

# Using Computerized Tests to Measure New Dimensions of Abilities: An Exploratory Study

**Charles H. Cory, Bernard Rimland**  
**Navy Personnel Research and Development Center**  
**Rebecca A. Bryson**  
**San Diego State University**

A battery of Graphic Information Processing Tests (GRIP) was developed to utilize the display characteristics of computer terminals in measuring abilities important for processing visually presented information. The GRIP battery was especially intended to assess five "real world" personal attributes which have been difficult to measure with paper-and-pencil tests. The experimental tests were administered to 385 Navy enlisted men and evaluated in conjunction with paper-and-pencil tests of the same attributes as well as with operational cognitive tests and biographical variables. The GRIP tests were found to be useful for measuring short-term memory and sequential reasoning abilities.

Whereas most of the research with computer-assisted test administration has been concerned with tailoring item difficulties to test takers (see Weiss and Betz, 1973, or Wood, 1974, for reviews), what appear to be important characteristics of computerized equipment for expanding dimensionality of measurement appear to have been largely ignored. Since paper-and-pencil tests are limited in terms of stimulus control and response mode, the nearly exclusive reliance on them for personnel selection has imposed restrictions on the types of abilities which can be measured. For example, using conventional

paper-and-pencil tests, it is difficult, if not impossible, to present a moving stimulus, obtain measures of tracking performance, control item exposure time, record response latencies, or sequence items as a function of prior responses. Computer terminals of the type ordinarily used for programmed instruction do have these capacities.

The battery of tests developed for the present research has been especially designed to exploit the special capabilities of computer terminals for pictorial display and movement, and has thus been designated the Graphic Information Processing (GRIP) series. A major interest of the research was in finding abilities important for on-job performance which computerized tests could measure accurately but paper-and-pencil tests could not.

As a starting point for the investigation, five traits of "real world" significance, as defined by Mecham and McCormick (1969), were selected. They were Short Term Memory, Perceptual Speed, Perceptual Closure, Movement Detection, and Dealing with Concepts/Information. Empirical data on the relative importance of these attributes for work performance are available from Mecham and McCormick (1969). The study was designed to provide comparisons of computerized and paper-and-pencil tests in-

---

*APPLIED PSYCHOLOGICAL MEASUREMENT*  
*Vol. 1, No. 1 Winter 1977 pp. 101-110*  
© Copyright 1977 West Publishing Co.

tended to measure these attributes, and to compare the computerized measures and the operational variables in terms of dimensionality and validity for job performance criteria.

The equipment used for computerized test administration consisted of an IBM 1500 system, a cathode ray tube (CRT) display unit, and a screen for film presentation linked on-line to an IBM 1130 computer. Subjects responded to visual stimuli presented on the CRT by touching a target with a light pen, or by entering a response into the typewriter keyboard. Programming was in the Coursewriter language.

### The GRIP Tests

The GRIP battery consisted of eight computer-administered tests, each designed to measure a major aspect of one of the five traits.

Illustrative items from each of the GRIP tests are shown in Figure 1.

1. *Memory for Objects.* Frames showing line drawings of common objects with simple one word names were flashed on the screen at an average exposure time of about one-half second per object per frame. Number of objects per frame ranged from three to nine. After the exposure period, subjects typed in the names of all of the objects remembered.

2. *Memory for Words.* The test was identical in intention and arrangement to Memory for Objects, but with words substituted for pictures. The object of this test was to compare the recall of given words with the recall of words generated by the candidates' recognition and labeling processes. Words were of two lengths: 3 letters and 5 letters. There were six frames for each word length.

3. *Visual Memory for Numbers Test.* This was a digit-span test using the same type of methodology as was used for the two preceding tests but having digits as stimuli. About 50 percent of the digits were presented sequentially and the other 50 percent were presented all at once, as a single stimulus. Scores for each of the first three tests consisted of the number of correct recalls.

4. *Comparing Figures.* In this test the frames contained sets of squares or circles presented as rows, vertical columns, and right and left slanted columns. Three to six stimulus pairs were shown on the screen at a time. Each stimulus had a crossbar, oriented either vertically or horizontally. Subjects were asked to record as true-false answers whether or not all crossbars of corresponding pairs in a set had the same orientations. Separate scores, consisting of total rights corrected for guessing, were computed for the true-false answers in the machine-paced and the self-paced sections.

