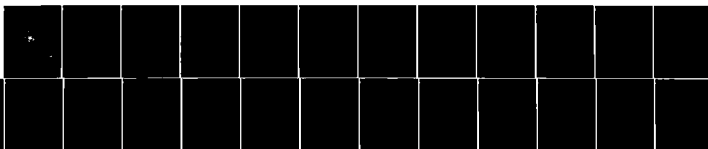AD-A115 676   NAVAL BIODYNAMICS LAB NEW ORLEANS LA   F/G 5/9
                ATTRIBUTE REQUIREMENTS FOR A SIMULATED FLIGHT SCENARIO MICROCOM--ETC(U)
                JAN 82   R H SHANNON, M KRAUSE, R C IRONS
UNCLASSIFIED   NBDL-82R004   NL

END
DATE
FILMED
7 82
DTIC

NBDL - 82R004

ATTRIBUTE REQUIREMENTS FOR A SIMULATED FLIGHT
SCENARIO MICROCOMPUTER TEST

RICHARD H. SHANNON, MICHELE KRAUSE, AND RICHARD C. IRONS

January 1982

DTIC
SELECTED
JUN 1 7 1982
F

NAVAL BIODYNAMICS LABORATORY
New Orleans, Louisiana

Approved for public release. Distribution unlimited.

82 06 17 019

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NBDL-82R004	2. GOVT ACCESSION NO. AD-A115 676	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Attribute Requirements for a Simulated Flight Scenario Microcomputer Test		5. TYPE OF REPORT & PERIOD COVERED Research Report
		6. PERFORMING ORG. REPORT NUMBER NBDL-82R004
7. AUTHOR(s) R. H. Shannon, M. Krause, R. C. Irons		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Biodynamics Laboratory Box 29407 New Orleans, LA 70189		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS MF58.524-02E-0002
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research and Development Command Bethesda, MD 20014		12. REPORT DATE January 1982
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

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

18. SUPPLEMENTARY NOTES

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

Repeated Measures, Human Performance Testing, Job Analysis, Microcomputer Testing, Test Battery, Practice Effects

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

Eighteen subjects practiced a video game of bombing and air combat maneuvering, Phantoms Five, on an APPLE microcomputer for 10 minutes a day for 15 days. The dependent variable was the combined score for number of hits and number of targets. Performance stabilized from Days 8 - 15 with a pooled reliability of .904. Eight reference tests which theoretically measure cognitive, perceptual, quantitative, and motor skills were selected and used as independent variables. Stabilized performance on these tests was observed after a period

ABSTRACT (Block 20) Continued

of practice which was predetermined from previous experimentation. Attributes
of the Phantoms Five[R] were isolated using a structured job analytic tool
(Position Analysis Questionnaire, PAQ). A principal components analysis of
the measures that correlated with the dependent variable resulted in a one
factor solution explaining 66 percent of the variance. This factor represented
the constructs of Flexibility of Closure, Perceptual Speed, and Spatial
Scanning. The synthetic validity approach using the PAQ indicated that Form
Perception, Perceptual Speed, Closure, and Spatial Visualization were the most
critical attributes of the Phantoms Five[R]. It was concluded that construct
validity was established since there was a strong similarity between the
attribute requirements attained by correlating the stabilized scores of
independent and dependent variables and by the PAQ analysis of task functions.

Accession For

NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	

Distribution/

Availability Codes

| | Avail and/or |
| | Special |

A

DTIC
COPY
INSPECTED
2

NBDL - 82R004

ATTRIBUTE REQUIREMENTS FOR A SIMULATED FLIGHT
SCENARIO MICROCOMPUTER TEST

RICHARD H. SHANNON, MICHELE KRAUSE, AND RICHARD C. IRONS

January 1982

Bureau of Medicine and Surgery
Work Unit No. MF58.524-02E-0002

Approved by   Released by

Channing L. Ewing, M. D.   Captain J. E. Wenger MC USN
     Chief Scientist   Commanding Officer

Naval Biodynamics Laboratory
Box 29407
New Orleans, LA   70189

SUMMARY PAGE

PROBLEM

The purpose of this research was to study a flight simulation task of bombing and air combat maneuvering (Phantoms Five) over a 15 day period in order to determine: (1) the amount of practice that this test requires to be stabilized, (2) the utility of pretesting the eight reference tests to a predetermined stabilization point, (3) the relationships between eight reference tests (those believed to measure a specific ability) and this complex criterion test, and (4) whether the PAO can be used to recommend a battery of tests that would predict performance.

FINDINGS

(1) The Phantoms Five test stabilized on days 8-15 with an intraclass reliability coefficient of .542 on each day and a pooled coefficient of .904 on the eight days. A significant linear trend was observed during these eight days with a daily increase of .63 and 1.31, respectively, in the number of hits and targets. (2) The utility of specifying and using predetermined periods of practice was demonstrated in this experiment by the reliabilities within a test and the correlational pattern between tests. (3) A principal components analysis of the independent variables that correlated with the Phantoms Five resulted in a one factor solution explaining 66 percent of the variance. This factor represented the constructs of flexibility of closure, perceptual speed, and spatial scanning. (4) The synthetic validity approach using the PAO indicated that form perception, perceptual speed, closure, and spatial visualization were the most critical attributes of the Phantoms Five.

RECOMMENDATIONS

(1) Tests that are to be used for repeated measurement should be practiced by the subjects prior to being used to obtain data. The required amount of practice should be determined from data obtained in a standard environment. (2) Differences in skill levels among subjects must be considered when pretesting periods are being established. (3) The PAO can be utilized to establish synthetic or job component validity.

Introduction

A recent article (Jones, Kennedy, & Bittner, 1981) discussed the merits of using the ATARI[R] Video Computer System game of Air Combat Maneuvering as a performance test. The authors claimed that this two-dimensional pursuit tracking task had substantial face validity to military jobs because of its similarity to radar and sonar interception. After analyzing the performance of 22 subjects over a 15 day period, the results indicated that the task stabilized after Day 6 with an average correlation among days of .927.

