings include weightlessness (Nicogassian & Parker, 1982), motion (McCauley & Kennedy, 1976), fatigue and sleep loss (Woodward & Nelson, 1976; Kiziltan, 1985), hypoxia (Bandaret & Burse, 1984), generalized stress (Lazarus & Cohen 1977; Lazarus & Launier, 1978), and noise (Poulton, 1978). The extent of the influence of these factors on critical job performance in space flight is unknown; however, research in more temporally based environments implies that such effects occur (Christensen & Talbot, 1986). Thus far, reports from space travelers aloft indicate that mission requirements have continued to be met, but future populations of space travelers may not be so well trained, and can be expected to include the casual passenger. Remedies typically employed in the relief of discomfort associated with such agents may prove inappropriate or even counterproductive in the space setting. For example, pharmacological treatments effective in remediating other forms of motion sickness (McCauley, Royal, Shaw, & Schmitt, 1979) have been recommended to alleviate symptoms accompanying the Space Adaptation Syndrome (SAS). However, the effects of these drugs on task performance during space flight has not been assessed. Although many factors have been identified that lead to performance decrements in more typical work environments, it would be presumptuous to assume that all such agents associated with the space flight environment have been identified. The exotic nature of space flight virtually insures that previous unknowns will eventually surface and exert their effects. Understanding of the nature and effect of these unknowns remains an important challenge to the space effort.

Performance Measurement Applications

Exposure of humans to exotic environments, drugs, and other treatments brings with it the requirement to determine whether and to what extent performance and well being are affected. A need exists for a standardized, automated, performance test battery to examine such effects. The battery should have tests which are stable, sensitive, and related to the tasks to be performed under operational conditions. The Army (Thorne, Genser, Sing, & Hegge, 1983; Bandaret & Burse, 1984), Air Force (O'Donnell, 1981; Christal, 1981), and the Navy (Kennedy & Bittner, 1978), all have programs of study. Governmental agencies as well as the military are currently involved in performance measurement development. Under NSF sponsorship, Kennedy, Dunlap,

Wilkes, and Lane (1985b) have related performance on a microcomputer battery to global measures of intelligence. Also, the Appletox program, sponsored by the Environmental Protection Agency (EPA), has developed an automated test battery to detect the effects of toxic substances on human performance (Gullion & Eckerman, 1985, in press). The primary test medium is an Apple II microcomputer. Tests identified by the cognitive experimental approach of J. B. Carroll (Carroll, 1980) have been selected for evaluation. More tasks are in process, some data have been collected, and refinement of tasks and technical equipment is ongoing (Eckerman, personal communication, June 1985).

In the private sector, neurobehavioral testing as a method for evaluating health effects of the workplace was introduced in the early 1970s. Neurobehavioral testing studies have since been used to establish standards designed to reduce health impairment following exposure to neurotoxins (Johnson & Anger, 1983). Baker and Letz (1984) report that neurobehavioral tests may be used for a variety of purposes with working populations. Baker, Letz, Fidler, Shalat, Plantamura, and Lyndon (1985) and Baker, Letz, and Fidler (1985) have developed a microcomputer testing system for use in epidemiologic field studies of human populations in the workplace or general environment.

Problems

The military and private sector have been quick to identify the advantages associated with microcomputer performance testing as an applied research tool. Unfortunately, the attraction of the approach has led to the employment of tests and systems without adequate prior assessment and evaluation. Kiziltan (1985) has noted that performance test batteries are often assembled largely for practical reasons, on short notice, by persons whose major interest is not performance testing. Kennedy, Dunlap, Wilkes, and Lane (1985a) have noted that establishing the reliability and validity of newly developed microcomputer tests has lagged far behind both the use and marketing of such tests and that established principles for constructing and validating tests have been virtually ignored by developers. Farrell (1983) has more recently admonished the developers of microcomputer test batteries that the establishment of test metric characteristics is a necessary requisite prior to use. In addition, Farrell has observed that the apparent evaluation of microcomputer tasks is infrequently seen in the literature. The importance of Farrell's observation has been underscored by Smith, Krause, Kennedy, Bittner, and Harbeson (1983) who have demonstrated that changing the method of testing (paper-and-pencil to microcomputer) can change the statistical attributes of the test. Feldman, Ricks, and Baker (1980) have stated that, "The principal difficulty in evaluating behavioral effects is the relative lack of available standardized neuropsychological tests which can be administered to exposed workers in a practicable period of time, and which can be scored or interpreted with reliability, accuracy, and reproducibility" (p. 224). Michael (1982) has noted that, though specified in the Toxic Control Act of 1976, no satisfactory behavioral battery is available for judging the safety of new chemicals. Similarly, Weiss (1983) has observed that there is "exclusion of behavior from food additive testing ... although one of the reasons for its exclusion is the lack of confidence in currently proposed behavioral tests" (p. 1185).

In another research domain, Guignard, Bittner, and Harbeson (1983) have decried the failure of previous batteries to separate the mechanical and mental effects of vibration due to problems of test instability. Indeed, some researchers (Kennedy, Bittner, and Harbeson, 1980) have called into serious question most previous environmental studies which have not addressed the question of stability over repeated-measures. They caution that unstable measures "cannot be used reliably to measure environmental change, or any other effects" (p. 3).

Research in exotic work environments demands that research tools receive intensely critical evaluation during development. The need for sensitive and metrically sound performance measures assumes even greater importance in the research environment of space travel. Space flight is characterized by a small number of subjects carrying out tasks that cannot be studied at leisure. Such limited opportunity for the assessment of the factors influencing performance necessitates the use of the repeated-measures screening approaches employing each subject as his own control. Repeated-measures designs are more efficient and economical than alternative approaches (Winer, 1971) and are ideally suited to experiments with small numbers of subjects. However, the compound symmetry requirement of the variance-covariance matrix for simple repeated-measures analysis of variance demands that intertrial correlations be unchanging (differentially stable) and that variances be homogeneous across baseline repetitions (Winer, 1971; Jones, Kennedy, & Bittner, 1981; Bittner, 1979; Lord & Novick, 1968). As noted by some (Bittner & Carter, 1981; Kennedy, Bittner, Harbeson, & Jones, 1981; Jones, Kennedy, & Bittner, 1981) close attention has not typically been paid to the statistical requirements of repeated-measures testing.

As amply identified in the literature, accurate assessment of the effects of environmental agents on performance can not be made until basic measurement properties have been established. Even so, lack of attention to test metric properties prior to research remains the single most important barrier to adequate performance assessment. Overall, the lack of a standardized, stable, and sensitive performance measures has significantly delayed progress in human performance assessment and undoubtedly confounded the understanding of environmental effects in general.

Solutions

Within the last decade significant advances have been achieved in performance testing. These advances form the basis for the development of sound human performance measures. Of particular importance are the contributions of the PETER (Performance Evaluation Tests for Environmental Research) program, initiated in 1977 by the Naval Aerospace Medical Research Laboratory Detachment, New Orleans, Louisiana (Kennedy & Bittner, 1977, 1978). The purpose of this program was to develop a repeated-measures test battery, effective in measuring human performance decrements over time, or in unusual work environments. To qualify as a candidate for PETER evaluation, a performance test was first determined to be appropriate for repeated-measures assessment (i.e., possess comparable alternative forms), and second, to measure mental work. Tests initially identified were then further reviewed relative to the following criteria: (1) sensitivity to disruptions in test performance due to an environmental stimulus (e.g., ship motion); (2) concurrence in the scientific literature that the test measured an

identifiable information processing or cognitive construct for which a theoretical basis was available; (3) ability to differentiate brain damaged individuals from normals on the basis of test results; (4) previous appearance in an established and/or factor analyzed battery; (5) inherent interest to the subject; (6) obvious face validity; and (7) availability, cost, and other practical considerations (Kennedy, Jones, & Harbeson, 1980). Almost no test met all criteria but most tests met several. Having qualified as a candidate for additional study a test was then subjected to the intense PETER evaluation procedure. Typically, a candidate test was administered to a group of subjects through a series of 15 trials over 15 successive days. These data were then subjected to rigorous analysis in order to study the metric characteristics of the test. Emphasis was directed at establishing the stability of the test and the total time (or number of trials) to stabilization. Reliabilities for stabilized tests were then determined and a procedure for standardizing and comparing tests was established. Only tests demonstrating "good" metric properties were endorsed for repeated-measures research. This engineering evaluation approach has come to be known as the "PETER paradigm" or "PETER approach," and is recognized as a critical necessity that must preface the use of a performance test in subsequent research. The critical nature of such evaluation is further underscored by the significant finding that 80% of the performance tests evaluated under the auspices of the PETER program did not meet minimum standards.

The issues and methodologies relevant to repeated-measures metric characteristics evaluation have been discussed in detail in previous works (Jones, 1969b, 1979, 1980; Bittner & Carter, 1981; Kennedy & Bittner, 1977; Kennedy, Bittner, & Harbeson, 1980; Harbeson, Bittner, Kennedy, Carter, & Krause, 1983; Bittner, Carter, Kennedy, Harbeson, & Krause, 1984). However, an abbreviated discussion of the PETER test selection criteria are presented below:

(a) Stability. Jones, Kennedy, and Bittner (1981) make the point that most subjects demonstrate improvement with practice for most performance tasks. Performance typically follows a pattern of negative acceleration (i.e., classic learning curve for acquisition) with most change occurring early in practice and less occurring late. In general, as practice continues, group means and individual subjects approach asymptote (i.e., remain constant or change in a linear manner over trials). An obvious consequence of such a pattern is that the obtained point measures for a subject may differ significantly over time. A second consequence of particular concern is the fact that different subjects may respond differently rather than uniformly to repeated exposures of the task. Therefore, the relative standings of subjects on the first measures may not resemble the relative standings on the final measure. Only after relative standings are clearly and consistently established between subjects (i.e., asymptotic performance with parallel curves for subjects) can the investigator place confidence in the adequacy of his measure. Such an instrument is said to have "stabilized" and results from a stable test may be readily interpreted, whereas results from unstable tests are ambiguous (Jones, 1979, 1980). Similarly, Jones suggests that 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.

Generally stated, a test is defined as stable when: (1) the group means for successive trials become constant (i.e., are level, asymptotic, or exhibit constant slope); (2) the between-subject variances for successive trials become constant (i.e., homogeneity of variance); (3) the correlation between a trial and subsequent trials becomes constant. This latter criterion of stability has been labeled "differential stability" (Jones, 1969a, 1972). If a task has not stabilized, the correlations among successive trials will very likely show "superdiagonal form" (Jones, 1969b). That is, the correlations are greatest between two immediately adjacent trials, with greater separation between trials resulting in progressively smaller correlations. Jones (1979) has summarized the superdiagonal form with the following algebraic inequalities:

$$r_{ij} > r_{jk}$$
and
$$r_{ik} < r_{jk}$$
$$(i < j < k)$$

Examination of an intertrial correlation matrix of an unstabilized task makes the pattern readily apparent. Correlations within rows decrease from left to right and correlations within columns increase from top to bottom. Therefore, the smallest intertrial correlation would be found in the upper right-hand corner of the matrix. When these correlations cease to change within a row and column and subsequent rows and columns of the matrix, differential stability has been achieved. Theoretically, correlations among stabilized trials are equal. Examples of applications in establishing test stability may be examined in Harbeson, Kennedy, and Bittner (1979), and Kennedy, Carter, and Bittner (1980). It is important to note that differential stability requires uniform intertrial correlations as well as unchanging means and standard deviations across trials.

(b) Stabilization Time. It may be necessary to evaluate highly transitory changes in performance when studying the effects of various treatments, drugs, or environmental stress. Data collected in such situations must clearly reflect effects on performance due to a specific factor, as opposed to confounded effects resulting from combined factors. Therefore, in addition to stability per se, "good" performance measures should reach stability "quickly" following short versus long periods of practice without sacrificing metric qualities. Clearly, rapidly stabilizing tasks are prime candidates for inclusion in a final battery. A task under consideration for environmental research must be represented in terms of the number of trials necessary to establish stability and/or the total amount of time necessary to establish stability. One task, Grammatical Reasoning (Baddeley, 1968), is representative of tasks that stabilize quickly. According to Carter, Kennedy, and Bittner (1981), Grammatical Reasoning can be expected to stabilize within five 60-second trials.

(c) Task Definition. Once differential stability has been achieved, the next requirement for a test is task definition. Task definition is the average reliability of the stabilized task (Jones, 1979, 1980). Higher average reliability improves power in repeated-measures studies when variances are constant. It is well known that 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.