5. *Recognizing Objects.* This was a test in which partially blotted-out pictures of common objects were presented. The first presentation showed 10 percent of the area, and more area was added in random increments of 10 per unit until 90 percent of the picture was exposed. Subjects entered the names of the stimuli on the keyboard. The score was the total number of frames shown before the objects were identified. Thus, the lower the score, the better the performance.

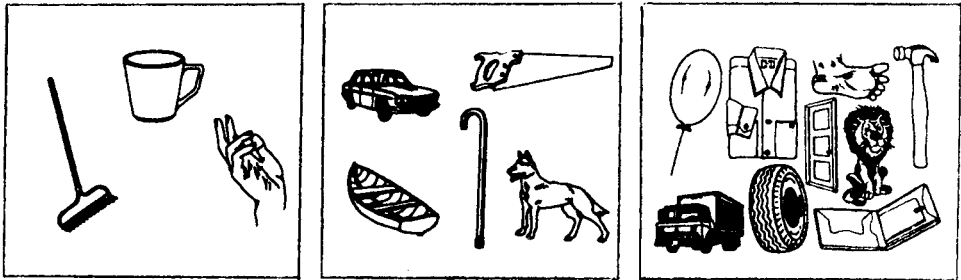
6. *Memory for Patterns.* In this test, patterns were formed by sequentially blinking dots. For some items subjects were asked to report whether or not two consecutive patterns were identical; for other items they were asked to reproduce given patterns on the CRT with a light pen. Two accuracy scores were computed: total rights corrected for guessing for the true-false items, and the total number of dots placed correctly for the free response items.

7. *Twelve Questions.* This test resembled the Twenty Questions game in that subjects were asked to guess the name of an object based on yes-no answers to questions supplied by the computer. It differs from Twenty Questions in that the questions were supplied in the test rather than being posed by the subject. The subject's objectives were to select those questions providing the quickest identification of the object and to avoid questions which are redundant or useless. Scores were sums of correct responses weighted by number and characteristics of the clues received.