The present paper extends this work by reviewing a similar task, Phantoms Five[R] (Gebelli, 1980) which is a simulation of bombing and air combat maneuvering using the APPLE[R] microcomputer. In addition, the results of eight performance tests previously studied at this Laboratory were utilized as reference or marker tests. The basic constructs of these tests had been isolated in Ekstrom, French, Harmon, and Derman (1976). Correlating performance on marker tests with performance on an unknown task such as the Phantoms Five[R] in order to determine the specific abilities being measured has been recommended by at least two researchers: Cattell (1966) and Fruchter (1966). In addition, attributes of the Phantoms Five[R] were isolated using a structured job analytic tool (Position Analysis Questionnaire, PAQ) developed by McCormick, Jeanneret, and Mecham (1972). McCormick (1979) claims that synthetic or job component validity can be established through the PAQ. A comparison between the abilities isolated through correlation and the attributes determined using the PAQ was performed. If this comparison is successful synthetic or job component validity then would acquire construct validity.

In remaining sections of the Introduction, the selection of the APPLE[R] computer system for psychological testing, stability requirements of a test, Position Analysis Questionnaire, and finally the purpose of this paper will be discussed.

Selection of an Automated Test System

An aim of this laboratory is to assess psychological performance while subjects are experiencing the effects of impact acceleration, ship motion, and vibration.

It was determined that developing an APPLE[R] microcomputer-based system would provide the most efficient means of measuring performance in these environments (Irons, Shannon, Krause, & Patsfall, 1981). In addition to providing automatic stimulus presentation and data collection, microcomputer-based testing has an added advantage of being adaptive to varying performance levels. After examining existing systems and reviewing available literature, the APPLE[R] system was chosen on the basis of several criteria: (a) low cost, (b) portability, (c) system independence, (d) availability of hardware/software, (e) color graphics capability, (f) available languages (e.g., BASIC and PASCAL), (g) voice input/output capability, (h) light pen input, (i) high speed serial and parallel input/output, and (j) analog to digital and digital to analog input/output.

To facilitate simultaneous testing at different "stations", a NESTAR Cluster/One Model A[R] was purchased. Each microcomputer is channelled through the NESTAR[R] system, which gives the added capability of having: (a) simul-

taneous testing on up to 64 "stations", (b) a centralized pool of psycho-
logical tests, (c) centralized data collection and analysis, (d) 67.2 MBytes
of information stored, and (e) testing as far away as 1000 feet from the
central unit, or at any location accessible by voice grade, telephone/
radiotelephone communications. The present system incorporated these advan-
tages within a psychological testing laboratory by having eight microcom-
puters in a network system that can be controlled by one experimenter.

Stability Requirements of A Test

When a test such as the Phantoms Five^R is administered on repeated days,
it will demonstrate the effects of practice. These effects may appear in
the daily means, variances, or correlations. There is a point, however, with
continued practice that the task becomes stabilized (Jones, 1980). Stabili-
zation occurs when the group daily means become asymptotic or increase with
a slight constant slope, the daily variances among subjects are constant, and
the intertrial correlations are equal. If a task does not become stabilized,
the assumption of compound symmetry is not met (Winer, 1971). In addition,
stability indicates that the performances of subjects are temporally gener-
alizable (Jensen, 1980), and that the task composition and the subjects'
abilities remain constant over time (Alvares & Hulin, 1972). The Steiger
MULTICORR computer program (Steiger, 1980) is used to test the hypothesis
of equal correlation. An average correlation of the hypothetically homo-
geneous matrix is determined and utilized as the null or comparison correl-
ation to all of the other correlations in the matrix.

The average correlation among the stabilized trials approximates the
intraclass correlation coefficient for each day. If either correlation is
placed in a Spearman-Brown prediction formula, the result is a pooled reli-
ability coefficient for $\underline{N}$ days (Winer, 1971; Nunally, 1967). If each stabil-
ized day or trial is considered to be a part of the total test as represented
by the total stable period of trials, then the total scores or means for each
subject are representative of an individual's performance for that test and
the pooled coefficient is the reliability for that test.

Position Analysis Questionnaire

The Position Analysis Questionnaire (PAQ) (McCormick, Jeanneret, &
Mecham, 1972) is a structured job analytic tool that is composed of 194 job
elements. A specific rating scale is designated to be used with each job
element. In general, "extent of use" and "importance to the job" are the
two scales that are most frequently used within this questionnaire having
anchor points from 0 to 5. The elements are of a worker-oriented nature
that tend to imply human activities that are involved in jobs. The job
elements in the PAQ are organized in the following six divisions: informa-
tion input, mental processes, work output, relationships with others, job
context, and other job characteristics. The PAQ element scores are con-
verted to 45 job factor or dimension scores by using factor loadings devel-
oped for 2200 jobs (Mecham et al., 1977). The 45 factors (dimensions)
include 32 and 13, respectively, for six divisions analyzed separately and
combined.

The PAQ is being used to establish a procedure for developing psycho-
logical batteries at this laboratory which will have synthetic, component

or construct validity. The concept of job component validity assumes that the human requirements of any given job are comparable with other jobs having equal amounts of similar work activities (McCormick, 1979). The procedure for establishing validity includes: (a) identification of the work functions and their relative importance, (b) determination of human attributes associated with successful performance of the work functions, and (c) combination of the attribute requirements associated with each function into an estimate of the requirements for the entire job. If the job component validation is successful, then the human attributes and work functions acquire construct validity. Of course, a job component validity effort presumes that a taxomony of work functions and a method for measuring all relevant human attributes are available. Both of these needs can be met through the use of the PAQ and the proper selection of psychological tests to measure human attributes.