(d) Reliability-Efficiency. Test reliability is known to be influenced by test length (Guilford, 1954). Tests with longer administration times and/or more items enjoy a reliability advantage over shorter test times. Therefore, test length must be equalized before meaningful comparisons can be made. A useful tool for making such relative judgments is the reliability-efficiency (also referenced as "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 or time 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. A nomogram is also available for easy calculation (Kennedy, Carter, & Bittner, 1980).

(e) Task Sensitivity. Task sensitivity may be conceptualized as a test's ability to discriminate differences between subjects on one testing occasion, or within subjects on repeated testing occasions. If tests are stable, insensitivity is proportional to the lack of reliability-efficiency. In a repeated-measures paradigm, each subject serves as his own control, and if between-subject differences are present, tests with retest reliabilities below $r = .25$ can be expected to be insensitive to change. Thus, while high task definition ($r > .707$) does not guarantee sensitivity, lack of it guarantees insensitivity.

(f) Task Ceiling. Tests may meet all of the previously stated criteria and yet be unsuitable candidates for inclusion in a performance battery. Group variability over trials should not decrease. If variability between individual scores decreases over repeated-measures, then tests are likely to possess ceilings. If all individual subjects asymptote at the same or near same levels of performance, then the test is said to have a ceiling or top (Jones, 1980). Ceilings are undesirable because they limit discrimination between subjects although discrimination would otherwise be possible; for example, overlearning could make performance quite resistant to the environmental treatment. When subjects perform equally well except for random error, between-trial correlations fall to zero. This collapse of nonerror variance has been described as "radical destabilization" by Jones (1979, 1980).

Microcomputer Testing

Attention solely to the adequacy of the performance measures may not satisfy all the testing demands encountered within exotic environments. The aerospace work situation requires objective, efficient, and convenient procedures for test material presentation, data collection, and data storage. Time factors are critical, necessitating rapid data analysis and immediate feedback of results. These concerns demand that innovative methods be explored. Microcomputer testing provides a vehicle that may relieve many of the problems common to exotic environment research. Table 1 presents a listing of the attributes associated with the microcomputer testing mode as opposed to paper-and-pencil. Even casual inspection suggests the overwhelming

superiority of the automated approach. Collectively, these advantages provide for more comprehensive assessment, enhanced reliabilities, and increased promise for new assessment paradigms and perspectives. It must be emphasized, however, that the benefits of microcomputer testing are only "potential" in nature. As in the case of individual performance measures, extensive test and evaluation of the system must first preface actual research application. Although time consuming, such efforts insure that potential benefits are fully realized and desired outcomes are achieved.

TABLE 1. ADVANTAGE OF MICROBASED TESTING
COMPARED TO PAPER-AND-PENCIL TESTING

(a)	Standardized testing conditions leading to higher test reliabilities.
(b)	Reduced variability between test procedures and administrators enhancing comparison of results between similar studies.
(c)	Accurate and objective response scoring, eliminating unintelligible responses, improper scoring, and subjective interpretation.
(d)	Complete automation of all testing, scoring, and data collection functions resulting in a reduction of problems associated with lost or misplaced data.
(e)	Utilization of a variety of response measures such as speed and latency.
(f)	Presentation of complex and innovative stimuli involving a variety of sensory modalities.
(g)	Capabilities for precise timing and control of stimulus materials.
(h)	Immediate scoring of responses with easy access to data for rapid analysis or feedback to the subject or administrator.
(i)	Automatic data storage with capabilities for handling quantities of diverse data over repeated trials, with large N's.
(j)	Self administration of interesting and challenging materials resulting in increased subject motivation and reduced boredom.
(k)	Increased convenience and efficiency in data collection reducing the need for highly skilled professionals or psychological technicians.
(l)	Portability of the system with the accompanying advantages of reduced size and weight.
(m)	Adaptive testing, where difficulty level changes with performance, can shorten testing time.

APT System

System Development. The major weakness within most early human performance research has been the inadequate evaluation of basic research tools prior to application. Both performance measures and test delivery systems have received criticism. Recently, through NASA sponsorship, we have attempted to combine critically evaluated performance tests with field assessed microprocessor delivery systems. The product of these efforts is the Automated Performance Test System (APT System) and it was specifically developed for use in human performance research and subjective status (Bittner, Smith, Kennedy, Staley, & Harbeson, 1984). System development was spurred by the general promise of microcomputers for human assessment, and the recent advent of the low-cost notebook-sized microprocessor (Kennedy, Bittner, Harbeson, & Jones, 1981). The APT System may be conceptualized as comprised of three subsystems: (a) hardware, (b) test programs, and (c) system control and is described in detail elsewhere (Bittner et al., 1984).

Recently, a preliminary field study of the APT System for compatibility with environmental testing was completed (Kennedy, Wilkes, Lane, & Homick, 1985). A microcomputer battery of six tests was administered in conjunction with a similar paper-and-pencil battery. The two batteries were found to be comparable resulting in strong endorsement for more extensive evaluation of the APT System. Overall, the APT System appears to be a potentially powerful tool for repeated-measures performance research in remote, unusual, or exotic environments.

System Applications and Prospects. Initially, the APT System was under development to provide a human assessment capability suitable for use in remote operational settings. Other researchers have recognized the benefits of using the well developed and versatile tool and more than a dozen laboratories and universities now employ the system. These studies are currently investigating a broad range of environmental effects on performance. Factors under examination include altitude, motion, sleep loss, workload/fatigue, pharmacological agents, and others. Appendix A provides summary information regarding the current status and tentative results of each study. Although most analyses are not complete, preliminary results are exceptionally encouraging. These preliminary findings, although conditional, consistently point to the sensitivity of the APT System. Furthermore, this sensitivity appears to be enjoyed across the broad range of environmental factors under examination. Considered collectively, these preliminary findings provide consistent evidence of the effectiveness of the System. The overall implication created is that the APT System can function as a sensitive and reliable indicator of human performance, with substantial prospects for future growth and development.

Purpose

The purpose of this research effort is to expand upon, refine, and continue with previous efforts to develop a fully portable automated battery of measures sensitive to change in human performance. The research plan for reaching these objectives entails examination of potential performance measures. These measures must be implemented on the microcomputer testing device and repeatedly administered to subjects. Selection of a particular test into the final battery will be based on demonstrated metric qualities, factor structure, and compatibility with microcomputer administration.

METHOD

Subjects

Twenty-eight Casper College students were recruited for participation. The subjects were solicited from introductory psychology classes on a voluntary basis in accordance with American Psychological Association principles for research with human subjects (American Psychological Association, 1983). Subject procurement and research procedures were reviewed by the Casper College Human Use Committee (Appendix B). The committee found the proposed study to be in compliance with established standards regarding the treatment, welfare, and dignity of research subjects. Subjects that completed the study were paid for their participation at an approximate rate of $4.00/hr. Seven of the original 28 volunteers attrited the study. Attrition for 3 of the subjects was related to personal decisions to withdraw from the academic setting. In the remaining 4 cases, the subjects were terminated from participation due to inability to comply with data collection criteria. Final analyses were based on the data from 21 subjects with 5 males and 16 females participating. The subjects ranged in age from 17 to 29, were in good physical and mental health, and varied from freshman to junior standing. Subject motivation was high with 32% of those solicited volunteering and 75% of those participating completing the study. Motivation for the research task appeared to remain high throughout the experimental sessions.

Materials

Bittner, Carter, Kennedy, Harbeson, and Krause (1984) provide a general menu of performance tests classified according to their efficacy for use in repeated-measures research. From the menu, and from other sources, 11 performance tests and one short form general measure of intelligence were selected for examination. Specific tests were selected on the basis of one or more of the following considerations: (1) demonstrated conformity to the criteria for "good" performance tests (see Table 2); (2) potential for improved metric qualities given revised methods of application; (3) indications representing well-differentiated factors associated with cognitive, perceptual, or motor skills; (4) present or potential compatibility with the microcomputer testing mode. The tests, complete with pre-existing individual summarized selection information, may be viewed in Table 3. Table 2 provides clarification and detailed descriptions of the selection information criteria presented in Table 3.

Five of the tests previously recommended as a "mini-battery" for environmental research (Bittner, Carter, Kennedy, Harbeson, & Krause, 1984) were included for additional examination. The first five tests identified in Table 3 comprise the mini-battery. These tests were recently examined with a limited PETER approach (Kennedy, Wilkes, Lane, & Homick, 1985), and judged to be excellent candidates for repeated-measures environmental research.

TABLE 2. DESCRIPTIONS OF TASK SELECTION CRITERIA

Selection Criteria	Descriptions
FACTOR	The factor(s) assessed by the measure as identified in the literature.
DOMAIN	Characterization of the domain(s) of assessment of the capability as cognitive, perceptual, or motor skills.
TESTING MODE	The task mode or modes of administration identified as paper-and-pencil, microbased, or both.
TIME TO STABLE Xs AND SD	The total amount of elapsed time (massed or distributed) required for task mean and standard deviation stabilization for paper-and-pencil and/or microbased testing mode.
TIME TO DIFFERENTIAL STABILITY	The total amount of elapsed time (massed or distributed) required for task intertrial correlation stabilization for paper-and-pencil and/or microbased testing mode.
TASK DEFINITION	The reliability (r) of the task following the occurrence of differential stabilization for paper-and-pencil and/or microbased testing mode.
RELIABILITY EFFICIENCIES	The reliability (r) of a stabilized task standardized to a 3-minute administration base for paper-and-pencil and/or microbased testing mode.
EVALUATION CATEGORY	A global judgment of the acceptability of a paper-and-pencil and/or microbased test for use in repeated-measures research. Tasks are judged as recommended, acceptable-but-redundant, marginal, or unacceptable.
EVALUATION REFERENCE	The relevant study of stability and the original source of the measure.

TABLE 3. STABILIZATION/RELIABILITY DATA, MICROBASED ADAPTABILITY, AND INFORMATION SOURCE FOR 12 TESTS SELECTED FOR STUDY

Task	Trial Mean Stabilizes	Trial SD Stabilizes	Trial r Stabilizes	Relability Efficiencies[a]	Microbased Adaptability[b]
1. Grammatical Reason.	2	2	3	.93	+++
2. Pattern Comparison	3	3	3	.93	+++
3. Code Substitution	4	4	4	.84	+++
4. Aiming[c]	9	5	12	.87	+
5. Spoke Control[c]	1	2	1	.95	+
6. Pattern Recognition	Data unavailable			.76	+++
7. Tapping (Averaged Order 3 Forms)	2	2	2	.94	+++
8. Short-Term Memory	5	5	5	.80	+++
9. Manikin	2	2	2	.79	+++
10. Dynamic Visual Acuity	New test - data unavailable				++
11. Choice Visual Reaction Time	8	1	8	.58	+++
12. Wonderlic	4	1	1	.70[d]	+

Evaluation References

1. Bittner et al., 1984
2. Bittner et al., 1984
3. Kennedy et al., 1985
4. Bittner et al., 1984
5. Bittner et al., 1984
6. Shannon, Carter, & Boudreau, 1981
7. Kennedy et al., 1985
8. Carter, Kennedy, Bittner, & Krause, 1980
9. Carter & Wolstad, in press
10. No references provided for 10 - New Test
11. Krause & Bittner, 1982
12. Mackaman, Bittner, Harbeson, Kennedy, & Stone, 1982

a Reliability efficiency: Reliability estimated for a 3-minute test computed using the Spearman-Brown formula (Bittner & Carter, 1981.

b Microbased adaptability: Rated adaptability of a task to the microbased testing mode. +++ = high, ++ = acceptable, + = low

c Stabilization data estimated from original data

d Task definition reported for Wonderlic

Where possible, the tests were administered in both the microcomputer and paper-and-pencil modes. Tests presently not adapted to the microcomputer testing mode were presented in paper-and-pencil form only. The paper-and-pencil task presentation order, practice times, individual trial times, and total times are presented in Table 4. The Wonderlic Personnel Test (Wonderlic, 1978), which was administered as a general measure of intelligence, and was not under consideration as a potential candidate for repeated-measures performance testing, does not appear in Table 4. Tests presented only in the microcomputer mode appear in Table 5 with presentation order, practice times, individual trial times and total times indicated. Each task listed in Tables 4 and 5 is described in summarized form below.

TABLE 4. PAPER-AND-PENCIL TEST BATTERY
ORDER, TASK, AND BATTERY TIME

TASK ORDER	Trials/ Battery	Practice Time	Trial Time	Total Task Time in a Battery Less Practice	Total Time on Task for 10 Replications of the Battery Less Practice
AIMING	2	15[a]	90	180	1800
SPOKE	2	15	30	60	600
PATTERN COMPARISON	1	15	75	75	750
GRAMMATICAL REASONING	1	15	90	90	900
CODE SUBSTITUTION	1	15	60	60	600
PATTERN RECOGNITION	2	15	75	150	1500
TOTALS		90		615	6150

[a] Times are reported in seconds.