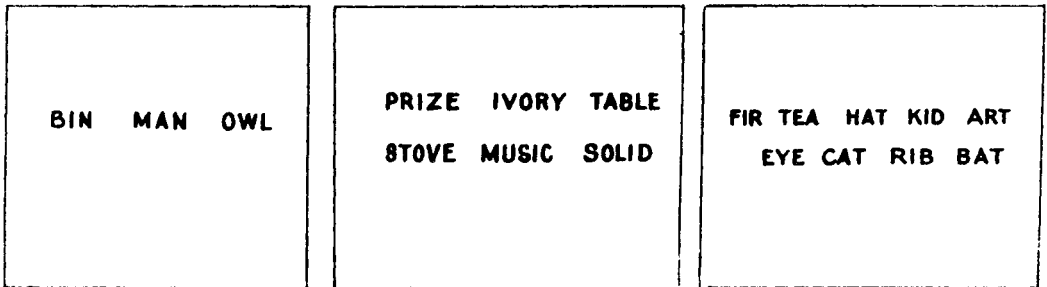
Figure 1

ILLUSTRATIVE ITEMS FROM THE EIGHT COMPUTERIZED TESTS

1. MEMORY FOR OBJECTS



2. MEMORY FOR WORDS

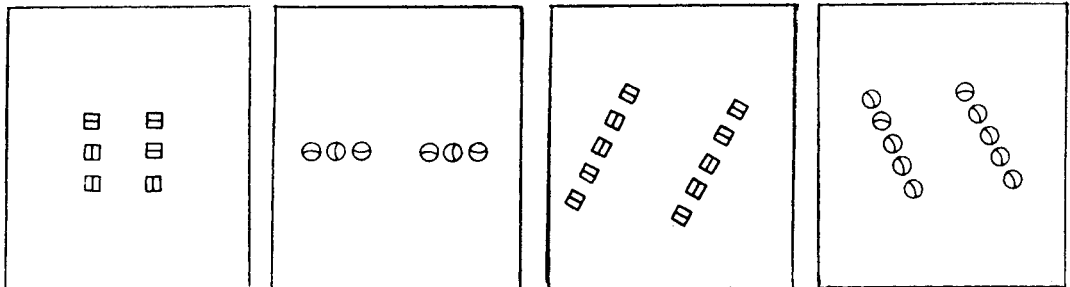


3. VISUAL MEMORY FOR NUMBERS TEST

2 5 1 6\*

124956387\*

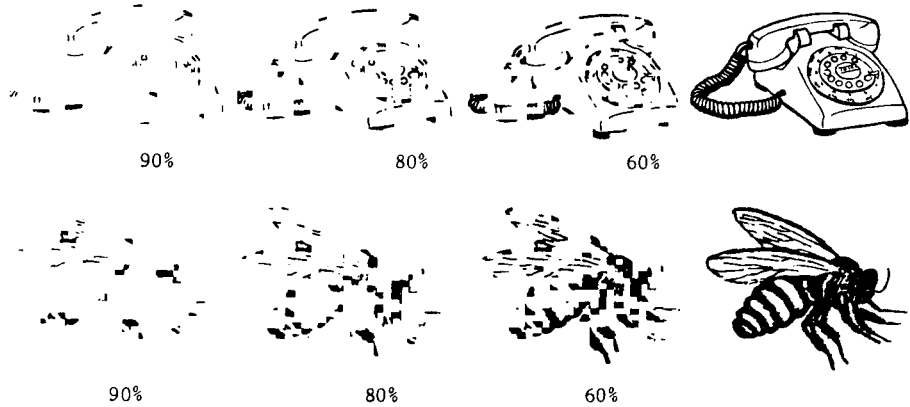
4. COMPARING FIGURES



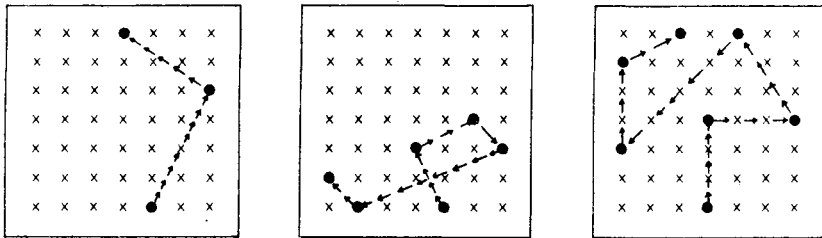
[Continued next page]

[Figure 1, cont'd]

5. RECOGNIZING OBJECTS



6. MEMORY FOR PATTERNS



7. COMPUTERIZED 12 QUESTIONS

Mineral  
Frequently larger than a glove

- |   |  |
|---|--|
| 1. Is it often used as clothing?        | 11. Is it sometimes used by magicians?         |
| 2. Is it made of a soft material?       | 12. Do men and women use it equally often?     |
| 3. Is it often used at meals?           | 13. Is it often used before a person goes out? |
| 4. Do people often wear it?             | 14. Can one use it with his eyes closed?       |
| 5. Does it have moving parts?           | 15. Must one touch it to use it?               |
| 6. Does it have a hard surface?         |  |
| 7. Is it always found on an auto?       | 16. Does it appear dark in the light?          |
| 8. Is it made at least partly of glass? | 17. Can it be used to send messages?           |
| 9. Does it have more than one use?      | 18. Can it improve one's appearance?           |
| 10. Does it use electricity?            |  |

(Mirror)

8. COMPUTERIZED PASSWORD

Metal Finger	Circle	Shiny	Wedding	(Ring)
Soaring Emblem	Feathers	Large	Bald	(Eagle)

8. *Password*. This test resembles the regular "Password" game, in that sets of words which suggest a target word were shown on the CRT. Five separate words were shown as clues. After the first two clues and each succeeding one, the name of the object may be typed on the keyboard. Scores were sums of correct responses weighted by number of clues received.

**Latency and Accuracy Variables**

In addition to direct measures of abilities, latency measures were computed for speed of response for the Memory for Words and the Comparing Figures tests and latency of Recognizing Objects responses (speed of closure). Also, a measure of the total extent to which the response patterns failed to duplicate the stimuli in the free response mode in the Memory for Patterns test was used (PAT-ERR).