Another study by McCormick and his associates (Marquardt & McCormick, 1972) at Purdue University was of assistance in determining the attribute requirements of a job. In this study, between 8 - 11 experts (psycologists who were members of APA) were asked to rate the relevance of 49 human attributes of an "aptitude" nature to 182 of the 194 items within the structured Position Analysis Questionnaire (PAQ). The following twelve PAQ numbered elements were not analyzed because they were open-ended with any response being possible: 44, 60, 127, 160, 181, 188 - 194. A 6-point scale (0 - 5) involving "the degree of relevance of an attribute to a job element" was used. The reliability coefficients of the pooled ratings for these attributes ranged from .796 to .964. The 49 abilities used in this analysis were very similar to abilities or attributes listed in other studies in the literature (Theologus, Romashko, & Fleishman, 1970; Pawlik, 1966; Ekstrom et al., 1976). A principal components analysis and varimax rotation of the matrix containing 49 attributes by 182 elements resulted in a seven attribute dimension structure (McCormick, 1979). This factor model is depicted in Shannon (1982b) with the following outline:

1) General Physical Skills
2) Cognitive Skills
3) Visual Perception/Interpretation
4) Psychomotor Skills
5) Chemical Senses
6) Physical Response/Coordination Versus Imaginative Orientation
7) Quantitative Skills

Purpose

The purpose of this research was to study a flight simulation task of bombing and air combat maneuvering (Phantoms Five[R]) over a 15-day period in order to determine:

1) the amount of practice that is required for performance on this test to stabilize.

2) the utility of pretesting on the eight reference tests to a predetermined stabilization point.

3) the relationships between eight reference tests (those believed to measure a specific ability) and the complex criterion test.

4) whether the PAQ can be used to recommend a battery of tests that would predict performance.

Method

Subjects

Eighteen Navy enlisted men were the subjects for this experiment. All subjects met or exceeded rigid medical standards set for environemntal research subjects as described by Thomas, Majewski, Ewing, and Gilbert (1978). National and international guidelines pertaining to voluntary informed consent were adhered to in this experiment.

Task Description

Eight tests reported in the research literature to measure cognitive, perceptual, or motor abilities were employed in this study. A Vertical Addition (VA) test similar to the numerical facility tests described by Ekstrom et al. (1976) was administered. Grammatical Reasoning (GR) modeled after Baddeley's test (1968) and Pattern Recognition (PR) based on Fitts' histoforms (Fitts, Weinstein, Rappaport, Anderson, & Leonard, 1956) were also used. These two tests resemble tests of logical reasoning and perceptual speed, respectively, as outlined by Ekstrom et al. (1976). Alternate forms of these three tests were randomly generated by computer programs which are publicly available (Carter & Sbisa, 1982). Three additional tests, Flexibility of Closure (FC), Speed of Closure (SC), and Visualization (V), each with 20 alternate forms, were provided by Moran (Moran, Kimble, & Mefferd, 1964). The FC of the Moran et al. (1964) battery corresponds to the FC construct described by Ekstrom et al. (1976); however, V and SC are described by Ekstrom et al. (1976), respectively, as Spatial Scanning and Verbal Closure. The seventh test, Hidden Figures (HF), was constructed in the manner of Ekstrom et al.'s (1976) Flexibility of Closure test. Fifteen alternate forms of HF were constructed by Shannon (1982a). Finally, a two-choice visual reaction time task was included. Tests described above were presented in a paper and pencil format, except for the reaction time test, which utilized a device constructed for this laboratory from schematics furnished by Teichner's Laboratory at New Mexico State University.

Phantoms Five[R], a more complex task, simulated air combat maneuvering (ACM) and ground target bombing (GTB) in two separate phases. Beginning in the GTB mode, the subject must direct his airplane (via paddle controller) and drop bombs on ground targets (via button on the paddle controller). Ten to 100 points are scored for bombing approximately 100 different targets, and either half or all points are lost for hitting two specific targets. Periodically throughout the task, the ACM mode will switch on. During this phase, the subject's perspective changes from controlling a distant aircraft (as in GTB mode) to controlling an airplane from the cockpit. Shots are fired (via controller button) at other aircraft occupying the airspace, and 10 points are scored for each plane hit. On the average, twenty percent of the total time on task is devoted to the ACM mode. Large variations in the proportion of time spent in each phase are attributable to the different skill levels of

each subject. More specifically, the variation appears to be grounded in the number of "good" targets bombed (i.e., those that add points to the total score), the number of "bad" targets hit (i.e., those that subtract points from the total score), and the number of times the aircraft is shot down by anti-aircraft guns during the GTB mode. Each subject begins the task with five aircraft and the task continues until all aircraft have been shot down.

Procedure

All tests in this study were introduced to the subjects two days prior to the beginning of the experiment. Some subjects were tested previously on a few tests. During this session, instructions for each task were clarified, practice problems were worked, and the purpose of the experiemnt was reviewed.

The seven cognitive tests (HF, V, SC, FC, GR, PR, and VA) along with the reaction time task, were administered once per day over an eight day period to stabilize each subjects' performance before comparison data were collected. Tasks requiring the most practice were administered throughout, whereas other tests were added to the sessions in time to be sufficiently practiced. The order of testing was randomized between days but remained the same within days. Each of the eight reference tests was administered in accordance with stabilization requirements determined by previous research at this laboratory (Bittner, Carter, & Krause, 1981; Krause, Bittner, & Carter. 1982; Shannon, 1982a):

HF	5	min.	Days 1 – 8
RT	5	min.	Days 1 – 8
V	3	min.	Days 2 – 8
GR	1	min.	Days 3 – 8
FC	3	min.	Days 5 – 8
PR	2	min.	Days 6 – 8
SC	2.5	min.	Days 6 – 8
VA	4	min.	Day 8

By the eighth day, the testing session was 30 minutes in length. On Days 9 and 10, the data to be used in the comparison with Phantoms Five[R] were collected.