TABLE 5. MICROCOMPUTER TEST BATTERY ORDER, TEST AND BATTERY TIME

TASK ORDER	Trials/ Battery	Practice Time	Trial Time	Total Task Time in a Battery Less Practice	Total Time (in secs) on Task for 10 Replications of the Battery Less Practice
PREFERRED HAND TAPPING	2	10a	10	20	200
PATTERN COMPARISON	1	15	75	75	750
TWO-HAND TAPPING	2	10	10	20	200
GRAMMATICAL REASONING	1	15	90	90	900
NON-PREFERRED HAND TAPPING	2	10	10	20	200
MANIKIN	1	10	60	60	600
SHORT-TERM MEMORY	1		60	60	600
CODE SUBSTITUTION	1	10	60	60	600
DYNAMIC VISUAL ACUITY	1		60	60	600
REACTION TIME	1		60	60	600
TOTALS		80		525	5250

(a) Time data report in seconds

Aim - The Aim task (Fleishman & Ellison, 1962) is accomplished by accurately marking a dot within a small oval-shaped target. The targets are 2mm in width and are repeated across the test page at the rate of 1/5mm. Subjects work continuously following the target trace. Performance is scored according to the number of targets correctly marked. Aim was presented in the paper-and-pencil mode only and is not directly adaptable to microcomputer testing. However, recent research (Kennedy, Wilkes, Lane, & Homick, 1985) indicated that microcomputer tapping correlates with Aim paper-and-pencil performance. Aim has been described as a perceptual motor task of manual dexterity with wrist-finger speed, and fine eye-hand coordination important to task performance (Carter, Kennedy, & Bittner, 1980). According to Bittner, Carter, Kennedy, Harbeson, and Krause (1984, p. 38), "Aim directly provides for assessment of environmental effects on fine eye-hand coordination and indirectly provides for the separating of such effects from other cognitive

measures." Previous studies with Aim (Bittner, Carter, Kennedy, Harbeson, & Krause, 1984; Kennedy, Wilkes, Lane, & Homick, 1985) have indicated that the task is highly recommended for use in repeated-measures research.

Spoke Control (C) Task - The Spoke Test (Bittner, Lundy, Kennedy, & Harbeson, 1982) is a modification of the Trail Making Test (Reitan, 1955). The subjects' task is to accurately make a mark within a circular target. The targets are 1cm in diameter, 9cm from a control point, and are evenly spaced on 32 imaginary radii eminating from the control point. Subjects accomplished the task by placing a mark within a target, returning to the control point, and proceeding to the following target. Performance is scored according to the number of targets correctly marked. Spoke was presented in the paper-and-pencil mode only and has not yet been adapted to microcomputer testing. However, recent research (Kennedy, Wilkes, Lane, & Homick, 1985) indicated that microcomputer tapping performance correlates with Spoke paper-and-pencil performance. Spoke is a psychomotor task with visual search as an important factor in performance (Bittner, Carter, Kennedy, Harbeson, & Krause, 1984, p. 38). Spoke "directly assesses arm movement speed and indirectly provides for distinction of gross environmental disruptions from disruptions in fine eye-hand coordination and cognition." Previous studies with Spoke, reviewed in Bittner, Carter, Kennedy, Harbeson, and Krause (1984), and Kennedy, Wilkes, Lane, and Homick (1985), have highly recommended the task for use in repeated-measures research.

Pattern Comparison. 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 is directly adaptable to microcomputer testing and is presented in both the microcomputer and paper-and-pencil testing modes. Pattern Comparison has been described as a spatial ability important to perceptual performance. According to Bittner, Carter, Kennedy, Harbeson, and Krause (1984, p. 38), Pattern Comparison "assesses an integrative spatial function neuropsychologically associated with the right hemisphere." A review of Pattern Comparison studies (Bittner, Carter, Kennedy, Harbeson, & Krause 1984) indicated that the task is acceptable for use in repeated-measures research. Recent field testing with a microcomputer adaptation of the task (Kennedy, Wilkes, Lane, & Homick, 1985) resulted in strong recommendations for inclusion of Pattern Comparison in repeated-measures microcomputer test batteries.

Grammatical Reasoning - 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 subjects' 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 directly adaptable to microcomputer testing and was presented in both the microcomputer and paper-and-pencil modes. 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, Carter, Kennedy, Harbeson, and Krause (1984, p. 38), Grammatical Reasoning "assesses an analytic cognitive neuropsychological function associated with the left hemisphere". Previous studies with Grammatical Reasoning identified in Bittner, Carter, Kennedy, Harbeson, and Krause (1984) have indicated that the task is acceptable for use in repeated-measures research. Recent field testing with a microcomputer version of the task (Kennedy, Wilkes, Lane, & Homick, 1985) resulted in strong recommendations for inclusion of Grammatical Reasoning in repeated-measures microcomputer test batteries.

Code Substitution - The Code Substitution Test (Ekstrom, French, Harmon, & Dermen, 1976) is derived by randomly assigning digits to nine letters. The subjects' task is to repeat the assigned digit code when presented with the test letters. Subjects are not permitted to inspect the letter digit codes prior to testing. Performance is scored according to the number correctly coded. Code Substitution is directly adaptable to microcomputer testing and was presented in both the microcomputer and paper-and-pencil modes. Code Substitution is described as cognitive and perceptual type task with visual search encoding and decoding, rote recall, and perceptual speed as important factors in performance. According to Bittner, Carter, Kennedy, Harbeson, and Krause (1984, p. 38), "Code Substitution is a mixed associative memory-perceptual speed task which provides for a traditional assessment of those components not otherwise covered by other measures." Previous studies of Code Substitution (Pepper, Kennedy, Bittner, & Wiker, 1980) have indicated that the task is acceptable for use in repeated-measures research. Recent field testing with a microcomputer version of the task (Kennedy, Wilkes, Lane, & Homick, 1985) resulted in strong recommendations for inclusion of Code Substitution in repeated-measures microcomputer test batteries.

Pattern Recognition - The Pattern Recognition Test (Fitts, Weinstein, Rappaport, & Leonard, 1956) is composed of a stimulus histogram pattern and a sample of nine similar histogram patterns. The subjects' task is to search the sample of nine histograms and identify the sample histogram that is equivalent to the stimulus. Histogram forms for both the stimulus and the samples are randomly generated. Performance is based on the number of stimulus patterns properly identified. Pattern Recognition was presented in the paper-and-pencil mode only; however, the task is directly adaptable to microcomputer testing. Pattern Recognition has been described as a perceptual task, with pattern recognition as an important factor in test performance (Bittner, Carter, Kennedy, Harbeson, & Krause, 1984). Previous studies with Pattern Recognition (Carter & Sbisa, 1982; Carter & Krause, 1983; Kennedy, Wilkes, Lane, & Homick, 1985) have indicated that the task is acceptable for use in repeated-measures research.

Tapping - The test is accomplished by alternatively pressing keys on the microprocessor keyboard. The task was administered in three different forms: (a) preferred-hand tapping; (b) two-hand tapping, and (c) non-preferred hand tapping. Performance is based on the number of alternate key presses made in the allotted time. Tapping was presented in the microcomputer mode only and has not been tested in other modes. In a recent study (Kennedy, Wilkes, Lane, & Homick, 1985), tapping was described as a psychomotor skill assessing factors common to both Aim and Spoke. Tapping was also highly recommended for inclusion in a repeated-measure microcomputer battery.

Short-Term Memory - 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 subjects' task is to determine if the probe digit was included in the positive set and respond with the appropriate key press. Performance is based on the number of probes correctly identified. The Short-Term Memory was only presented in the microcomputer mode. Short-Term Memory is described as a cognitive-type task which reflects short-term memory scanning rate (Bittner, Carter, Kennedy, Harbeson, & Krause, 1984). Previous research with the task (Carter, Kennedy, Bittner, & Krause, 1980) has indicated that Short-Term Memory is acceptable for use in repeated-measures research.

Manikin Test - The Manikin Test (Benson & Gedye, 1963) involves the presentation of a simulated human figure in either a full-front or full-back facing position. The figure is shown to have two easily differentiated hand-held patterns. One of the two patterns is the matched pair to a pattern appearing below the figure. The subjects' task is to determine which hand of the figure holds the matching pattern and respond by pressing the appropriate microprocessor key. Pattern type, hand associated with the matching pattern and front-to-back figure orientation are randomly determined for each trial. Performance is based on the number of correctly matched pairs. The Manikin Test was presented in the microcomputer mode only. The Manikin Test is a perceptual measure of spatial transformation of mental images and involves spatial ability (Carter & Woldstad, in press). Bittner, Carter, Kennedy, Harbeson, and Krause (1984) recommended the use of the Manikin Test when latency scores are reported, and Kennedy, Wilkes, Lane and Homick (1985) identified the Manikin Test for inclusion in microcomputer repeated-measures batteries.

Dynamic Visual Acuity Test - This test (Higgins & Stultz, 1950) entails the presentation of a moving stimulus object (a Landolt C) with four possible orientations of the ring break. The cardinal position of the ring break is randomly determined for each trial. However, speed of travel of the figure is adaptively contingent upon the subjects' past performance. Faster and more accurate responses on the part of the subject result in faster rates of stimulus travel. Poor performance generates slower rates of travel. The subjects' task is to determine the orientation of the ring break and respond to the orientation. Performance is measured in terms of the fastest asymptotic velocity. The Landolt C was presented in the microcomputer mode only.

Reaction Time - The Visual Reaction Time Test (Donders, 1868) involves the presentation of a visual stimulus and measurement of a response latency to the stimulus. The subjects' task is to respond as quickly as possible with a key press to a simple visual stimulus. The visual stimulus is prefaced by an auditory signal and no decision making (disjunctive) regarding the stimuli is necessary. Reaction time is measured from the onset of the visual stimulus to the key press and was presented only in the microcomputer mode. 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, Carter, Kennedy, Harbeson, & Krause, 1984).

Wonderlic - The Wonderlic Personnel Test (Wonderlic, 1978) is a brief measure (50 questions/12-minute administration time) of general mental ability or "g." The test assesses the primary or general factor among many factors comprising intellectual capacity. General "g" is conceptualized as a condition that overlaps specific abilities to promote learning, problem solving, and communication. The test has been successfully used in the selection and placement of personnel and to predict achievement in the academic setting. Question types cover a broad spectrum and range from analogies to clerical items. The questions are arranged in order of difficulty and scoring is accomplished by summing the total correct. Sixteen similar forms have been produced and judged metrically comparable. A variety of validity studies have been reported, with coefficients varying from r = .10 (education) to r = .67 (professional occupation). Reported test-retest reliabilities range from r = .82 to r = .94. The Wonderlic was presented in the paper-and-pencil mode only and in its present form is not suitable for microcomputer testing. Previous research with the Wonderlic (Mackaman, Bittner, Harbeson, Kennedy, & Stone, 1982) has indicated that the test is suitable for use in repeated-measures research.

Apparatus

Microcomputer testing was accomplished with the Essex Corporation APT System, 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. The wide variety of auxiliary components that augment the system may be viewed in Figure 1. 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 light weight (3.8 lbs), compact (110 W x 40 H x 130 D mm), and fully portable with rechargeable nickel cadmium batteries permitting up to four hours of continuous operation. Table 6 abstracts the technical feature of the system which are more fully described in NEC (1983) and Essex (1985).

Procedure

Prior to testing, subjects received a brief introduction to the purpose of the study and were advised regarding the general procedures associated with data collection. Subjects were advised to work quickly, accurately, and to the best of their abilities. Attempts to raise motivation and reduce test anxiety were made by pointing out that the test batteries were the focus of the study, as opposed to the subjects themselves. In our judgment, the subjects were motivated to perform, but not adversely affected by performance anxiety.

