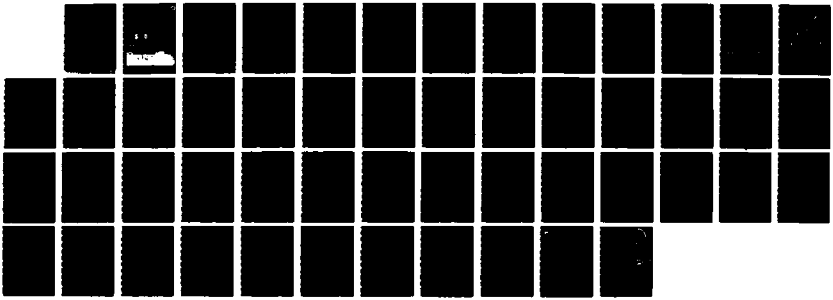
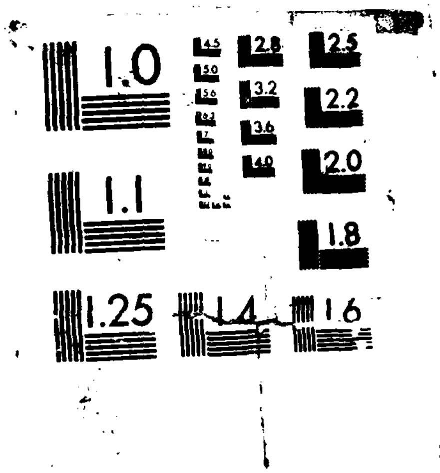
AD-A183 971 COMPUTER-CONTROLLED TESTING OF VISUAL-SPATIAL ABILITY 1/1
(U) WASHINGTON UNIV SEATTLE DEPT OF PSYCHOLOGY
E HUNT ET AL AUG 87 91-6001-537 NPRDC-TR-87-31
UNCLASSIFIED N66001-85-C-0017 F/G 5/9 NL





MICROCOPY RESOLUTION TEST CHART

Navy Personnel Research and Development Center

San Diego, CA 92152-6800 TR 87-31 August 1987

Computer-Controlled Testing of Visual-Spatial Ability

Because many jobs require the ability to reason about moving objects, these new tests hold promise for improving personnel classification.

NPRDC TR-87-31

August 1987

Computer-Controlled Testing of Visual-Spatial Ability

Earl Hunt, Ph.D.
University of Washington

James W. Pellegrino, Ph.D.
Ronald Abate
David L. Alderton, Ph.D.
University of California at Santa Barbara

Simon A. Farr
Robert W. Frick
University of Washington

Thomas P. McDonald
University of California at Santa Barbara

Reviewed and approved by
John J. Pass, Ph.D.
Director, Personnel Systems Department

Released by
B. E. Bacon
Captain, U.S. Navy
Commanding Officer
and
J.S. McMichael, Ph.D.
Technical Director

Approved for public release;
distribution unlimited.

Navy Personnel Research and Development Center
San Diego, CA 92152-6800

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AD-A183971

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S): 91-6001-537	5 MONITORING ORGANIZATION REPORT NUMBER(S): TR 87-31
6a NAME OF PERFORMING ORGANIZATION: Psychology Department, University of Washington; 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION
6c ADDRESS (City, State, and ZIP Code): Seattle, WA 98195	7b ADDRESS (City, State, and ZIP Code)
8a NAME OF FUNDING/SPONSORING: Navy Personnel Research and Development Center; 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code): San Diego, CA 92152-6800	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO 62763; PROJECT NO F63521; TASK RF63521-804; WORK UNIT 521-804-040

11 TITLE (Include Security Classification)
Computer-Controlled Testing of Visual-Spatial Ability

12 PERSONAL AUTHOR(S)
E. Hunt, J. W. Pellegrino, R. Abate, D. L. Alderton, S. A. Farr, R. W. Frick, and T. P. McDonald

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM 85 Jan TO 86 Jan	1987 August	48

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
05	09		Testing, spatial ability, ASVAB

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Five "static" tasks were constructed involving stationary displays and six "dynamic" tasks were constructed involving moving displays. These tasks, together with conventional paper-and-pencil tests of spatial-visual ability, were given to 170 subjects. The results of a multi-dimensional analysis suggested that several spatial-visual abilities that are not measured in the conventional paper and pencil format might be measured by computer-controlled testing. These include separate measures for speed and accuracy and a measure of an ability to reason about moving objects.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL: David L. Alderton, Ph.D.	22b TELEPHONE (Include Area Code): (619) 225-2191; 22c OFFICE SYMBOL: Code 622

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

FOREWORD

This report describes an investigation of computer-based tests of spatial aptitude conducted for the Computerized Testing Technologies (CTT) project (Work Unit No. 620521.040.03.06; Job Order No. 6822163GAD3). The objective of the CTT project is to develop and apply information processing methods and models to aptitude constructs and to evaluate their potential for computerized testing. Other CTT projects have investigated paragraph and mechanical comprehension.