A portion of the subject pool in this experiment had been tested on some of the reference tests during a previous experiment. Prior performance was taken into account here, and is reflected in the analysis. In this way, carry-over effects could be studied and compensated for. Subjects in the current study who were practiced on one or more of the reference tests are referred to as "non-naive". Likewise, those exposed to these tests for the first time at this laboratory are labeled "naive" throughout this paper. In this study, the number of non-naive subjects (in parentheses) by test were: Hidden Figures (6), Visualization (3), Flexibility of Closure (3), Speed of Closure (3), Vertical Addition (3), Pattern Recognition (9), Grammatical Reasoning (3), and Reaction Time (7). Previous experimental data for the non-naive subjects for the appropriate criterion days were used in the comparison with the Phantoms Five[R] simulation. However, the second set of data was also collected for comparison with the first set. For example, practice on the Pattern Recognition test in the present experiment (Phantoms

Five^R) was given on Days 6 - 8, while stable performance data were collected on Days 9 and 10. Comparable data on the previous experiment involving Pattern Recognition were Days 1 - 3 for practice and Days 4 and 5 for stable performance measurements.

The Phantoms Five^R test was individually administered in a four by six foot booth to each subject, who sat approximately two feet from a 13 inch square color monitor that presented the task. Each individual was instructed to record his score at the end of each trial, and reset the task by pushing the button on the paddle controller. At the end of 10 minutes, a buzzer sounded signalling the subjects to stop the task and record his last score. Ten minutes of training was given on Phantoms Five^R for 15 consecutive workdays. All testing was conducted in the mornings with the seven paper and pencil tests followed by reaction time and Phantoms Five^R.

PAQ Analysis of Phantoms Five^R

A structured job analysis of the Phantoms Five^R was conducted independently by two analysts. The instrument used was the Position Analysis Questionnarie (PAQ) developed by McCormick, Jeanneret, and Mecham (1972). Interrater reliability across the 194 elements was .876. The two analysts then discussed differences in their scores, which resulted in a pooled set of PAQ ratings. These PAQ element scores were then converted to 31 dimension scores using the factorial model outlined in Mecham, McCormick, and Jeanneret (1977). One dimension (#28) score was not computed since it was composed of job elements that did not have attribute ratings. Scores were also not determined for the 13 dimensions involving the combined divisional analysis. A mean was computed for each dimension (a sample of PAQ elements) and compared with the population mean (182 PAQ ratings). A series of t-tests with a correction for sampling from a finite population was conducted using a .1 alpha level, one-tail. Since Type II error was considered more important than Type I at this stage of analysis, the alpha level was thought to be appropriate, with three out of 31 dimensions expected to be significant by chance. Within the eight significant dimensions, an element rating of 2.5 and above was labeled as critical. This cut-off rating is the midpoint on the 0 - 5 scale. Critical elements and significant dimensions are listed in Appendix A.

The next phase of the analysis was the identification of significant attributes and attribute dimensions for the critical elements of the Phantoms Five^R. This information is outlined in Appendix B. The procedure for collating scores and isolating critical attributes followed from the sums of attribute ratings for the critical PAQ elements. A sample mean was determined for each attribute across the critical elements and is compared with the population average (all 182 ratings within an attribute). Statistical significance was computed using a t-test with a correction for sampling from a finite population. A .005, one-tail level of significance was used to correct for possible Type I error among the 49 attributes compared.

Results

Phantoms Five^R

Table 1 depicts the means and standard deviations for the three measures on the Phantoms Five^R test: number of hits (air combat maneuvering, ACM), number of targets (ground target bombing, GTB), and number of hits plus number

of targets (ACM & GTB combined). ACM and GTB are in actual unit scores, while the combined measure is the Z scores of both ACM and GTB added together. Differential stability was achieved on each of the three criterion measures by Days 8 - 15, as shown in Table 2, using the Steiger MULTICORR program. Since the three measures are highly intercorrelated (hits/targets = .899, combined/hits = .975, combined/targets = .974), further discussion of the data will mainly be concerned with the combined hits & targets scores. Table 3 contains the intercorrelations among Days 8 - 15 (stable period) for this combined measure. The average correlation is .553. If Day 15 is ignored, the average correlation is .593 indicating that there was a lowering of the reliability on the last day. Table 4 Depicts an anlysis of variance for Days 8 - 15 with the following results:

(1) a significant linear trend over days (p$\leq$.01) for the combined score which explained 84% of the daily variance with a slight increasing slope of .17 each day (this value in actual score units is # hits = .63 increase per day and # targets = 1.31 increase per day).

(2) homogeneous daily variances for the combined score ($F_{max}(18, 8)$ = 1.61, NS).

(3) the unbiased intraclass reliability coefficient for each day of .542 (p$\leq$.05) and the pooled reliability for Days 8 - 15 (total test) of .904. This pooled estimate is based upon 80 minutes of testing and 70 minutes of practice for each subject over the 15 days.

Reference Tests

Table 5 contains the means and standard deviations for each of the tests on the two observation days as well as both days combined. The experimental data for the non-naive subjects during the first session are combined in this table and the tables that follow with the naive subject data of the second session. Since the earlier data for Grammatical Reasoning were not available, the three non-naive subjects were omitted from the computations on this test (n = 15). In addition, the reliability of the performance on both days is listed in Table 5 with a low of .709 on Pattern Recognition and a high of .931 on Vertical Addition. Number of corrected responses minus a correction for guessing was recorded for Grammatical Reasoning and Hidden Figures (1.0 and .25 was subtracted for errors, respectively). Reaction times were measured in milliseconds. On the remaining five tests, number correct was the score used.