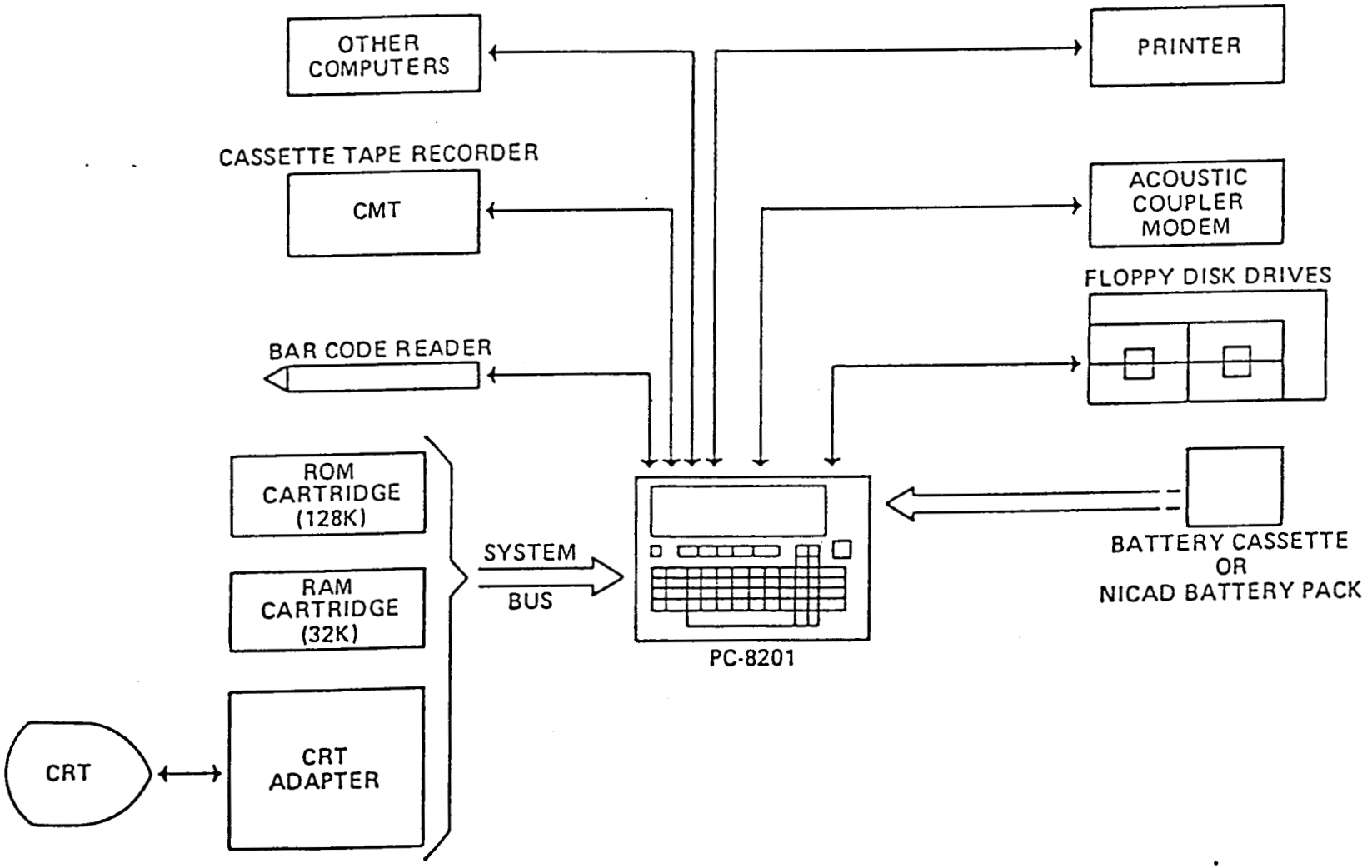

Figure 1. NEC 8201A auxiliary components.

TABLE 6. NEC PC8201A TECHNICAL SPECIFICATIONS

FEATURES	SPECIFICATIONS
SIZE	30 C M (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) - 128 K (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 NON-RECHARGEABLE BATTERIES, OR RECHARGEABLE NICKEL-CADMIUM PACK, OR AC ADAPTER 50/60 HZ @ 120 VAC, OR EXTERNAL BATTERY SYSTEMS (E.G., 8 AMP HR)

Subjects were examined over a six-week period in a modified PETER approach. On all occasions subjects were first administered the paper-and-pencil test battery, followed by the microcomputer test battery. Practice was provided preparatory to the first exposure of each paper-and-pencil test, with no further practice provided thereafter. Occurrence and amount of practice varied with each individual microcomputer test. Testing periods were arranged to occur on a weekly basis. During the first testing session subjects were tested in pairs to encourage individual questions, resolve problems, and provide explicit directions. In the first testing session, two back-to-back administrations of the paper-and-pencil and microcomputer test batteries were administered (i.e., AB AB). General instructions, statement of purpose, questions, answers, and test battery practice lengthened the initial test period by approximately 15 minutes for the average subject. Subsequent weekly testing was divided into a paper-and-pencil mode testing session and a microcomputer mode testing session. In the paper-and-pencil mode session, three consecutive back-to-back paper-and-pencil batteries were administered in a group setting. Group size varied from 3 to 5 subjects. Subjects were allowed to select and attend, at their convenience, one of three administration times on the designated test day. In general, group testing with the paper-and-pencil batteries could be accomplished within 40-50 minutes. In the microcomputer testing mode, subjects were required to self-administer three consecutive back-to-back

microcomputer batteries. Subjects selected test times convenient to their personal schedules with the requirement that microcomputer testing occur after the corresponding paper-and-pencil test battery, but prior to the administration of the pending paper-and-pencil test battery. Therefore, subjects enjoyed a 7-day option in which to fulfill the microcomputer testing obligation. In general, the average subject could complete the three microcomputer batteries in approximately 30-35 minutes. Following the initial back-to-back sessions, the paper-and-pencil and microcomputer testings were repeated over a three-week period, resulting in a total of 10 measures for each subject in both testing modes. During the fourth week of testing, procedures were slightly altered to include the administration of two forms (T11 and B) of the Wonderlic Personnel Test (Wonderlic, 1978). Subjects were required to complete two additional forms of the Wonderlic (EM and T21) the following week, resulting in a total of four measures/subject.

Analyses

The group means, standard deviations, and intertrial correlation matrices were calculated for each individual paper-and-pencil and microcomputer test over the first ten trials. Group means and standard deviations were examined for evidence of test stabilization, and intertrial correlations were assessed for evidence of differential stability. Rapid stabilization was predicted.

"Construct validity" for the battery was examined via correlation between the original and the computerized versions of the tests, and between the computerized versions and the Wonderlic. Such analyses enabled direct comparison and evaluation of the metric properties of individual tests and across test modes. Means, standard deviations, and intertrial correlation matrix for the Wonderlic Personnel Test were established, and cross-correlation between the microcomputer battery and the Wonderlic were calculated.

RESULTS

Analyses of Paper-and-Pencil Subtest Stabilities

Stability of Means. Inspection of Table 7 suggests that for the paper-and-pencil subtests stability of group means was achieved, or mean score improvement was significantly slowed, to imply stability well within the 10 trials. Group means appear to have stabilized for Code Substitution and Pattern Comparison by Trial 4, for Spoke and Aiming by Trial 5, for Grammatical Reasoning by Trial 6, and for Pattern Recognition by Trial 7.

Stability of Standard Deviations. The group subtest standard deviations are largely constant within each test. Examination of Table 7 suggests that standard deviation stability was achieved relatively quickly, with only slight changes occurring across the 10 trials. Grammatical Reasoning, Code Substitution, and Pattern Recognition all show slight increases, while Pattern Comparison remains relatively unchanged. Both Aiming and Spoke show decreases; however, the reduction in standard deviations across the 10 trials is too slight to imply problems with test ceilings.

TABLE 7. MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) OF SIX PAPER-AND-PENCIL TESTS ACROSS TEN TRIALS

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Aiming	179.3	206.3	203.8	213.9	221.8	201.8	221.7	229.8	220.0	227.7
	(34.7)	(33.2)	(26.6)	(25.0)	(23.8)	(24.4)	(25.8)	(24.3)	(26.0)	(27.1)
Spoke	32.1	34.6	34.2	36.8	37.4	34.5	38.9	38.4	38.3	39.4
	(5.3)	(5.8)	(4.7)	(4.7)	(5.2)	(9.1)	(4.7)	(5.3)	(4.3)	(4.7)
Pattern Comparison	40.8	48.3	56.8	58.4	57.3	55.6	60.8	53.0	60.9	61.4
	(9.2)	(9.4)	(8.9)	(9.6)	(8.3)	(10.8)	(9.7)	(13.8)	(7.8)	(10.8)
Grammatical Reasoning	21.0	24.6	25.4	25.4	25.5	28.0	27.3	27.5	27.7	29.5
	(6.3)	(5.8)	(5.9)	(8.4)	(8.4)	(9.0)	(8.3)	(7.9)	(8.6)	(7.7)
Code Substitution	37.9	39.1	39.7	40.1	38.1	38.8	39.7	42.3	40.4	39.4
	(4.6)	(4.6)	(5.8)	(4.7)	(4.6)	(5.6)	(6.3)	(6.1)	(6.2)	(5.2)
Pattern Recognition	18.3	18.7	21.6	22.4	21.8	22.1	23.6	23.0	26.1	25.0
	(2.5)	(3.2)	(3.5)	(3.6)	(3.4)	(6.7)	(4.2)	(6.6)	(4.2)	(3.9)

Differential Stability. Examination of the intertrial correlation matrices for the six subtests (Table 8) suggests that, in general, differential stability was established more rapidly than stability of group means. Aiming, Spoke, and Grammatical Reasoning are established as differentially stable by Trial 3. Questionable intercorrelations were obtained for Pattern Comparison, Code Substitution, and Pattern Recognition during Trials 6, 7, and 8. These correlations slightly complicate the determination of differential stability and require further consideration. A review of data collection procedures identified Trials 6, 7, and 8 as occurring during the same data collection session (session #3). Furthermore, inspection of the intercorrelation matrices for the same tests presented in the microcomputer mode (Table 13) do not indicate similar degraded correlations during the noted trials. Lastly, the tests in question all give indications of differential stability by Trial 4, with Code Substitution a possible exception. Therefore, it is reasonable to assume that the suspect correlations were a product of data collection discrepancies (probably timing of administration) and are not reflections of problems inherent within the tests.

TABLE 8. INTERTRIAL CORRELATIONS FOR SIX PAPER-AND-PENCIL SUBTESTS

Trial-to-Trial Intercorrelations of Paper-and-Pencil Tests (decimals omitted)

	Trials								
1	2	3	4	5	6	7	8	9	10
Aim									
100									
93	100								
85	92	100							
77	83	91	100						
73	82	91	94	100					
70	76	85	79	85	100				
66	75	84	82	86	95	100			
59	72	82	87	88	84	90	100		
58	69	82	78	84	88	95	90	100	
55	66	80	78	82	81	90	89	95	100
Spoke									
100									
95	100								
91	94	100							
80	87	90	100						
82	91	93	97	100					
24	25	34	23	28	100				
73	79	83	84	87	41	100			
76	80	80	86	84	30	90	100		
77	84	89	90	92	19	92	85	100	
64	75	77	84	84	21	89	87	92	100
Pattern Comparison									
100									
75	100								
59	79	100							
56	67	72	100						
43	70	64	80	100					
64	72	52	49	59	100				
54	80	67	77	85	73	100			
04	-06	27	39	24	00	07	100		
29	58	53	65	84	59	88	20	100	
35	47	53	67	78	47	73	43	85	100

TABLE 8. INTERTRIAL CORRELATIONS FOR SIX PAPER-AND-PENCIL SUBTESTS (CONT'D)

Trial-to-Trial Intercorrelations of Paper-and-Pencil Tests

Grammatical Reasoning

100									
65	100								
73	72	100							
69	80	83	100						
61	83	85	78	100					
72	79	82	73	86	100				
62	80	79	76	89	88	100			
68	84	70	80	78	85	78	100		
64	76	66	69	76	82	89	80	100	
60	78	78	80	86	85	84	86	84	100

Code Substitution

100									
86	100								
79	74	100							
87	80	80	100						
61	55	62	66	100					
62	61	35	40	28	100				
79	68	73	81	66	51	100			
79	78	91	76	64	42	81	100		
72	73	73	71	74	52	80	84	100	
62	54	56	65	74	13	60	56	50	100

Pattern Recognition

100									
51	100								
20	49	100							
42	65	85	100						
42	68	64	81	100					
11	02	18	11	19	100				
23	48	77	80	73	05	100			
30	13	18	24	25	00	21	100		
13	54	69	74	76	16	61	30	100	
25	52	72	79	79	08	73	32	88	100

TABLE 9. INDICATORS OF TEST STABILITY IDENTIFIED BY TRIAL AND ESTIMATED TIME TO ESTABLISH DIFFERENTIAL STABILITY FOR SIX PAPER-AND-PENCIL TESTS

TEST	Trial Mean Stabilizes	Trial SD Stabilizes	Trial Differential Stability Demonstrated	Time (in secs) to Establish Differential Stability
AIMING	5	3	3	900
SPOKE	5	3	3	180
PATTERN COMPARISON	4	2	4	300
GRAMMATICAL REASON.	6	4	3	270
CODE SUBSTITUTION	3	3	4	180 to 240
PATTERN RECOGNITION	7	3	4	600

In summary, the six paper-and-pencil tests may be viewed as demonstrating indications of rapid test stability. A comparison of the typical indicators of stability may be reviewed for the tests in Table 9.