**Paper-and-Pencil Experimental Tests, Biographical Variables, and Operational Tests**

Together with the GRIP battery, eight paper-and-pencil tests largely drawn from the ETS Kit of Reference Tests of Cognitive Factors (French et al., 1963), and a motion picture test (Drift Direction by Gibson, 1947) composed the set of experimental tests. In addition, data were obtained for two biographical variables and for the nine tests which are routinely administered and used for Navy personnel decisions. These included the Armed Forces Qualification Test (a test of general mental ability), and tests of vocabulary and verbal reasoning, arithmetic reasoning, perceptual speed, mechanical principles, tool knowledge, electrical knowledge, and memory for pitches and sound patterns.

These variables are described in more detail in Table 1.

In Table 1, the last eight variables—tests in the Navy enlisted classification battery—were scored as Navy Standard Scores and have means of 50 and standard deviations of 10 for an unbiased Navy population. Scores for the Armed Forces Qualification Test are recorded as percentiles and normally range from 10 to 99.

**Sample**

The experimental battery was administered to students at the Naval Training Center, San Diego, during May and June 1972. The sample consisted of 385 male enlisted personnel at or above the 50th percentile of the ability distribution of recruits. Ages ranged from 17 to 19. Personnel selected were either (a) in the first two weeks of specialized training leading to the equivalent of apprenticeship status in one of three specific occupations; or (b) had qualified for these but were still in their final week of recruit training.<sup>1</sup> The subsamples are listed below:

A-School students in first two weeks of training:	
Personnelman	82
Electrician's Mate	107
Sonar Technician	58
School qualified recruits	138
Total	385

**Results**

Submatrices of the intercorrelations of the variables intended to be measures of the same ability are shown in Table 2. Computerized tests did not correlate highly with paper-and-pencil tests measuring the same attribute. Intercorrelations for tests measuring Movement Detection were considerably lower than those for tests measuring the other abilities.

Rotated loadings for the seven principal components derived from 26 operational and experimental variables<sup>2</sup> are shown in Table 3. These factors accounted for 56.5 percent of the total variance of the variables. The two factors (3 and 6) primarily defined by computerized tests

<sup>1</sup>The tests were subsequently validated against supervisory evaluations of on-job performance after seven to nine months of duty in the fleet. These data are presently being analyzed and a report on them is in preparation.

<sup>2</sup>To preclude duplicate coverages from contributing undue weight to the factor-structure, the Auditory Memory for Numbers; Comparing Figures, self-paced; and Memory for Patterns, T-F tests were eliminated from the final factoring.

Table 1  
Description of Non-Computerized Experimental and Operational Variables

Variable	Acronym	Content Area
<b>Experimental Tests</b>		
Memory for Numbers		Short term recall of groups of digits.
Tests from the Kit of Reference		
Tests for Cognitive Factors		
Object Number		Short term associative memory.
Gestalt Completion		Perceptual closure in the recognition of objects from fragmentary details.
Concealed Words		Perceptual closure in the recognition of words from fragmentary details.
Hidden Patterns		Inspection of complex patterns to recognize simple target patterns.
Nonsense Syllogisms		Syllogistic reasoning using nonsense words.
Inference		Evaluation of the correctness of inference from given statements.
Drift Direction		Detection of the direction of minimal movements.
Counting Numbers		Scanning rows of digits to identify specified numbers and count their frequencies.
<b>Operational Variables</b>		
Year of Birth	YRBI	
Years of Education	YRED	
Armed Forces Qualification Test	AFQT	Vocabulary, arithmetic reasoning, spatial reasoning, and mechanical knowledge.
General Classification Test	GCT	Word meanings and the ability to reason verbally.
Arithmetic Reasoning Test	ARI	Quantitative aptitude, including mathematical reasoning and problem solving.
Mechanical Test	MECH	Basic mechanical and electrical knowledge and mechanical principles.
Clerical Test	CLER	Perceptual speed and accuracy.
Sonar Pitch Memory Test	SONR	Ability to perceive and remember small differences in tonal pitches.
Radio Code Aptitude Test	RADO	Ability to learn, remember and use sound patterns as symbols.
Electronic Technician Selection Test	ETST	Mathematics, science, electricity and electronics.
Shop Practices	SHOP	Knowledge of tools and shop equipment.