The current research was contracted to Dr. Earl Hunt at the University of Washington (Contract No. N66001-85-C-0017), and part of it was subcontracted to Dr. James Pellegrino of the University of California, Santa Barbara. The work was accomplished during a 1-year period beginning 1 January 1985. The purpose of this work was to develop computer-based tests of spatial-visual ability to be used in further research by the Navy as possible classification tests. It is believed that spatial ability tests, which are not currently represented among military job assignment tests, may improve the assignment of enlisted personnel to selected technical ratings.

This work describes the development and evaluation of the test battery on a college population and will serve as the basis for further work that will be conducted at the Navy Personnel Research and Development Center on Navy enlisted personnel. Several of the contract-developed spatial ability tests are currently being administered to Navy machinist mates under the Tri-Services Performance-Based Personnel Classification Project, and there are plans to experimentally append several of these tasks in a field test of the computerized adaptive version of the Armed Services Vocational Aptitude Battery (CAT-ASVAB) in FY88.

B. E. BACON
Captain, U.S. Navy
Commanding Officer

J. McMICHAEL
Technical Director

SUMMARY

Problem

Identifying people who have high spatial-visual ability would facilitate the assignment of individuals to occupations where success depends on those skills. The major facilitation would be expected for jobs requiring machinery operations and/or the reading of analog displays and diagrams. Traditionally, spatial-visual ability has been tested by asking people to reason about pictures presented in a conventional paper-and-pencil format. The advent of computer-controlled testing makes it possible to make much finer measures of how people reason about a visual scene, and to measure reasoning about absolute and relative motion.

Purpose

The purpose of this research is to (1) develop tests of spatial-visual reasoning that take advantage of computer technology, (2) determine if these tests measure any dimensions of spatial-visual ability not measured by current tests, and (3) provide these new tests to the Navy for further investigation as tools in personnel classification.

Approach

Eleven computer-administered tasks requiring spatial-visual ability were developed. Six of these took advantage of the computer's ability to present moving objects. Five took advantage of the computer's ability to measure reaction time for individual problems. These tasks and eight conventional paper-and-pencil tests were given to 170 college students. Scores were correlated, and multivariate factor analyses were conducted.

Results and Discussion

The results indicate two advantages for computerized test administration. First, computer capabilities allow the development of tests for previously unmeasureable human abilities. The data strongly indicate that the ability to deal with objects in motion is separate from the ability to deal with the stationary visual displays used on conventional tests. Because many jobs require the ability to reason about moving objects, these new tests hold promise for improving personnel classification.

Computer based tests also allow the separate measurement of the speed and accuracy of answering test items; conventional tests combine these factors. Prior research has shown that there is a substantial difference between the speed of problem solving and the accuracy of responding. Separating these two dimensions may improve personnel selection and classification.

In summary, the results show that computerized tests of spatial-visual ability have advantages over conventional tests and have potential for improving the prediction of job performance.

Recommendations

The battery should be used in studies of Navy personnel to determine if the abilities measured by these tests predict performance in Navy jobs.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Problem

Spatial-visual ability is the ability to reason about visual displays. This ability is useful in a variety of mechanical tasks and in machinery operation tasks. Identifying people who have high spatial-visual ability is important in personnel classification, as it facilitates the assignment of individuals to occupations compatible with their skills. Traditionally, spatial-visual ability has been tested by asking people to reason about pictures presented in a conventional paper-and-pencil test. The advent of computer-controlled testing makes it possible to extend testing of spatial-visual ability to provide finer measures of how people reason about a visual scene and to examine people's ability to reason about absolute and relative motion. The development of finer measures of these abilities would be a first step toward improved personnel classification and would augment the highly verbal orientation of present paper-and-pencil tests.

Purpose

This study had three goals. The first was to determine whether computer-administered spatial-visual tests involving static displays evaluate the same abilities as paper-and-pencil spatial-visual tests. The second was to determine whether or not tests involving dynamic displays evaluate a new dimension of spatial-visual ability. As subsidiary goals related to these questions, we examined the use of within-problem reaction time measures that can be obtained in computerized testing but cannot be obtained in paper-and-pencil testing. Finally, a computer-administered test battery containing both static and dynamic tasks was to be constructed for subsequent research on the prediction of job performance.

Background

Spatial-visual ability is the ability to reason about visual scenes. Examples of this ability are ubiquitous, ranging from the performance of pedestrians deciding to cross busy streets to the performance of jigsaw puzzle addicts as they piece together a picture.