Mean performance levels at various periods of time on seven of the eight tests (Grammatical Reasoning is omitted) by the non-naive subjects are depicted in Table 6. From these data even with the small samples involved, it can be seen that: (1) mean performance on the two criterion days improved from session one to session two, and (2) mean performance on the last six days of session one were more similar to the criterion days of session two than session one. These observations have important implications for future experiments at this laboratory because of the potential retention of skills and abilities over large periods of time. Session one for the Hidden Figures, Vertical Addition, and Choice Reaction Time tests was conducted six months prior to session two, while the time difference between both sessions on the other four tests was one year.

Table 1: Means and Standard Deviations of Scores (Number of Hits,
Number of Targets, Combined Hits and Targets) Among 18 Subjects on the Phantoms
Five Test Over 15 Days

	Scores		
Days	Hits (M/SD)[*]	Targets (M/SD)[*]	Combined (M/SD)[**]
1	11.6/ 8.0	45.0/14.5	−1.90/.92
2	15.4/12.4	52.7/17.4	−1.39/1.24
3	14.8/ 8.6	64.8/20.5	−1.05/1.22
4	16.4/10.4	70.5/20.4	− .75/1.33
5	18.0/ 8.8	72.1/17.3	− .59/1.11
6	20.3/12.0	78.7/27.5	− .21/1.67
7	20.4/11.8	78.3/19.8	− .22/1.43
8	23.6/15.7	82.5/32.0	.14/2.03
9	24.3/13.5	88.9/27.5	.39/1.77
10	26.7/16.0	93.2/34.3	.71/2.19
11	26.3/15.8	87.5/30.2	.49/2.04
12	26.9/13.7	98.6/33.3	.89/1.99
13	26.7/16.6	93.7/36.8	.72/2.31
14	30.8/13.1	97.6/30.8	1.13/1.85
15	34.1/15.1	105.9/34.1	1.63/2.00
8-15 combined[***]	27.4/15.0	93.5/32.5	.76/2.03

  * Hits and targets in actual score units
 ** Combined score = Z score of hits + Z score of targets
*** Average mean and standard deviation of Days 8 − 15.

Table 2: Results of the Steiger MULTICORR Program Analysis
for the 3 Phantoms Five Criterion Values Among 18 Subjects over the Stabilized
Period (Days 8 - 15)

Measure	x^2	df	ρ	$\bar{r}$
# Hits (ACM)	27.5	27	.44	.545
# Targets (GTB)	18.26	27	.90	.545
# Hits & # Targets	22.4	27	.72	.553

Table 3: Correlations Between Days 8 Through 15 (Stable Period) for the
Combined # Hits & # Targets Scores on the Phantom Fives Test
Over 18 Subjects

Days	8	9	10	11	12	13	14	15
8								
9	.476							
10	.518	.600						
11	.625	.727	.781					
12	.684	.458	.608	.627				
13	.602	.541	.525	.535	.509			
14	.589	.681	.656	.595	.435	.675		
15	.558	.606	.278	.446	.334	.222	.596	

Table 4: Analysis of Variance for the Combined # Hits & # Targets Scores
on the Phantoms Five Test During Days 8 - 15 (Stable Period)
Among the 18 Subjects

Source	SS	df	MS	F	P
Subjects	337.6	17	19.9	10.6	$>.001$
Days	27.0	7	3.85	2.07	$>.05$
linear	22.6	1	22.6	12.1	$>.01$
nonlinear	4.4	6	.73	.39	NS
Residual	222.0	119	1.87		
Total	586.6	143			

Comparison of Tests

Table 7 depicts the product-moment correlations between the eight refer-
ence tests and the three criterion measures of the Phantoms Five$_R$ simulation.
A principal components analysis and varimax rotation were conducted on the
eight reference tests and the combined score of the Phantoms Five$_R$. Three
factors resulted: visual perception explaining 34 percent of the total var-
iance$_R$ (Hidden Figures, Flexibility of Closure, Pattern Recognition, Phantoms
Five$_R$), (2) cognitive/quantitative skills with 25 percent of the total variance
(Speed of Closure, Visualization, Vertical Addition, Grammatical Reasoning),
and (3) Reaction Time with 14 percent of the total variance. The Flexibility
of Closure (FC), Hidden Figures (HF), and Pattern Recognition (PR) tests of
the first factor are significantly related ($p \leq .1$, two tailed) to the three
simulation scores as well as with each other. The average correlation among
the three tests and the combined measure of the Phantoms Five is .608. These
three paper and pencil tests represent the perceptual factors outlined by
Ekstrom and his associates (1976) as:

(1) Flexibility of Closure: "the ability to hold a given visual percept
or configuration in mind so as to dissembed it from other well defined per-
ceptual material" (FC and HF tests).

(2) Perceptual Speed: "speed in comparing figures or symbols. scanning
to find figures or symbols, or carrying out other very simple tasks involving
visual perception" (PR seems to measure this factor).

The second factor has an average correlation of .438 with the associations
between GR/V and GR/VA being nonsignificant. Of the four variables, only$_R$
Visualization (V) correlated significantly with the three Phantoms Five$_R$
measures. In addition, V also had significant relationships with the other

Table 5: Means, Standard Deviations, Correlations of Scores
on the Reference Tests Among 18 Subjects* Over 2 Days

Test (Measure)	Day 1 (M/SD)	Day 2 (M/SD)	Combined Days (M/SD)	Correlation Days 1/2
Visualization (# correct)	47.6/ 9.4	48.2/ 8.7	47.9/ 8.6	.803
Flexibility of Closure (# corr)	12.1/ 5.7	12.4/ 4.9	12.3/ 5.1	.806
Speed of Closure (# correct)	25.1/ 7.8	27.3/ 6.7	26.2/ 6.8	.751
Hidden Figures (# corr - .25 # errors)	4.6/ 2.7	6.0/ 4.3	5.3/ 3.3	.760
Gram. Reasoning* (# corr - # errors)	12.6/ 6.9	13.0/ 8.1	12.8/ 7.2	.861
Vert. Addition (# correct)	35.7/11.6	37.2/11.4	36.4/11.3	.931
Pattern Recog. (# correct)	25.1/ 7.4	23.9/ 5.6	24.4/ 5.9	.709
Choice Reaction Time (msecs)	237.0/36.9	239.0/33.9	238.4/34.4	.889