Table 2  
Intercorrelations and Distribution Statistics for  
Tests Measuring Five Cognitive Abilities

	2	3	4	5	Mean	S.D.	N
<b>Short Term Memory</b>							
1. Object Number	09	10	16	16	12.48	6.66	370
2. Memory for Objects		50	21	13	38.22	8.96	373
3. Memory for Words			40	32	32.09	8.79	373
4. Memory for Numbers (V)				52	82.46	17.38	373
5. Memory for Numbers (A)					117.45	14.65	347
<b>Perceptual Speed</b>							
1. Counting Numbers	27	11	48		30.00	6.25	370
2. Comparing Figures, Machine-paced (M)		33	37		34.02	8.39	373
3. Comparing Figures, Self-paced (S)			12		13.92	2.54	373
4. CLER					54.88	8.60	319
<b>Closure</b>							
1. Gestalt Completion		29	32	-24	16.42	3.31	360
2. Concealed Words			29	-11	23.39	5.64	369
3. Hidden Patterns				-14	63.91	20.05	361
4. Recognizing Objects					15.95	4.13	373
<b>Movement Detection</b>							
1. Drift Direction		11	13		13.13	7.41	368
2. Memory for Patterns, True-False			24		5.13	3.58	367
3. Memory for Patterns, Free Response					16.53	8.03	367
<b>Dealing with Concepts/Information</b>							
1. Nonsense Syllogisms	34	17	12	30	5.43	6.84	366
2. Inference		28	31	58	11.28	4.38	369
3. Twelve Questions			33	32	23.70	28.00	336
4. Password				40	22.24	5.72	358
5. GCT					60.78	6.20	319

Note: Decimal points omitted from correlation coefficients. Correlations of .12 or greater are significant at  $p \leq .05$ ; correlations of .16 or greater are significant at  $p \leq .01$ .

represented 12 percent of the total and 21 percent of the common variance. In addition, computerized tests loaded heavily on two other factors (2 and 7) which accounted for 13 percent of the total and 24 percent of the common variance. Loadings which round to .30 or greater were interpreted.

*Factor 1 ("g").* With AFQT, GCT, ARI, Nonsense Syllogisms, and Inference tests loading heavily on this factor, it appears to represent the broad-based cognitive ability originally identified by Spearman as "g."

*Factor 2 (Perceptual Speed).* The heavy loadings of CLER and Counting Numbers and the absence of significant loadings for these tests on other factors indicate that Factor 2 is a measure of Perceptual Speed. Comparing Figures loaded significantly but less heavily on this factor. Speed of performance, which is important for ARI and Hidden Patterns, may be the reason these tests load significantly on Factor 2. For this sample, age was associated with Perceptual Speed, a finding also reported in previous research (Cory, 1971).



Table 3  
Rotated Factor Loadings of Twenty-Six Operational  
and Experimental Variables  
(N=373)