Analyses of Microcomputer Subtest Stabilities

Stability of Means. Inspection of Table 10 indicates that continued improvement occurred within the microcomputer subtests means over the 10 trials. However, improvement appears to be sufficiently slowed in nine of the 10 tests to warrant stabilization. Group means appear to have stabilized for Preferred-Hand Tapping by Trial 2 and by Trial 3 for Pattern Comparison, Two-Hand Tapping, Grammatical Reasoning, Non-Preferred Hand Tapping, Short-Term Memory, and Code Substitution. Manikin and Reaction Time appear stabilized by Trial 5, however, the Dynamic Visual Acuity Test did not appear to stabilize with respective means showing improvement through the last trial. Comparisons of the group means for all similar paper-and-pencil and microcomputer tests (Table 7 vs. Table 10) suggest that, in general, the group means for microcomputer tests stabilized more quickly than the group means for corresponding paper-and-pencil tests.

TABLE 10. MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) OF TEN MICROCOMPUTER TESTS ACROSS TEN TRIALS

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Preferred-Hand Tapping									
34.6	36.7	36.6	36.7	36.8	37.7	37.8	37.8	38.6	38.9
(7.5)	(8.1)	(9.1)	(7.6)	(7.6)	(7.1)	(8.3)	(7.0)	(7.6)	(8.0)
Pattern Comparison									
55.2	53.9	56.7	55.9	56.1	57.1	58.6	57.6	57.4	58.7
(7.1)	(8.4)	(7.3)	(6.9)	(7.2)	(7.5)	(8.4)	(8.8)	(8.6)	(8.6)
Two-Hand Tapping									
40.6	39.3	41.5	40.7	41.7	42.7	41.2	42.6	43.1	43.2
(7.0)	(9.1)	(8.0)	(8.2)	(7.8)	(7.8)	(7.8)	(7.8)	(7.4)	(7.7)
Grammatical Reasoning									
23.0	22.5	25.5	25.1	25.9	24.8	25.8	27.1	25.8	25.8
(5.8)	(6.9)	(8.5)	(7.3)	(8.8)	(7.7)	(8.6)	(7.4)	(6.8)	(7.2)
Non-Preferred Hand Tapping									
31.0	32.1	33.8	33.4	32.9	34.7	33.9	34.5	34.9	35.5
(7.8)	(8.4)	(8.1)	(7.7)	(8.3)	(6.4)	(7.8)	(6.7)	(5.8)	(7.0)
Manikin									
29.0	34.3	36.9	35.9	38.0	38.4	39.8	41.6	40.5	41.6
(9.2)	(8.5)	(8.1)	(7.5)	(8.2)	(9.1)	(9.0)	(9.4)	(9.6)	(9.7)
Short-Term Memory									
30.5	31.7	32.6	31.8	32.9	32.0	32.4	32.9	33.9	33.2
(4.0)	(3.3)	(4.0)	(4.0)	(3.9)	(4.4)	(2.9)	(4.4)	(3.4)	(4.3)
Code Substitution									
25.0	25.7	29.7	28.2	29.3	28.8	31.5	31.7	30.8	31.8
(5.3)	(5.1)	(4.4)	(5.9)	(5.2)	(4.8)	(5.9)	(6.3)	(5.3)	(5.3)
Dynamic Visual Acuity									
134.2	162.1	154.7	129.8	153.5	176.4	180.3	181.7	186.2	198.9
(64.4)	(86.2)	(74.1)	(52.7)	(60.4)	(96.3)	(96.8)	(95.9)	(104.8)	(108.4)
Reaction Time									
450.3	336.3	336.2	383.6	333.6	344.0	325.3	342.5	335.4	321.6
(307.9)	(78.6)	(104.7)	(203.2)	(121.8)	(139.7)	(87.4)	(105.6)	(80.8)	(81.6)

Stability of Standard Deviations. Standard deviations for eight of the microcomputer subtests demonstrated only slight or no change across the 10 trials (Table 10). Preferred-Hand Tapping, Manikin, Short-Term Memory, and Code Substitution show virtually no change. Pattern Comparison, Two-Hand Tapping, and Grammatical Reasoning show slight increases. Non-Preferred Hand Tapping showed a slight insignificant decrease. Standard deviations for

Reaction Time decrease until about trial 7 which may indicate less skew on later trials, but we do not feel this represents a floor effect. The standard deviations for the Dynamic Visual Acuity Test show increases through the 10 trials. The lack of stability in standard deviations and, as previously noted, in the means for the Dynamic Visual Acuity indicate lack of test stability. This is believed to be due primarily to an artifact of the algorithm used for adaptive variation of difficulty, and not to any inherent instability of the phenomenon itself.

Differential Stability. Inspection of the intertrial correlation matrices (Table 11) for the 10 subtests suggests that differential stability is established relatively quickly for most of the tests. Preferred and Non-Preferred Hand Tapping give evidence of differential stability as early as Trial 2; Two-Hand Tapping and Code Substitution stabilized by Trial 3; Manikin and Short-Term Memory stabilized by Trial 4; and both Pattern Comparison and Grammatical Reasoning stabilized by Trial 5. Dynamic Visual Acuity Test again lacked indications of stability. Reaction time appears to have stabilized by Trial 6.

TABLE 11. INTERTRIAL CORRELATIONS FOR TEN MICROCOMPUTER SUBTESTS
(decimals omitted)

Trials									
1	2	3	4	5	6	7	8	9	10
Preferred-Hand Tapping									
100									
87	100								
79	93	100							
87	95	94	100						
84	94	92	94	100					
87	87	87	88	89	100				
84	84	88	84	89	86	100			
82	84	84	86	91	87	90	100		
81	86	81	81	88	87	89	94	100	
88	89	88	89	90	92	88	94	95	100
Pattern Comparison									
100									
90	100								
84	79	100							
89	77	77	100						
88	82	84	87	100					
82	80	70	76	85	100				
80	63	58	78	75	76	100			
82	73	74	70	82	82	72	100		
74	70	67	79	83	84	77	75	100	
84	63	66	78	82	77	78	78	70	100
Two-Hand Tapping									
100									
59	100								
84	70	100							
85	62	89	100						
85	60	91	95	100					
79	51	84	86	91	100				
83	52	89	88	93	96	100			
79	76	86	79	81	88	89	100		
83	54	83	78	82	86	93	87	100	
87	47	86	89	91	88	94	81	93	100
Grammatical Reasoning									
100									
91	100								
69	61	100							
78	67	85	100						
78	62	75	86	100					
80	67	72	78	89	100				
76	72	64	67	82	87	100			
79	67	49	56	82	80	83	100		
78	62	74	72	88	79	75	87	100	
60	78	78	80	86	85	84	86	84	100

TABLE 11. INTERTRIAL CORRELATIONS FOR TEN MICROCOMPUTER SUBTESTS (CONT'D)

Non-Preferred Hand Tapping

100									
95	100								
94	93	100							
93	92	94	100						
95	93	93	96	100					
91	88	85	89	91	100				
95	97	95	95	97	80	100			
90	91	83	88	90	95	91	100		
87	87	83	89	89	95	88	96	100	
93	94	90	92	93	93	93	93	93	100

Manikin

100									
81	100								
76	80	100							
64	78	84	100						
77	83	90	92	100					
60	76	84	85	92	100				
61	75	83	89	91	86	100			
65	84	84	86	91	90	94	100		
54	77	72	80	83	81	85	88	100	
60	81	86	90	92	88	95	95	93	100

Short-Term Memory

100									
76	100								
81	63	100							
75	68	76	100						
55	51	81	76	100					
69	53	76	69	74	100				
74	55	64	86	68	65	100			
60	55	74	51	56	40	60	100		
66	63	69	73	68	49	67	65	100	
63	69	69	80	75	68	77	70	70	100

Code Substitution

100									
52	100								
58	63	100							
55	66	70	100						
47	54	70	55	100					
75	58	90	79	68	100				
52	56	78	75	85	78	100			
51	54	75	69	87	77	92	100		
37	25	66	63	66	63	79	77	100	
50	57	73	67	61	72	74	69	61	100

TABLE 11. INTERTRIAL CORRELATIONS FOR TEN MICROCOMPUTER SUBTESTS (CONT'D)

Dynamic Visual Acuity

100									
39	100								
27	19	100							
47	37	32	100						
51	51	48	63	100					
07	24	72	47	47	100				
50	48	60	54	41	69	100			
-11	03	67	18	20	69	43	100		
04	14	41	25	15	43	40	23	100	
28	42	06	31	46	00	33	-21	-03	100

Reaction Time

100									
24	100								
55	81	100							
-07	43	22	100						
08	59	63	15	100					
19	67	68	15	56	100				
28	64	55	63	18	63	100			
22	48	33	09	24	69	57	100		
50	59	57	23	19	58	79	78	100	
41	66	70	17	25	60	65	43	62	100

In summary, 9 of the 10 microcomputer tests may be viewed as demonstrating indications of fairly rapid test stability. Preferred-Hand Tapping, Pattern Comparison, Two-Hand Tapping, Grammatical Reasoning, Non-Preferred Hand Tapping, Manikin, Short-Term Memory, Code Substitution, and Reaction Time are recommended for inclusion in future microcomputer test batteries. The Dynamic Visual Acuity Test is not recommended for inclusion in future microcomputer test batteries without significant formatting changes and further research. A comparison of the typical indicators of test stability for the 10 microcomputer tests may be viewed in Table 12.

Comparison of Paper-and-Pencil and Microcomputer Subtests. A review of the paper-and-pencil and microcomputer data for similar subtests suggests that the two modes of testing are generally comparable. Comparisons of the Trial 9 reliabilities for paper-and-pencil and corresponding microcomputer tests (Table 8 and Table 11) indicate the following: (a) Pattern Comparison and Grammatical Reasoning reliabilities for paper-and-pencil testing are higher than the corresponding microcomputer reliabilities; (b) The Microcomputer-based Code Substitution reliability is higher than corresponding paper-and-pencil reliability; and (c) Aiming and Spoke, the motor ability paper-and-pencil tests, demonstrate reliabilities comparable to the microcomputer tapping tests. Comparison of the trials at which differential stability is first established (Table 9 and Table 12) also supports the notion of comparability between the testing modes. Pattern Comparison, Grammatical Reasoning, and Code Substitution are differentially stabilized for both testing modes between Trials 3 and 5. Differential stability was established for the microcomputer tapping tests by Trials 2 to 3, and by Trials 2 to 5 for the paper-and-pencil equivalents, Spoke and Aiming.

Validation of Microcomputer Subtests

Construct validation of the microcomputer battery was accomplished by correlating performance on the microcomputer subtests with performance on similar paper-and-pencil subtests. This type of construct validity is known as "convergent validity," and is said to occur "when a test or other measure of a proposed trait correlates strongly with instruments of other kinds designed to measure the same trait or that are thought to measure it" (Guilford & Fruchter, 1978, p. 437). To examine the cross-correlations, performances over the last three trials of each test were averaged. The resultant scores represent differentially stable performance indices for both modes of testing, with the Dynamic Visual Acuity Test and Reaction Time the exceptions. Table 13 presents the cross-correlations between the paper-and-pencil and microcomputer subtest batteries. Of particular interest are the correlations between similar tests. Correlations between the two modes for Pattern Comparison, Grammatical Reasoning, and Code Substitution were respectively 0.66, 0.93, and 0.76. These high correlations may be interpreted as evidence in support of the convergent validity of the microcomputer subtests. Correlations between the subtests for motor abilities were disappointingly low, with a peak correlation of 0.48 between Aiming and Two-Hand Tapping. Aiming and Spoke did, however, correlate moderately with Reaction Time (-0.59 and -0.50, respectively). Other interesting correlations were also surfaced. For example, paper-and-pencil Code Substitution correlated highly with microcomputer Pattern Comparison (0.73) and Manikin (0.61), while Aiming correlated highly with Pattern Comparison (0.74). Pattern Comparison correlated moderately with Reaction Time (-0.54). Caution is advised in interpretation because of the small sample.