* (grammatical reasoning, N = 15)

Table 6: Means and Standard Deviations of Reference Test Scores
on Two Testing Sessions Among the Non-Naive Subjects

| Test | Session One | | Session Two |
	Stable Performance Days M/SD	Last 6 Days M/SD	Stable Performance Days M/SD
Pattern Recog. (N = 9)	22.8/ 6.7 (days 4/5)	28.9/ 6.7 (days 10–15)	31.7/ 6.9 (days 4/5)
Reaction Time (N = 7)	218.6/14.1 days (9/10)	217.9/13.5 (days 10–15)	214.6/13.1 (days 9/10)
Flex. of Closure (N = 3)	16.0/ 8.4 (days 5/6)	20.8/ 5.3 (days 15–20)	19.5/14.4 (days 5/6)
Visual- ization (N = 3)	41.7/ 8.1 (days 8/9)	49.2/ 4.4 (days 15–20)	47.8/12.4 (days 8/9)
Speed of Closure (N = 3)	32.0/ 6.1 (days 4/5)	39.6/ 2.3 (days 15–20)	34.8/ 5.5 (days 4/5)
Vertical Addition (N = 3)	32.3/ 6.7 (days 2/3)	38.1/ 6.7 (days 10–15)	39.2/12.8 (days 2/3)
Hidden Figures (N = 6)	7.4/ 3.5 (days 9/10)	data not available	12.8/ 8.1 (days 9/10)

variables in the first factor (V/FC = .521; V/HF = .534; V/PR = .521). The Visualization test is categorized under the Spatial Scanning factor which is defined as "the speed in exploring visually a wide or complicated field" (Ekstrom et al., 1976).

The average correlation among the four reference tests, which are significantly related to the three measures of the Phantoms Five[R], and the combined score of the flight simulation is .573. A principal components analysis of this five variable matrix results in a one factor solution explaining 66 percent of the variance and having the following loadings:

Visualization	.761
Flexibility of Closure	.845
Hidden Figures	.768
Pattern Recognition	.863
Phantoms Five[R]	.821

These reference tests are represented by the constructs of Flexibility of Closure, Perceptual Speed, and Spatial Scanning.

Appendix B depicts the Visual Perception/Interpretation dimension and its attributes of Visual Form Perception, Perceptual Speed, Closure, and Spatial Visualization as being the most important to Phantoms Five[R] performance. Marquardt and McCormick (1972) defined these attributes as:

1) Visual Form Perception - "Ability to perceive pertinent detail or configuration in a complex visual stimulus."

2) Perceptual Speed - "Ability to make rapid discriminations of visual detail."

3) Closure - "Ability to perceptually organize a chaotic or disorganized field into a single perception."

4) Spatial Visualization - "Ability to manipulate visual images in two or three dimensions mentally."

Therefore, one can conclude that the PAQ analysis of the Phantoms Five[R] identified those constructs which would have the highest correlations with performance on the simulation task.

Discussion

Four goals of this paper were listed in the Introduction. Each of these goals will be discussed in this section under its own heading.

Stability of the Phantoms Five[R].

Phantoms Five[R] is a complex scenario or simulation involving air combat maneuvering and bombing. Stability of the means, variances, and intertrial correlations was achieved on Days 8 - 15. However, the average correlation (.553) and the intraclass reliability coefficient (.542) on each day were moderate. Therefore, to improve test reliability, the eight days of the stable period were pooled resulting in a reliability coefficient of .904.

Table 7: Correlations.* Among the Means for the 8 Reference Tests (2 Days) and the 3 Phantoms Five
Criterion Values During the Stabilized Period (Days 8-15) Over 18 Subjects

	V	FC	SC	HF	GR**	VA	PR	RT	HIT	TGT	COMB
V											
FC	.521										
SC	.479	.631									
HF	.534	.600	.365								
GR	.398	.390	.596	.582							
VA	.560	.292	.470	.132	.122						
PR	.521	.701	.594	.523	.295	.325					
RT	.060	-.196	-.212	.022	-.064	.147	.020				
HIT	.452	.506	.352	.486	.259	.288	.713	.088			
TGT	.543	.677	.424	.408	.183	.452	.757	-.084	.899		
COMB	.510	.607	.398	.460	.228	.378	.755	.003	.975	.974	

* (.401 = .1, .468 = .05, .590 = .01, using two tails for N = 18)

** (.441 = .1, .514 = .05, .641 = .01, using two tails for N = 15)

This pooled estimate is based upon 80 minutes of testing and 70 minutes of practice for each subject over the 15 days (10 minutes per day). In addition, within the stable period, there was a significant linear trend with increases of .63 and 1.31, respectively, in the number of hits and the number of targets per day.

If one decided that 80 minutes of testing was too long and wanted to use 30 minutes, what then would be the reliability of the scores over three days of testing? Using the Spearman-Brown prophesy formula, the pooled reliability would be .780, which is a respectable estimate. The problem, therefore, is not just to determine how stable or consistent the test scores are from one day to the next, as indicated by the constancy of the relative standing of subjects on the same test, but to achieve a specific level of internal consistency or reliability. For example, if the stable period of the Phantoms Five^R was separated into two 30 minute test periods (Days 8 - 10 and 12 - 14) with Day 11 separating both groups, the pooled internal reliability within both sessions would be .773 and .779, while the stability coefficient between both sessions would be .855 (Jensen, 1980).