Variable	Factor							h <sup>2</sup>
	1	2	3	4	5	6	7	
<b>Operational Variables</b>								
AFQT	<u>56</u>	01	05	<u>47</u>	<u>31</u>	-02	<u>30</u>	73
GCT	<u>66</u>	02	24	<u>30</u>	<u>09</u>	<u>32</u>	<u>06</u>	70
ARI	<u>71</u>	<u>30</u>	03	<u>00</u>	-02	-10	25	67
MECH	<u>19</u>	-09	04	<u>79</u>	21	-01	13	73
CLER	17	<u>74</u>	-01	-15	15	-07	-12	64
SONAR	03	-06	08	23	11	13	<u>63</u>	48
RADIO	20	<u>35</u>	16	-01	<u>03</u>	-17	<u>60</u>	58
ETST	<u>58</u>	<u>14</u>	27	<u>44</u>	-02	-04	<u>11</u>	63
SHOP	<u>11</u>	-04	-08	<u>83</u>	-09	18	05	76
Year of Birth	-00	-46	<u>38</u>	-16	03	-36	-06	52
<b>Paper-and Pencil Tests</b>								
Hidden Patterns	18	<u>43</u>	00	14	<u>47</u>	14	21	52
Counting Numbers	02	<u>74</u>	05	-07	<u>03</u>	-02	05	56
Gestalt Completion	06	<u>00</u>	15	10	<u>77</u>	17	-06	66
Concealed Words	11	25	03	-20	<u>65</u>	-13	05	56
Object Number	<u>33</u>	-02	-21	-46	<u>11</u>	22	21	47
Nonsense Syllogisms	<u>55</u>	00	-04	-05	16	15	-04	36
Inference	<u>70</u>	14	12	04	-02	<u>33</u>	-03	63
Drift Direction	<u>01</u>	06	03	11	<u>36</u>	-03	<u>31</u>	24
<b>Computer Administered Tests</b>								
Comparing Figures	07	<u>54</u>	11	04	15	14	19	40
Twelve Questions	20	-01	07	06	01	<u>66</u>	10	50
Memory for Words	14	03	<u>56</u>	-13	06	<u>26</u>	<u>38</u>	57
Visual Memory for Numbers	06	17	13	-12	-05	<u>40</u>	<u>57</u>	56
Memory for Patterns	15	<u>39</u>	<u>54</u>	16	-11	<u>07</u>	<u>19</u>	54
Password	17	<u>07</u>	<u>30</u>	02	09	<u>62</u>	04	52
Memory for Objects	01	-10	<u>61</u>	-13	15	09	<u>34</u>	55
Recognizing Objects	-01	-13	-58	-23	-25	-26	21	58
Eigenvalue	5.33	2.41	1.81	1.48	1.34	1.23	1.08	
Percentage of Variance	20.50	9.28	6.97	5.68	5.18	4.72	4.17	
(Total Percentage = 56.50)								

Note: Decimal points are omitted from the factor loadings and the commonalities. Factor loadings  $\geq |.30|$  are underlined. Small percentages of missing data are present for some of the Operational Variables.

**Factors 3 and 7 (Short Term Memory).** Both Factor 3 and Factor 7 measure short term memory abilities. However, the abilities are primarily distinguished by their degree of associational content. Thus, the memory tests loading most heavily on Factor 3 require the direct recall of stimuli having high associational values or substantial verbal mediation—i.e., words and the names of objects. On the other hand Factor

7 was primarily defined by stimuli requiring rote reproduction of musical pitches, sound patterns, and numeric digits, with little if any associational content.

**Factor 4 (Technical Knowledge).** This factor, with MECH and SHOP having the highest loadings, primarily represents mechanical reasoning and tool knowledge. Thus, it consists of a crystallized ability having primarily prac-



tical rather than academic content. However, AFQT, ETST, and GCT—tests which require academic crystallized abilities—also loaded significantly on it.

*Factor 5 (Perceptual Closure).* The high loadings of Gestalt Completion and Concealed Words, and the moderate loading of Hidden Patterns on Factor 5 indicate it to be a measure of Perceptual Closure. Drift Direction and AFQT probably loaded significantly on this factor by virtue of their visual elements.

*Factor 6 (Sequential Reasoning).* This factor had moderate loadings on reasoning tests which have been identified as measuring general verbal facility (GCT), memory (Visual Memory for Numbers), and reasoning abilities (Inference). However, its principal loadings on Password and Twelve Questions suggest that Factor 6 primarily involves serial, deductive reasoning requiring frequent updates of the information pool. This unique ability appears not to have been previously identified (French et al., 1963; Guilford, 1967; Messick, 1973). It is somewhat similar to the factor identified as Serial Integration by Siebert and Snow (1965, p.161), who used a motion picture rather than a computer terminal mode of administration.

These factors are no doubt somewhat distorted from those derivable from a full range sample. An obvious distortion is a “g” factor which is underdefined as a result of the low intercorrelations among operational tests, resulting from the restricted nature and the relatively high ability level of the sample. However, the analysis was believed to be useful for establishing the dimensionality of the predictor battery.