Virtually every major theory of intelligence acknowledges the existence of spatial-visual ability and distinguishes it from verbal ability and general reasoning ability (Carroll, 1982). Closer examination shows that spatial-visual ability is better thought of as a domain of abilities than as an isolated skill. Factor analytic studies of the domain have identified three different spatial ability factors (Lohman, 1979; McGee, 1979). *Spatial orientation* is the ability to imagine how a stimulus or stimulus array will appear from various perspectives (Guilford & Zimmerman, 1947). *Spatial relations* is the ability to move objects "in the mind's eye," such as when "mentally rotating" an object about its center (Shepard & Cooper, 1982). Conventional psychometric tests of spatial relations include the Primary Mental Abilities Space test (Thurstone & Thurstone, 1949) and a psychometric analog of the laboratory rotation task (Lansman, 1981). *Spatial visualization* is the ability to deal with complex visual problems that require imagining the relative movements of internal parts of a visual image. Solving jigsaw puzzles is a good example. Psychometric tests tapping spatial visualization include the folding task in the Differential Aptitude Test battery (DAT; Bennett, Seashore, & Wesman, 1974) and the Minnesota Paper Form Board test (Likert & Quasha, 1970).

Tests of spatial orientation, spatial relations, and spatial visualization are typically correlated across individuals. Therefore, in a multidimensional analysis, the scores from several spatial-visual tests can often be placed in a two- rather than a three-dimensional space. More precisely, usually three dimensions are required for an excellent fit, but two dimensions will be "almost" sufficient.

On logical grounds alone, one might expect tests of spatial-visual ability to predict performance in nonacademic fields where a person is required to deal with visual objects. Spatial relations and spatial visualization tests are reliable predictors of performance in architecture and engineering courses (McGee, 1979). More detailed studies have shown that spatial ability tests predict performance on problems involving analysis of engineering drawings (Pellegrino, Mumaw, & Shute, 1985). Within the military, a spatial ability test that was formerly (but is not now) included in the Armed Services Vocational Aptitude Battery (ASVAB) was found to correlate with performance in several situations that involve understanding mechanical operations (Navy Personnel Research and Development Center, 1979).

The above remarks apply to conventional paper-and-pencil tests of spatial-visual ability, in which a person is shown a picture and asked to reason about it. Hunt and Pellegrino (1985) pointed out that the conventional paper-and-pencil format restricts testing of spatial-visual ability severely, because the visual scene the examinee must reason about cannot contain moving elements. Also, although it is possible to determine how many items an examinee can pass in a fixed time, it is not possible to determine how long the examinee spends on an individual item or the time spent on various subproblems within an item. This is a serious issue because speed and accuracy in solving different parts of a problem may reflect different psychological skills (Pellegrino & Kail, 1982). More generally, people may make trade-offs between speed and accuracy of performance in different ways, so measures of both speed and accuracy may be needed to assess ability accurately (Pachella, 1974). Hunt and Pellegrino (1985) pointed out that both of these aspects can be measured by computer-administered tests. Visual displays with moving elements (dynamic displays) can be presented in computer-controlled testing. Speed-accuracy relations can be assessed by recording latency and accuracy separately for each item, or even for subparts of an item. In some cases it is possible to avoid the speed-accuracy problem by adaptive testing, in which one finds out the level of difficulty at which an examinee can maintain a fixed level of accuracy.

Hunt and Pellegrino added two cautions. First, one can hypothesize that reasoning about dynamic displays is different from reasoning about static displays, but there is at present no evidence to show that this is the case. Second, while it is possible to design static display problems that appear to be related to (and that are correlated with) performance on paper-and-pencil tests of spatial ability, it is also possible that the very fact of computer-controlled testing itself taps a new dimension of ability, the ability to deal with computer-controlled displays *per se*. Hunt and Pellegrino noted that the evidence concerning the existence of such an ability is sparse and somewhat contradictory.

Should tests of spatial-visual ability be redesigned to take advantage of the flexibility of computer-controlled displays? The answer to this question depends on the answers to three related questions. Does computerized testing involving static displays evaluate the same abilities as paper-and-pencil testing? Can dynamic displays reveal a dimension of ability that is different from the ability evaluated using static displays? Finally, do the additional measures available through computerized testing make possible better prediction of on-the-job performance? Of course, the last question is of most interest in applied psychology. An attempt to answer it directly, however, could be both fruitless and extremely expensive unless the first two questions are examined first.

APPROACH

Subjects

Subjects were recruited through newspaper advertisements directed towards the campus community at the University of Washington (UW) and the University of California, Santa Barbara (UCSB). At UW, 83 subjects were tested; 87 were tested at UCSB. All subjects were at least 18 years old and spoke fluent English.

Apparatus

The computer-administered tasks were performed on Apple II+ or IIe computers. Six of the tasks required a Mountain Hardware clock card, and one task required a joystick.

Each computer controlled a 13 mm X 19 mm green monochrome screen that was 192 pixels high and 280 pixels wide. The screen was refreshed every 33 msecs.