To summarize, future experiments using a complex scenario such as Phantoms Five^R may also have moderate intraclass reliablities for each trial. The pooling of data, therefore, may be necessary for specific levels of reliability and stability to be achieved. If the goal is to measure specific attributes or abilities, then the consistency and retentive qualities of performance are essential. Finally, daily performance may continually improve on Phantoms Five^R because of its complexity at a slow but constant linear rate. This increase in mean performance over time further underscores the need for stable performance measurements.

Utility of Pretesting to Specified Levels

Previous research at this Laboratory had determined the appropriate levels of practice or pretesting that was necessary for stabilized performance. The utility of using this information can be measured by the intertrial correlations between the two criterion days, and the consistency of performance among the tests. The reliability and/or stability of performance on both days for the eight reference tests was highly satisfactory: Pattern Recognition (.709), Speed of Closure (.751), Hidden Figures (.760), Visualization (.803), Flexibility of Closure (.806), Grammatical Reasoning (.861), Choice Reaction Time (.889), and Vertical Addition (.931). These correlations are comparable to the coefficients of the previous research on each test. The consistency among the correlations in Table 7 is further evidence that pretesting was successful. There were expected significant relationships among the perceptual constructs as well as a high commonality among these tests.

Although the sample sizes were small for those individuals who had been observed previously on the same tests, there was consistent evidence in seven reference tasks of a retention of skill levels over a 6 - 12 month period between testing. Mean performance on the two measured criterion days improved from session one to session two; and the scores on the last six days of session one were more similar to session two than to session one.

To summarize, there is a definite need for tests to be practiced if stabilized performance is desired. Also, one must consider differences in

skill levels among the subjects when pretesting periods are being established. In this experiment, only data from naive subjects on a particular test was used for comparison purposes. The utility of the specified pretesting levels of performance utilized in this research was demonstrated by the reliabilities within a test and the correlational pattern between tests.

Correlational Pattern of Tests

Eight reference or marker tests were selected which theoretically measure the Flexibility of Closure (Flexibility of Closure, Hidden Figures), Spatial Scanning (Visualization), Verbal Closure (Speed of Closure), Logical Reasoning (Grammatical Reasoning), Perceptual Speed (Pattern Recognition), Numerical Facility (Vertical Addition) and Reaction Time (Choice Reaction Time). These tests were used to determine whether the specific constructs that they purport to measure are found in the Phantoms Five[R] simulation, as evidenced by the correlational coefficient. This methodology of using marker variables that have been shown to mark the location of a given concept and then to observe the relationships between these independent variables with a dependent, criterion task is supported by Cattell (1966) and Fruchter (1966). Table 7 indicates that there are four tests which significantly correlate with the three measures of the Phantoms Five[R], the criterion or dependent variable. A principal components analysis between these four independent measures and the combined score of the simulation task resulted in only one factor, which explained 66 percent of the variance with an average correlation of .573.

To summarize, the correlational pattern among the reference tests and the Phantoms Five[R] indicates that this simulation task is composed of the following perceptual factors or constructs: Flexibility of Closure, Perceptual Speed, and Spatial Scanning.

Construct Validity Using the PAQ

McCormick (1979) claims that synthetic or job component validity can be established using the Position Analysis Questionnaire (PAQ). If the job component validation phase is successful, then the human attributes and work functions acquire construct validity. In this study, the procedure that he outlined was followed in our analysis of the Phantoms Five[R]: (1) the relative importance of the PAQ elements was established (Appendix A), (2) the attribute requirements of the total task were determined from the critical elements (Appendix B), and (3) correlational procedures were used to assess the fit between proposed and observed attributes or constructs. Appendix B depicts the Visual Perception/Interpretation dimension and its attributes of Visual Form Perception, Perceptual Speed, Closure, and Spatial Visualization as being the most important to Phantoms Five[R] performance.

In summary, construct validity appears to have been established using the PAQ. There is a strong similarity between the attribute requirements attained by correlating reference and criterion variables and by PAQ analysis of task functions.

References

Alvares, K, & Hulin, C. Two explanations of temporal changes in ability-skill relationships: A literature review and theoretical analysis. Human Factors, 1972, 14, 295-308.

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

Bittner, A. C., Jr., Carter, R. C., & Krause, M. Performance Tests for Repeated Measures: Moran and Computer Batteries. (Research Report No. NBDL-81R012). New Orleans: Naval Biodynamics Laboratory, 1981.

Carter, R. C., & Sbisa, H. Human Performance Tests for Repeated Measurements: Alternate Forms of Eight Tests by Computer. (Research Report No. NBDL 82R003), New Orleans: Naval Biodynamics Laboratory, 1982.

Cattell, R. B. The meaning and strategic use of factor analysis. In R. G. Catell, Ed. Handbook of Multivariate Experimental Psychology. Chicago: Rand McNally, 1966.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. Manual for Kit of Factor-Referenced Cognitive Tests. Princeton, N. J.: Educational Testing Service, 1976.

Fitts, P., Weinstein, M. Rappaport, M., Anderson, N., & Leonard, J. Stimulus correlates of visual pattern recognition: A probability approach. Journal of Experimental Psychology, 1956, 1-11.

Fruchter, B. Manipulative and hypothesis-testing factor-analytic experimental designs. In R. G. Cattell, Ed. Handbook of Multivariate Experimental Psychology, Chicago: Rand McNally, 1966.

Gebelli, N. Phantoms Five[R], Sacramento, California: Sirius Software, Inc., 1980.

Irons, R. C., Shannon, R. H., Krause, M., & Patsfall, M. Automated psychological testing and generic workstation development utilizing general purpose microcomputers. Proceedings of the International Workshop on Research Methods in Human Motion and Vibration Studies, New Orleans, 1981, in press.

Jensen, A. R. Bias in Mental Testing. New York: Free Press, 1980.

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

Jones, M. N. Stabilization and Task Definition in a Performance Test Battery (Monograph No. NBDL-M001). New Orleans: Naval Biodynamics Laboratory, 1980.