### Discussion and Conclusions

It is clear that the experimental battery represents an increase in the breadth of abilities covered beyond those in the operational Navy battery, a considerable amount of which is attributable to the GRIP tests. GRIP tests were the primary sources of two factors which accounted for 12 percent of the total and 21 percent of the common variance and had substan-

tial loadings on factors which accounted for an additional 13 percent of the total and 24 percent of the common variance. Computerized tests apparently provided measures of several attributes which were different from those measured by paper-and-pencil tests.

The unique measurement characteristics of the GRIP tests appear to be as follows:

1. Computer administration of tests of short-term recall using a variety of stimuli is feasible, and appears to offer advantages in ease of data collection and processing over paper-and-pencil tests measuring the same attributes.
2. Computerized administration of perceptual speed tests, as implemented in the GRIP battery, was only marginally different from paper-and-pencil measures of perceptual speed.
3. Recognizing Objects, the GRIP test for perceptual closure, appears to be measuring a substantially different attribute than that measured by the paper-and-pencil tests which have been previously used to define this ability. The computerized measure appears to place greater reliance on short-term memory factors than do paper-and-pencil tests of closure, but the extent and character of the other differentiations between the modes are not clear at the present time. More research on this ability is needed.
4. The two experimental tests designed to measure Movement Detection were not closely related to one another and therefore did not provide evidence of a Movement Detection factor. Instead these tests loaded on short-term memory factors, Perceptual Speed, and Perceptual Closure. Clearly, further research is needed for the development of definitive measures of Movement Detection.
5. Facility in Sequential Reasoning was apparently an ability which was uniquely measurable by computer-administered tests.

6. Although short-term memory tests were the major defining elements in two factors, the results of this study were in general agreement with findings of Christal (1958) and Thurstone (1946), among others, in failing to find clear-cut evidence for memory factors defined by data content. The different memory abilities which emerged corresponded with the high and low associational characteristics of the stimuli.

### References

- Cristal, R.E. Factor analytic study of visual memory. *Psychological Monographs: General and Applied*, 1958, 72(13), 1-24.
- Cory, C.H. *A comparison of the Porteus and Navy Maze Tests*. San Diego: Naval Personnel and Training Research Laboratory, June 1971. (Technical Bulletin STB 71-11)
- French, J.W., Ekstrom, R.B., & Price, L.A. *Manual for Kit of Reference Tests for cognitive factors* (Rev. 1963). Princeton: Educational Testing Service, 1963.
- Gibson, J.J. (Ed.). *Motion picture testing and research: Army Air Forces aviation psychology program, report no. 7*. Washington, D.C.: U.S. Government Printing Office, 1947.
- Guilford, J.P. *The nature of human intelligence*. New York: McGraw-Hill, 1967.
- Mecham, R.C., & McCormick, E.J. *The rated attribute requirements of job elements in the Position Analysis Questionnaire*. Occupational Research Center, Purdue University, January 1969. Report No. 1, AD-682 490. Prepared for Office of Naval Research under Contract Nonr-1100(28).
- Messick, S. *Multivariate models of cognition and personality: The need for both process and structure in psychological theory and measurement*. Princeton: Educational Testing Service, 1973.
- Seibert, W.F., & Snow, R.E. *Studies in cine-psychometry I: Preliminary factor analysis of visual cognition and memory*. Lafayette, Indiana: Purdue University, Audio Visual Center, June 1965. (Grant No. 7-12-0280-184, U.S. Office of Education, Department of Health, Education and Welfare.)
- Thurstone, L.L. Theories of intelligence. *Science Monographs*, 1946, 62, 101-112.
- Weiss, D.J., & Betz, Nancy E. *Ability measurement: Conventional or adaptive?* Psychometric Methods Program, University of Minnesota, February 1973. Report No. 73-1. Prepared for Office of Naval Research under Contract NR No. 150-343.
- Wood, R. Response-contingent testing. *Review of Educational Research*, 1974, 43(4), 529-544.

### Acknowledgements

*This research was supported by the Personnel and Training Research Programs of the Office of Naval Research under Project Order Number RR042-04 NR 150-335. The opinions expressed are those of the authors and do not necessarily reflect those of the Department of the Navy.*

### Author's Address

Charles H. Cory, Attn: Code 310, U.S. Navy Personnel Research and Development Center, San Diego, CA 92152.