TABLE 12. INDICATORS OF TEST STABILITY IDENTIFIED BY TRIAL AND ESTIMATED TIME TO ESTABLISH DIFFERENTIAL STABILITY FOR TEN MICROCOMPUTER TESTS

TEST	Trial X Stabilizes	Trial SD Stabilizes	Trial Differential Stability Demonstrated	Time (in secs) to Establish Differential Stability
PREF. HAND TAP	2	2	2	40
PATTERN COMP.	3	2	5	375
TWO-HAND TAPPING	3	2	3	60
GRAMMAT. REASON.	3	3	5	450
NON-PREFERRED HAND TAPPING	3	2	2	40
MANIKIN	5	2	4	240
SHORT-TERM MEM.	3	3	4	240
CODE SUBSTITUTE.	3	2	3	180
DYNAMIC VISUAL ACUITY	DOES NOT STABILIZE	DOES NOT STABILIZE	DOES NOT STABILIZE	DOES NOT STABILIZE
REACTION TIME	5	7	6	300

TABLE 13. CROSS-CORRELATIONS OF PAPER-AND-PENCIL SUBTESTS WITH MICROCOMPUTER SUBTESTS

	Microcomputer Tests									
P&P	PHTAP	Pattrn Comp.	THTAP	Reason	NPHTp	Mankn	Short-Term	CodSub	DyVis Acuity	RxnTm
Aiming	22	74	48	25	22	50	51	48	45	-59
Spoke	14	57	32	26	23	33	32	36	28	-50
Patt. Comp.	21	66	37	38	33	38	31	47	38	-54
Reason	10	46	13	93	28	35	40	51	16	03
Code Subtitut.	27	73	11	43	06	61	48	76	-01	-18
Pattern Recog.	40	52	30	55	42	51	45	44	41	-38

Analyses of Wonderlic Test Data

Means, Standard Deviations, and Intertrial Correlations. The means and standard deviations for four administrations of the Wonderlic Personnel Test are presented in Table 14. Comparison of the means demonstrate a consistent increase in scoring across the first three administrations and a return to the initial level with the last administration. The corresponding standard deviations are unremarkable and do not imply that a test ceiling is approached within the four administrations.

TABLE 14. MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) FOR FOUR ADMINISTRATIONS OF THE WONDERLIC PERSONNEL TEST

T1	T2	T3	T4
26.2	27.7	29.	26.5
(3.9)	(5.4)	(4.6)	(4.5)

Table 15 reflects the intertrial correlations for the four administrations of the test. Correlations of Test #4 with the previous tests are low and erratic. This pattern suggests that discrepancies occurred during the fourth Wonderlic testing period (the day prior to spring vacation), and supports the conclusion that the data from the fourth testing period should be disregarded. This conclusion is further strengthened by the unusual pattern observed in the mean scores of Table 14. Intercorrelations for the first three administrations of the test imply that differential stability has not been achieved by Trial 3. Also, it should be noted that the intertrial correlations for the first three trials are somewhat lower than those reported in previous research (Mackaman, Bittner, Harbeson, Kennedy, & Stone, 1982).

TABLE 15. INTERTRIAL CORRELATIONS FOR FOUR ADMINISTRATIONS OF THE WONDERLIC PERSONNEL TEST

100			
50	100		
75	62	100	
00	38	14	100

Wonderlic and Microcomputer Subtests Cross-Correlations. The intercorrelations between the Wonderlic Personnel Test and the microcomputer battery subtests are presented in Table 16. For analysis purposes, the microcomputer subtest performances were averaged over the last three trials, and the Wonderlic was averaged across all four administrations. The resultant scores for the microcomputer subtests represent differentially stable performance indices, excepting Dynamic Visual Acuity Test and Reaction Time.

Conversely, the score representing average performance on the Wonderlic does not reflect differential stability and was further negatively influenced by the administration discrepancies previously discussed. Therefore, the correlations presented in Table 16 may not accurately reflect relationships between the Wonderlic and the microcomputer subtests. The only noteworthy correlation presented in Table 16 (0.47) suggests a relationship between the Wonderlic and Grammatical Reasoning. The verbal nature of these two tests may provide the basis for the finding; however, the relationship is most likely underrepresented. Due to these preliminary findings, and because of the complications discussed earlier, it is recommended that subsequent research reexamine the relationship between the Wonderlic and the microcomputer subtests, with special interest in Grammatical Reasoning.

TABLE 16. WONDERLIC AND MICROCOMPUTER SUBTESTS CROSS-CORRELATIONS

Cross-Correlations Between the NEC Tests and the Wonderlic

PHTAP	PatCm	THTAP	Reasn	NPHTp	Mankn	Short	Codsb	Dyn.Vis.Acuity	RxnTm
14	35	24	47	25	23	22	30	35	-30

DISCUSSION

Completed Analyses

Past efforts to demonstrate the effects of exotic work environments on performance have often proved inadequate. In general, lack of success has been directly related to insufficient attention in establishing the basic psychometric characteristics of performance measures prior to employment in research. Only through the systematic development and evaluation of performance measures and automated delivery modes can past inadequacies be avoided. Such a systematic approach is tedious and costly and has been avoided, perhaps for these reasons. The initial step in such a process entails the identification of performance measures that are psychometrically sound. To accomplish this objective subjects were repeatedly measured with paper-and-pencil and microcomputer performance batteries. Subjects were also administered several short-form measures of general intelligence. The sub-subtests in each battery were scored for the number of items correct, and assessed for stability of means, stability of standard deviations, differential stability, and time to establish differential stability. Nine of the 10 microcomputer tests gave evidence of rapid stabilization (median trial = 3.5) with relatively high reliabilities. These data indicate that a battery of microcomputer tests can be formed with stable means, standard deviations, and intertrial correlations. Furthermore, all eight tests can be expected to stabilize with minimal amounts of practice. The specific microcomputer tests represented by sound psychometric qualities include: Preferred-Hand Tapping, Pattern Comparison, Two-Hand Tapping, Grammatical Reasoning, Non-Preferred Hand Tapping, Manikin, Short-Term Memory, Code Substitution and Reaction Time. These subtests are highly recommended for inclusion in microcomputer

performance batteries employed in repeated-measures research. Dynamic Visual Acuity was found not to be psychometrically stable. This test cannot be recommended for inclusion in a repeated-measures battery and should either be discarded or revised. However, factors tapped by this test can only be sampled through microcomputer testing. Loss of such a unique aspect strongly recommends that revision and further research be considered prior to abandoning this subtest.

A second objective of the reported research entailed establishing a form of construct validity for the microcomputer performance tests. To establish validity, similar paper-and-pencil and microcomputer performance measures were intercorrelated. The validity coefficients obtained for Pattern Comparison, Grammatical Reasoning, and Code Substitution provide strong evidence that the constructs measured by the paper-and-pencil tests were unaffected by adaptation to the microcomputer mode. Correlations between the paper-and-pencil tests of motor ability (Aiming and Spoke), and the microcomputer tests of motor ability (Tapping) proved disappointing. However, subsequent analyses may surface factors unique to the tapping task not previously identified. It may be concluded from these analyses that construct validity (convergent type) was established for three of the microcomputer performance tests. Factor analysis with the subtests in question may result in encouraging evidence. Guilford and Fruchter (1978) strongly recommend establishing the "factorial validity" of a measure as the "best" solution in addressing construct validity. Factorial validity is established by identifying the loading of a test on the factor that it represents. Such analyses are planned and further discussion of factor analysis is reviewed in the Discussion section of this paper under "Proposed Analyses."

A third objective of the research effort was to examine the Wonderlic Personnel Test as a correlate with the microcomputer performance measures. Four forms of the Wonderlic were administered and the trial means, standard deviations, and intertrial correlations were examined for evidence of test stability. Analysis of the Wonderlic data did not indicate strong support for test stability. Intercorrelations of averaged Wonderlic scores with microcomputer subtests proved especially unremarkable; however, the Wonderlic did correlate relatively highly with Grammatical Reasoning. Data collection discrepancies associated with the fourth administration of the Wonderlic should be considered in evaluating the test's stability and subtest intercorrelations. It is recommended that further research with the Wonderlic be undertaken prior to final decisions regarding test disposition. Specifically, the measure should be examined with larger N's over four to five replications for indications of test stability and improved task definition. If evidence of sound psychometric characteristics is obtained, then further subtest intercorrelations are recommended. In particular, an examination of the relationship between Grammatical Reasoning and the Wonderlic may prove beneficial. A related study (Kennedy, Dunlap, Wilkes, & Lane, 1985b) has shown strong relationships between some of these same tests and individually administered tests of intelligence.

Proposed Analyses

A number of important analyses remain to be carried out in order to establish a complete picture of the subtest psychometric characteristics. The task definition (reliability of the test following the establishment of

differential stability), and reliability efficiency (reliability of a test standardized to a 3-minute base) for each subtest must be determined employing scores based on the number of correct responses. Factor structure for each subtest and for the total battery must also be established employing these data. Because factor analysis requires a larger sample size than provided in the present study more data must be available for this purpose. Microcomputer subtests should be factor analyzed separately for each trial with corresponding analyses for the paper-and-pencil subtests, but all tests for all trials should be factor analyzed in a single analysis.

The versatility of the microcomputer testing approach provides for data collection typically ignored in traditional testing. Latency speed and throughput are interesting and important measures yet requiring analysis. These measures must also be systematically examined for stability, task definition, reliability efficiency, and factor structure. Such analyses shall be included in following reports.

Recommended Research

Findings to date suggest that at least two general areas of future effort may prove fruitful. The first and most obvious is the development, implementation, and assessment of potential subtests for inclusion in the APT System test batteries. Initial selection of candidate tests should be based on one or more of the following criteria: (a) Certain tests currently employed in the APT System battery should be refined and reexamined for improved psychometric properties. Grammatical Reasoning, which has been demonstrated to be an excellent subtest in previous form, could be further improved with attention to standard item length (Dunlap, 1986, unpublished observation). Landolt C may establish adequate psychometric characteristics with changes to presentation format and instructions. (b) Tests should be selected that have previously demonstrated good paper-and-pencil psychometric characteristics and are easily adapted to the microcomputer testing mode. Pattern Recognition is suggested as the most immediate candidate. (c) Tests should be selected that are likely to tap factors unidentified with previous measures and/or correlate highly with a standard measure of "g." Tests should be selected that are "enriched by" or "unique to" the microcomputer testing mode. Generally, "enriched" tests are appropriate for microcomputer testing because of their complexity or other features impossible to control with simple paper-and-pencil testing. Examples include tests of complex decision-making where the intensity of the stimulus presentation and the difficulty level of the appropriate response continuously change as a function of past performance; and tests of three-dimensional spatial ability (Cooper & Shepard, 1984) where computer-simulated objects may be rotated simultaneously about three axes. Such tests may prove rich in factors impossible to obtain through traditional paper-and-pencil testing. (d) The latency scores currently available on the microcomputer shall be analyzed as well as rights minus wrongs and other measures.

The second area of future research must include systematic efforts to demonstrate and document the sensitivity of the APT System batteries to factors known to compromise performance. These research efforts must first be performed under highly controlled laboratory conditions, followed by field testing in actual work settings and other "real world" environments. Specific environmental variables of interest include fatigue, work load, altitude,

motion sickness, and motion sickness drug therapy. Without such sensitivity documentation, interpretation of field results may be difficult or unclear. Even without such documentation researchers have been quick to employ APT System technology. Field studies as diverse as the identification of learning disabled children and cancer chemotherapy effects on performance are appearing in the literature. We have included as Appendix A a table which, while incomplete, constitutes a sort of status report of who is using the APTS microcomputer and their most recent findings. Such efforts may, to some degree, be regarded as preliminary attempts to establish the sensitivity of the APT System batteries. Furthermore, the rush to employ the APT System may be interpreted as an indication of the widespread need for such performance testing technology. Although the ultimate worth of the APT System will be reflected in the field identification of performance influencing factors, substantial sensitivity assessment and documentation remain important intervening tasks.

REFERENCES

American Psychological Association (1983). Ethical principals of psychologists (Revised). American Psychologist, 36, 633-638.

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

Baker, M. D., & Letz, R. (1984). Neurobehavioral testing in monitoring hazardous workplace exposures. Paper presented at Conference on Medical Screening in the Workplace, Cincinnati, OH.

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 and Teratology, 7, 369-377.

Banderet, L. E., Benson, K. P., MacDougall, D. M., Kennedy, R. S., & Smith, M. (1984). Development of cognitive tests for repeated performance assessment. Proceedings of the 26th MTA Conference, Munich, Germany.

Banderet, L. E., & Burse, R. L. (1984). Cognitive performance at 4500 meters simulated altitude. Paper presented at the Annual Meeting of the American Psychological Association. Toronto, Canada.