Krause, M., Bittner, A. C., Jr., & Carter, R. C. Repeated Measures on a Choice Reaction Time Task (Research Report No. BNDL-82R006). New Orleans: Naval Biodynamics Laboratory, 1982.

Marquardt, L. D., & McCormick, E. J. Attribute Ratings and Profiles of the Job Elements of the Position Analysis Questionnaire (PAQ). West Lafayette, Indiana: Occupational Research Center, Department of Psychological Sciences, Purdue University, 1972.

Mecham, R. C., McCormick, E. J., & Jeanneret, P. R. Users Manual for the Position Analysis Questionnaire (PAQ) System II. Logan, Utah: PAQ Services, Inc., 1977.

McCormick, E. J. Job Analysis: Methods and Applications. New York: AMACOM, 1979.

McCormick, E. J., Jeanneret, P. R., & Mecham, R. C. A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). Journal of Applied Psychology Monograph, 56, 1972.

Moran, L., Kimble, J., & Mefferd, R. Repetitive psychometric measures: equating alternate forms. Psychological Reports, 1964, 14, 335-338.

Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill Book Co., 1967.

Pawlik, K. Comcepts on calculations in human cognitive abilities. In R. G. Cattell, Ed. Handbook of Multivariate Experimental Psychology. Chicago: Rand McNally, 1966.

Shannon, R. H. Development of Alternate Forms Within a Generic Battery of Tests. (Research Report No. NBDL-82R005). New Orleans: Naval Biodynamics Laboratory, 1982a.

Shannon, R. H. Translation of Functional Tasks of a Naval Workstation to Structured Job Elements. (Research Report No. NBDL-82R007). New Orleans: Naval Biodynamics Laboratory, 1982b.

Steiger, J. H. Tests for comparing elements of a correlational matrix. Psychological Bulletin, 1980, 87, 245-251.

Theologus, G., Romashko, T., & Fleishman, E. Development of a Taxonomy of Human Performances: A Feasibility Study of Ability Dimensions for Classifying Human Tasks. American Institutes for Research, Washington, D. C., 1970.

Thomas, D. J., Majewski, P. L., Ewing, C. L., & Gilbert, N. S. Medical qualifications for hazardous-duty aeromedical research. AGARD Conference Proceedings No. 231, Nuilly-Sur-Seine, France: AGARD, 1978, A3, 1-13, 1982.

Winer, B. J. Statistical Principles in Experimental Design. New York: Mc Graw-Hill, 1971.

APPENDIX A

Significant Dimensions and Critical

Elements for the PAQ Analysis of the Phantoms Five

A. <u>Information Input Dimensions and Elements</u>[*]:

   1.   <u>Interpreting What Is Sensed (1.57)</u>[**]

       5.   Visual displays [***]   (5.0)

      13.   Events on circumstances   (5.0)

      23.   Color perception   (3.0)

      29.   Estimating speed of moving objects (3.5)

   2.   <u>Using Various Sources of Information</u> (2.80)

       3.   Pictorial materials   (5.0)

      20.   Near-visual differentiation   (3.0)

      35.   Estimating time   (2.5)

   5.   <u>Being Aware of Environmental Conditions</u> (2.34)

      11.   Man-made features of environment   (4.0)

      13.   Events or circumstances   (5.0)

      23.   Color perception   (3.0)

      29.   Estimating speed of moving objects (3.5)

      34.   Estimating size   (2.5)

B. <u>Mental Processes Dimensions and Elements:</u>

   7. <u>Making Decisions</u> (1.63)

      36.   Decision making   (2.5)

      37.   Reasoning in problem solving   (2.5)

      39.   Combining information   (2.5)

      40.   Analyzing information or data   (2.5)

   8. <u>Processing Information</u> (1.49)

      37.   Reasoning in problem solving   (2.5)

      39.   Combining information   (2.5)

      40.   Analyzing information or data   (2.5)

F. Other Job Characteristics Dimensions and Elements:

   30. Working Under Job-Demanding Circumstances (2.27)

      173. Time pressure of situation   (3.0)

      174. Precision   (2.5)

      175. Attention to detail   (3.0)

      176. Recognition   (4.0)

   31. Performing Structured Work (2.52)

      169. Specified work pace   (4.0)

      170. Repeated activities   (3.0)

   32. Being Alert to Changing Conditions (1.62)

      176. Recognition   (4.0)

      178. Vigiliance: continually changing

           events   (5.0)

---

* Divisions listed in PAQ sequential order from A - F

** Significant dimensions listed in PAQ sequential order from 1-32 with

t-scores in parentheses ($\underline{P} \leqslant$ .1, one tail)

*** Critical elements listed in PAQ sequential order from 1-194 with average

mean in parentheses (Critical rating = 2.5 and above)

APPENDIX B

Significant Attributes and Attribute Dimensions for the PAQ

Analysis of the Phantoms Five Using the 20 Critical Elements in Appendix A

1. <u>Visual Perception/Interpretation Dimension</u> *

    12.   Visual form perception   (4.40)**

    13.   Perceptual speed   (4.78)

    14.   Closure   (4.73)

    16.   Spatial visualization   (4.35)

    17.   Near visual acuity   (3.26)

    18.   Far visual acuity   (3.26)

    19.   Depth perception   (2.88)

    20.   Color discrimination   (5.25)

    41.   Mechanical ability   (3.64)

    45.   Spatial orientation   (2.88)

2. <u>Cognitive Skills</u> Dimension

    14.   Closure   (4.73)

    47.   Time sharing   (2.88)

3. <u>Quantative Skills</u> Dimension

    4.   Numerical computation   (3.03)

---

*   Attribute dimensions listed in order of importance

** Significant attributes listed in sequential order from 1-49 with t-score

    in parentheses ($\underline{P} \leqslant$ .005, one tail)

TE

MED

8