Bandaret, L. E., MacDougall, D. M., Roberts, D. E., Tappan, D., Jacey, M., & Gray, P. (1984). Effects of dehydration or cold exposure and restricted fluid intake upon cognitive performance. Published Conference Proceedings of the National Academy of Sciences (Committee on Military Nutrition Research) Workshop, "Workshop on Knowledge Needed for the Development of Predictive Models of Military Performance Decrements Resulting from Inadequate Nutrition." Washington, DC.

Benson, A. J., & Gedye, J. L. (1963). Logical processes in the resolution of orientation conflict (Report 259). Farnborough, UK: Royal Air Force Institute of Aviation Medicine.

Bittner, A. C., Jr. (1979). Statistical tests for differential stability. Proceedings of the 23rd Annual Meeting of the Human Factors Society (pp. 541-545). Santa Monica, CA: Human Factors Society (Research Report No. NBDL-81R010, 10-14). New Orleans, LA: Naval Biodynamics Laboratory (1981). (NTIS No. AD A111086)

Bittner, A. C., Jr., & Carter, R. C. (1981). Repeated measures of human performance: A bag of research tools (Research Report No. NBDL-81R011). New Orleans, LA: Naval Biodynamics Laboratory. (NTIS No. AD A113954)

Bittner, A. C., Jr., Carter, R. C., Kennedy, R. S., Harbeson, M. M., & Krause, M. (1984). Performance Evaluation Tests for Environmental Research (PETER): Evaluation of 112 measures (Research Report NBDL-84R006). New Orleans, LA: Naval Biodynamics Laboratory. (NTIS No. AD A152317) and Perceptual and Motor Skills, In press, 1986.

Bittner, A. C., Jr., Lundy, N. C., Kennedy, R. S., & Harbeson, M. M. (1982). Performance Evaluation Tests for Environment Research (PETER): Spoke Tasks. Perceptual and Motor Skills, 54, 1319-1331.

Bittner, A. C., Jr., Smith, M. G., Kennedy, R. S., Staley, C. F., & Harbeson, M. M. (1984). Automated Portable Test (APT) System: Overview and prospects. Proceedings of the 14th Annual Meeting of the Society for Computers in Psychology. San Antonio, TX.

Bittner, A.C., Jr., Smith, M.G., Kennedy, R.S., Staley, C.F. & Harbeson, M.M. (1985). Automated portable test system (APTS): Overview and prospects. Behavior Research Methods, Instruments and Computers, 17, 217-221.

Carroll, J. B. (1980). Individual difference relations in psychometric and experimental cognitive tasks (Contract No. N00014-77-C-0722). Personnel and Training Research Programs, Psychological Services Division, Office of Naval Research. (NTIS No AD A0860057)

Carter, R. C., Kennedy, R. S., & Bittner, A. C., Jr. (1980). Selection of Performance Tests for Environmental Research. Proceedings of the 24th Annual Meeting of the Human Factors Society (pp. 320-324). Santa Monica, CA: Human Factors Society. (NTIS No. AD A111296)

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

Carter, R. C., Kennedy, R. S., Bittner, A. C., Jr., & Krause, M. (1980). Item recognition as a Performance Evaluation Test for Environmental Research. Proceedings of the 24th Annual Meeting of the Human Factors Society, Santa Monica, CA: Human Factors Society, 340-344. Also, Naval Biodynamics Laboratory, New Orleans, 1981, 47-50 (Research Report No. NBDL-81R008). (NTIS No. AD A111296)

Carter, R. C., & Krause, M. (1983). Reliability of slope scores for individual slope score (Research Report No. NBDL 83R003). New Orleans, LA: Naval Biodynamics Laboratory. (NTIS No. AD A130252)

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, LA: Naval Biodynamics Laboratory. (NTIS No. AD A115021).

Carter, R. C., & Woldstad, J. C. Repeated measurements of spatial ability with the Manikin Test. Manuscript submitted for publication.

Christal, R. E. (1981). The need for laboratory research to improve the state-of-the-art in ability testing. Paper presented at the National Security Industrial Association, DoD Conference on Personnel and Training Factors in Systems Effectiveness, San Diego, CA.

Christal, R. E., Payne, D. L., Weissmuller, J., & Anderson, M. S. (1982). Learning abilities measurement program. Paper presented at the 23rd Annual Meeting of the Psychonomic Society, Minneapolis, MN.

Christensen, J. M., & Talbot, J. M. (1986). A review of the psychological aspects of space flight. Aviation, Space and Environmental Medicine, 57(3), 201-298.

Cooper, L. A., & Sheppard, R. N. (1984). Turning something over in the mind. Scientific American, 251(6), 106-114.

Donders, F. C. (1868). Die schnelligkeit psychischer processe. Archiv fur Anatomie und Physiologie und Vissenchaftliche Medizin, 657-681.

Ekstrom, R. B., French, J. W., Harmon, H. H., & Dermen, D. (1976). Manual for kit of factor referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Essex (1985). Ongoing research utilizing Essex' Automated Performance Test Systems: Portable microcomputer performance batteries on a NEC PC8201A. APTS News, 1(1).

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

Farrell, A. D. (1983). When is a computerized assessment system ready for distribution? Computers in Psychiatry/Psychology, 5, 9-11.

Feldman, R. G., Ricks, N. L., & Baker, E. L. (1980). Neuropsychological effects of industrial toxins: A review. American Journal of Industrial Medicine, 1, 211-227.

Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, J. A. (1956). Stimulus correlates of visual pattern recognition: A probability approach. Journal of Experimental Psychology, 51, 19-24.

Fleishman, E. A., & Ellison, G. D. (1962) A factor analysis of fine manipulative tests. Journal of Applied Psychology, 46, 96-105.

Guignard, J. C., Bittner, A. C., Einbender, S. W., & Kennedy, R. S. (1980). Performance Evaluation Tests for Environmental Research (PETER) Landolt C. Reading Test. Proceedings of the 24th Human Factors Society, Los Angeles, CA, 13-17, 355-339.

Guignard, J. C., Bittner, A. C., Jr., & Harbeson, M. M. (1983). Current research at the Naval Biodynamics Laboratory on human whole-body motion and vibration. Proceedings, U.K. Informal Group on Human Response to Vibration. London, England. (Research Report No. NBDL-83-R008, July 1983). (NTIS No. AD A138367)

Guilford, J. P. (1954). Psychometric methods (2nd ed.). New York: McGraw-Hill, 400-402.

Guilford, J. P., & Fruchter, B. (1978). Fundamental statistics in psychology and education. New York, NY: McGraw-Hill.

Guillion, C. M., & Eckerman, D. A. (in press). Field testing for neurobehavioral toxicity: Methods and methodological issues. To appear in Z. Annau (Ed.), Behavioral Toxicology.

Hanninen, H. (1979). Psychological test methods: Sensitivity to long-term chemical exposure at work. Neurobehavioral Toxicology, 1, 157-161.

Harbeson, M. M., Bittner, A. C., Jr., Kennedy, R. S., Carter, R. C., & Krause, M. (1983). Performance Evaluation Tests for Environmental Research (PETER): Bibliography. Perceptual and Motor Skills, 57, 283-293.

Harbeson, M. M., Kennedy, R. S., & Bittner, A. C. (1979). A comparison of the Stroop test to other tasks for studies of environmental stress. 12th Annual Human Factors Association of Canada Meeting, Bracebridge, Ontario, Canada, 1-9.

Johnson, B. L., & Anger, W. K. (1983). Behavioral toxicology. In W. R. Rom (Ed.), Environmental and occupational medicine. Boston.

Johnson, J.H., Kennedy, R.S., Lilienthal, M.G., & Merkle, P.J. (1985). Microprocessor based field testing for human performance assessment. Paper presented at the 27th Annual Military Testing Association Conference, San Diego, CA.

Johnson, J.H., Kennedy, R.S., Smith, M.G., & Dutton, B. (1985). On the use of portable microprocessors as field data collection units. Paper presented at the Annual Scientific Meeting of the Aerospace Medical Association, San Antonio, TX.

Jones, M. B. (1969a). Differential processes in acquisition. In E. A. Bilodeau and I. McD. Bilodeau (Eds.), Principles of skills acquisition. New York: Academic Press.

Jones, M. B. (1969b). Knowledge of results and intertrial correlations in a simple motor task. Journal of Motor Behavior, 1, 331-340.

Jones, M. B. (1972). Individual differences. In R. N. Singer (Ed.), The psychomotor domain (107-132). Philadelphia, PA: Lea & Febiger.

Jones, M. B. (1979). Stabilization and task definition in a performance test battery (Final Report, Contract No. N0023-79-M-5089). New Orleans, LA: U.S. Naval Aerospace Medical Research Laboratory.

Jones, M. B. (1980). Stabilization and task definition in a performance test battery (Final Report on Contract No. N0023-79-M-5089, Monograph No. NBDL-M001). New Orleans, LA: Naval Biodynamics Laboratory. (NTIS No. AD A099987)

Jones, M. B. (1981). Convergence-divergence with extended practice: Three applications. Proceedings of the 24th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society. Pp. 359-362. Also, Naval Biodynamics Laboratory, New Orleans: September 1981, Pp. 6-9. (Research Report No. NBDL-81R010). (NTIS No. AD A111086)

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. (1984, April). Objective measures of human capabilities. Presented at the Workshop on Advances in NASA-Relevant Minimally Invasive Instrumentation, Pacific Grove, CA.

Kennedy, R. S., & Bittner, A. C., Jr. (1978). The stability of complex human performance for extended periods: Applications for studies of environmental stress. Aerospace Med. Assoc. Meeting, New Orleans, LA.

Kennedy, R. S., & Bittner, A. C., Jr. (1978). Progress in the analysis of Performance Evaluation Tests of Environmental Research (PETER). Proceedings of the 22nd Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, 29-35. (NTIS No. AD A060676) Also, Naval Biodynamics Laboratory, New Orleans, 1981, 14-21 (Research Report No. NBDL-82R004). (NTIS No. AD A111180)

Kennedy, R. S., & Bittner, A. C., Jr. (1981). The development of a Navy performance evaluation test for environmental research (PETER). In L. T. Pope & D. Meister (Eds.), Productivity enhancement: Personnel performance assessment in Navy systems (pp. 391-408). Naval Personnel Research & Development Center, San Diego, CA. (NTIS No. AD A056047)

Kennedy, R. S., Bittner, A. C., Jr., Carter, R. C., Krause, M., Harbeson, M. M., McCafferty, D. B., Pepper, R. L., & Wiker, S. F. (1981). Performance Evaluation Tests for Environmental Research (PETER): Collected papers (NBDL-80R008). New Orleans, LA: Naval Biodynamics Laboratory.

Kennedy, R. S., Bittner, A. C., Jr., & Harbeson, M. M. (1980). An engineering approach to the standardization of Performance Evaluation Tests for Environmental Research (PETER). Proceedings of the 11th Annual Conference of the Environmental Design and Research Association (EDRA), Charleston, SC. Also, Naval Biodynamics Laboratory, New Orleans, 1-7 (Research Report No. NBDL-82R004). (NTIS No. AD A111180)

Kennedy, R. S., Bittner, A. C., Jr., Harbeson, M. M., & Jones, M. B. (1981). Perspectives in Performance Evaluation Tests for Environmental Research (PETER): Collected papers (Research Report No. NBDL-80R004). New Orleans, LA: Naval Biodynamics Laboratory. (NTIS No. AD A111180)

Kennedy, R. S., Carter, R. C., & Bittner, A. C., Jr. (1980). A cataloque of performance evaluation tests for environmental research. Proceeding, Human Factors Society, Los Angeles, 334-348.

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

Kennedy, R.S., Dunlap, W.P., Wilkes, R.L., & Lane, N.E. (1985b). Factor analysis of a fifteen minute microcomputer-based performance test battery. Final report. National Aeronautics and Space Administration. Report No. 85-2. Orlando, FL: Essex Corporation.

Kennedy, R.S., Dunlap, W.P., Jones, M.B., & Wilkes, R.L. (1985). Portable human assessment battery: Stability, reliability, factor structure and correlation with intelligence tasks. Orlando, FL: Essex Corporation. National Sciences Foundation, Final Technical Report 85-3. In preparation.

Kennedy, R.S., Dunlap, W.P., Jones, M.B., Wilkes, R.L., & Bittner, A.C., Jr. (1985). Automated portable test system (APTS): A performance envelope assessment tool (Tech. Paper 851775). Society of Automotive Engineers Technical Paper Series, Longbeach, CA.

Kennedy, R. S., Jones, M. B., & Harbeson, M. M. (1980). Assessing productivity and well-being in Navy workplaces. Proceedings of the 13th Annual Meeting of the Human Factors Association of Canada. Rexdale, Ontario, Canada: Human Factors Association of Canada, 108-113. Also (Research Report No. NBDL-82R004), Naval Biodynamics Laboratory, New Orleans, 1981, 8-13. (NTIS No. AD A111180)

Kennedy, R.S., Wilkes, R.L., Lane, N.E., & Homick, J.L. (1985). Preliminary evaluation of a microbased repeated measures testing system. Final report. National Aeronautics and Space Administration. Report No. 85-1. Orlando, FL: Essex Corporation.

Kiziltan, M. (1985). Cognitive performance degradation on sonar operator and torpedo data control unit operator after one night of sleep deprivation. Unpublished Masters' thesis, Naval Post Graduate School, Monterey, CA.

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

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

Lazarus, R. S., & Cohen, J. B. (1977). Environmental stress. I. Altman, J. F. Wohlwill (Eds.). In Human behavior and the environment: Current theory and research. New York: Plenum.

Lazarus, R. S., & Launier, R. (1978). Stress-related transactions between person and environment. In L. A. Pervin & M. Lewis (Eds.) Perspectives in international psychology. New York: Plenum.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Mackaman, S. M., Bittner, A. C., Jr., Harbeson, M. M., Kennedy, R. S., & Stone, D. A. (1982). Performance Evaluation Tests for Environmental Research (PETER): Wonderlic Personnel inventory. Psychological Reports, 51, 635-644.

McCauley, M. E., & Kennedy, R. S. (1976). Recommended human exposure limits for very low-frequency vibration. Department of the Navy, Pacific Missile Test Center, Point Mugu, CA: Technical Publication TP-76-36.

McCauley, M. E., Royal, J. E., Shaw, J. E., & Schmitt, L. G. (1979). Effect of transdermally administered scopolamine in preventing motion sickness. Aviation, Space and Environmental Medicine, 50, 1108-1111.

Michael, J. M. (1982). The second revolution in health: Health promotion and its environmental base. American Psychologist, 37, 936-941.

Miller, J. W. (1959). Dynamic visual acuity in applied settings (Report No. 16). Pensacola, FL: Naval School of Aviation Medicine.

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

Nicogossian, A. E., & Parker, J. F. (1982). Space physiology and medicine. NASA, Scientific and Technical Information Branch.

O'Donnell, R. D. (1981). Development of a neurophysiological test battery for workload assessment in the U.S. Air Force. Proceedings of the International Conference on Cybernetics and Society (pp. 398-402), IEEE, Atlanta.

Payne, D. L. (1982). Establishment of an experimental testing and learning laboratory. Paper presented at the Fourth International Learning Technology Congress and Exposition of the Society for Applied Learning Technology. Orlando, FL.

Pepper, R. L., Kennedy, R. S., Bittner, A. C., Jr., & Wiker, S. F. (1980). Performance Evaluation Tests for Environmental Research (PETER): Code substitution test. Proceedings of the 7th Psychology in the DoD Symposium (USAFA-TR-80-12). Colorado Springs, CO: USAF Academy, 451-457. Also Naval Biodynamics Laboratory, New Orleans, LA, 1981. (Research Report No. NBDL-80R008) (NTIS No. AD A111296)

Poulton, E. C. (1978). A new look at the effects of noise upon performance. British Journal of Psychology, 69, 435-437.

Reid, G. B., Shingledecker, C. A., Nygren, T. E., & Eggemeier, T. S. (1981). Development of multidimensional subjective measures of workload. Proceedings of the International Conference on Cybernetics and Society, Atlanta, GA: IEEE, Systems, Man, and Cybernetics Society.

Reitan, R. M. (1955). Investigation of the validity of Halstead's measures of biological intelligence. AMA Archives of Neurology and Psychiatry, 73(6).

Shannon, R. H., Carter, R. C., & Boudreau, A. Y. (1981). A systematic approach to battery development and testing within unusual environments. Meeting on "Research Methods in Human Motion and Vibration Studies," New Orleans, LA.

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

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

Thorne, D., Genser, S., Sing, H., & Hegge, F. (1983). Plumbing human performance limits during 72 hours of high task load. The Human as a Limiting Element in Military Systems. Toronto, Canada: DRG Seminar Defense and Civil Institute of Environmental Medicine.

Weiss, B. (1983). Behavioral toxicology and environmental health science. American Psychologist, 11, 1174-1187.

Winer, B. J. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Wonderlic, E. F. (1978). Wonderlic Personnel Test Manual. Northfield, IL: Wonderlic.

Woodward, D. P., & Nelson, P. D. (1976). A user oriented review of the literature on the effects of sleep loss, work-rest schedules, and recovery on performance. Office of Naval Research. Naval Medical Research and Development Command.

APPENDIX A

Summary of Currently Active Nonpublished Studies Involving the Automated Performance Test System (APTS) in Performance Research

General Areas of Research	Principal Investigator	Method of investigation	Study Status	Variables to be Examined or Manipulated	Number of Subjects	Subject Organismic Variables or Selection Factors	Performance Score	APTS Subtests	Number of Trials	Number of Battery Applications	Sensitivity/Significance	Potential Correlates
Simulated Altitude	Schiflett	Experiment	Complete	Sea Level vs 18000' vs 25000'	N = 6	NA(a)	Number Correct	PHT THT NPHT GR PC	3 3 3 6 4	18000'= 6 25000'= 3 min	18000' / 25000' 5% Loss / Total Disruption 5% " / " " 5% " / " " 40% " / " " 20% " / " "	NA
Simulated Altitude	Schiflitt	Experiment	Report in Preparation	Sea Level vs 13000'	N = 12	NA	Number Correct	PHT THT NPHT GR PC	3 3 3 6 4	13000'= 6	13000' NE(b) " " "	Battery from two other computer systems
Simulated Altitude	Coussens	Experiment	Analysis	Sea Level vs 11000' vs 14000'	N = 30	Above Average IQ 1SD above	Number Correct and mean Time Between Responses	PHT Mean THT NPHT GR PC CS SB MNK VM LC RT	4 4 4 4 4 4 4 Slow Slow No Stab 4	4 11000'= 4 14000'= 4	No.Correct: 11000' NE, 14000' NE; Mean Time: 11000' NE, 14000' NE " " " " " " " " " " " SE(c) " UK(d) " " " NE	Anxiety Inventory; Environmental Symptoms Questionnaire; Various Physiological Measures; General Intelligence

General Areas of Research	Principal Investigator	Method of investigation	Study Status	Variables to be Examined or Manipulated	Number of Subjects	Subject Organismic Variables or Selection Factors	Performance Score	APTS Subtests	Number of Trials	Number of Battery Applications	Sensitivity/Significance	Potential Correlates
Simulated Altitude	Banderet	Experiment	Complete	Sea Level vs 15000' and Adaptation to 14000' vs No Adaptation	N = 23	US Army Special Forces Personnel	Mean Number Problems Correct/ Minute	GR PC CS PR NC NM	UK " " " " "	15000'= 6	0' vs 15000' Adapted vs Non p < .01 No difference " " " " " " " " " " " " " " " "	Computer Interaction Test
Altitude	Banderet	Experiment	Analysis	Sea Level to 28000'	N = 8 with 2 dropping	Paramedical	# Correct				% Loss at Altitude Baseline is Prior to 15000 ft.	Other Psychological Measures

	15	18	19	20	22	25	25	Post	p
GR	4	+6	4	3	5	8	13	10	.05
ST	3	6	3	4	5	13	14	4	.001
PC	+1	2	5	6	6	16	16	12	.001
CS	3	+2	+2	2	9	16	22	13	.01
PHT	+1	3	+7	+3	2	+1	3	5	NS
NPT	+4	+1	+4	+6	0	+1	+2	4	NS
THT	+2	+2	+3	–	0	1	7	6	NS

Pharmacological Agents	Kohl (Researcher Unavailable)	UK	-	-	-	-	-	-	-	-	-	-
Pharmacological Agents	Schiflett	Experiment	Proposal (Classified Study)	Tamazepam (Sleep Enhancement) vs No Tamazepam	N = 34	US Air Force SAC Flight Crews	To be Established	PHT THT NPHT GR PC CS RT(4)	To be Evaluated	To be Established	To be Evaluated	EEG
Pharmacological Agents	Wood	Experiment	Analysis	Dexedrine (Alertness Enhancement) Hyoscine Scopalamine (Motion Sickness Treatment)	N = 16	NA	UK	PHT THT NPHT GR PC CS SB MNK ACM	Battery Practiced for 6 Trials	UK	To be Evaluated	NA

General Areas of Research	Principal Investigator	Method of Investigation	Study Status	Variables to be Examined or Manipulated	Number of Subjects	Subject Organismic Variables or Selection Factors	Performance Score	APTS Subtests	Number of Trials	Number of Battery Applications	Sensitivity/Significance	Potential Correlates
Pharmacological Agents	Lieberman (Bandaret)	Experiment	Data Collection	Tyrosine (Neurotransmitter Precursor/ Potential Altitude Sickness Treatment)	N = 8	UK	To be Established	To be Established	To be Evaluated	To be Established	To be Evaluated	UK
Pharmacological Agents	Parth	Experiment	Analysis	Bone Marrow Transplant and Chemo/ Radiation Therapy	N = 22 Experimental N = 8 Control	Cancer Patients	Number Correct and Number Correct Minus Incorrect	GR PC SB CS MNK RT(1) RT(4)	UK	UK	Experimental groups show performance decrement 40 days following treatment. Original performance levels are reached 90 days following treatment.	NA
Pharmacological Agents	DeGioanni	-	-	-	-	-	-	PHT THT NPHT GR PC CS SB MNK ACM	-	-	-	-
Sleep Deprivation	Kiziltan	Experiment	Complete	36 Hours Sleep Deprivation and Time Period for Data Collection (0900, 1400, 2000)	N = 24	Foreign Military Personnel	Number Correct and Number Correct Minus Incorrect	CS SB VM ACM	UK	3	Sleep Dep: $p < .02$, $p < .08$, NE, NE; Time Per: NE, $p < .023$, $p < .04$, NE; DepXPer: NE, NE, $p < .017$, $p <$	NPRU Mood Scale Questionnaire
Sleep Deprivation	Hutchins	Experment	Complete	40 Hours Sleep Deprivation	-	-	-	-	-	-	-	-

General Areas of Research	Principal Investigator	Method of investigation	Study Status	Variables to be Examined or Manipulated	Number of Subjects	Subject Organismic Variables or Selection Factors	Performance Score	APTS Subtests	Number of Trials	Number of Battery Applications	Sensitivity/Significance	Potential Correlates
Fatigue and Workload	McCauley	Experimental	Analysis	Fatigue = 0 to 14 Hours Underway at Sea	N = 100 to 124	US Coast Guard Personnel	To be Established	PHT THT NPHT GR PC CS MNK	Approximate 6 (10 Practice Trials)	15	To be Evaluated	Subjective Fatigue Rating Endogenous Eyeblinks
Fatigue and Workload	May	Correlation	Analysis	Eye Movement and Workload	N = 5	NA	UK	ABP	To be Evaluated	UK	r = 0.35 to r = 0.99 Xr = 0.67	NA
Motion	McComas (Hutchins)	Experiment	Data Collection	Simulated Ship Motion	N = 16	UK	To be Established	CS SB VM ACM	UK	UK	To be Evaluated	UK
Motion	Jones	Experiment	Data Collection	Hull Type and Sea State	N = 30 to 40	US Coast Guard Personnel	Number Correct and Mean Time Between Responses	PHT THT NPHT GR PC	To be Evaluated	UK	Significant difference found for all subjects with mean time between responses as measure. Number Correct is not significantly influenced.	Motion Sickness Questionnaire Motion History Questionnaire
Learning Disabilities	Mellard (Researcher Unavailable)	Controlled Observation	UK	-	-	-	-	-	-	-	-	-
Learning Disabilities	Wilkes	Controlled Observation	Proposal	Acquisition Curves for Normal and LD Children to be Compared	N = 40	Subjects Selected Due to Global LD Characteristics and Paired with Normal	To be Established	To be Establish-	To be Evaluated	15	To be Evaluated	WISC Draw-a-Person SAT
Learning Disabilities	Williams (Researcher Unavailable)	UK	-	-	-	-	-	-	-	-	-	-

(a) Not Applicable (NA)
(b) No Effect (NE)
(c) Significant Effect (SE)
(d) Unknown (UK)