Design

Subjects were given 19 tests. Eight were standard paper-and-pencil tests of spatial-visual ability, verbal ability, or general reasoning. Eleven tests were presented under computer control. Five of these used static tasks, and six used dynamic displays. All computer tests measured some aspect of spatial-visual ability. The tests are described below. Tests were given in a fixed order, which is listed in Table 1. Table 1 also presents the number of items per test and an estimate of the time required for the test. Subjects were tested for 5 consecutive days for a maximum of 2 hours per day.

Table 1

Order and Duration of Testing

Test Sequence	Type	Items	Time (min.)
Monday			
Path Memory	Dynamic	72	20
Arrival Time–1 object	Dynamic	80	10
DAT Space	Paper	60*	25
Mental Rotation	Static	280	30
Tuesday			
Raven's Matrices	Paper	20*	25
Identical Pictures	Paper	96*	5
Integrating Details	Static	48	30
Extrapolation	Dynamic	108	20
Wednesday			
Intercept	Dynamic	72	15
Perceptual Comparisons	Static	60	25
Shape Memory	Paper	32*	20
Vocabulary Test	Paper	100*	15
Thursday			
Adding Detail	Static	60	30
Spatial Orientation	Paper	60*	15
PMA Space	Paper	150*	10
Arrival time–2 objects	Dynamic	250	25
Friday			
3-D Mental Rotation	Paper	96*	10
Arrival time–4 objects	Dynamic	64	10
Surface Development	Static	192	50

Note: Paper = Paper-and-Pencil
*All problems were not necessarily attempted.

Paper-and-Pencil Tests

The paper-and-pencil tests covered the domain of spatial-visual abilities as presently tested. Two analytical tests, a vocabulary test and a general intelligence (*g*) test, were included as markers of abilities outside the spatial-visual domain.

For almost all of the paper and pencil tests, the dependent measure was the number correct. For all of the tests except the DAT Space test, a fraction of the number wrong (the number wrong divided by the number of alternatives) was subtracted from the number correct to compensate for guessing. For the PMA Space test, the score was the number of correct "match" responses minus the number of incorrect "match" responses.

DAT Space

This paper-and-pencil test of spatial-visualization ability is part of the Differential Aptitude Test Battery (Bennett, Seashore, & Wesman, 1974). The subject was shown a flat shape that could be folded into a three-dimensional object. Four possible objects were shown. The subject selected which object the flat shape could be folded into. The flat shape often had shadings on some of the sides, so the subject selected the alternative that had both the correct shape and correct shadings. Figure 1 presents a sample of this type of problem. (In this and all examples from paper-and-pencil tests, we present representative problems that are not actual samples from the tests.)

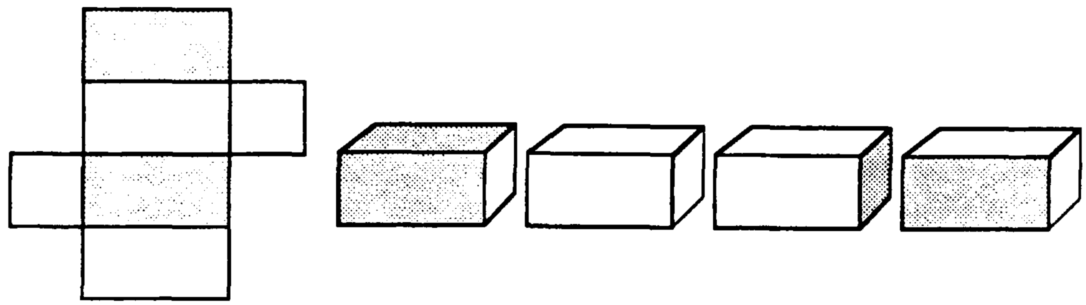


Figure 1. Facsimile problem from the DAT Space test.

Raven's Advanced Progressive Matrices

Raven's Advanced Progressive Matrices (Raven, 1962) is a paper-and-pencil test of general reasoning ability. The examinee was shown a pattern with a piece missing and told to determine which of eight alternatives best completed the pattern. The pattern was a 3 X 3 array with the bottom right element missing. The test had 12 practice problems, followed by 40 test problems presented in order of difficulty. The 20 odd-numbered problems were used. Examinees were allowed 20 minutes (half of the normal time) to complete the test.

Identical Pictures

This is a test of the perceptual speed factor taken from Educational Testing Service's (ETS) Reference Kit of cognitive abilities (Ekstrom, French, & Harman, 1979). The subject saw a target object and selected which of five objects matched the given object. All objects were two-dimensional line drawings, and all alternatives were in the same orientation as the target stimulus. The test was divided into two sections. Each section had 48 problems; 1-½ minutes were allowed. An example is presented in Figure 2.

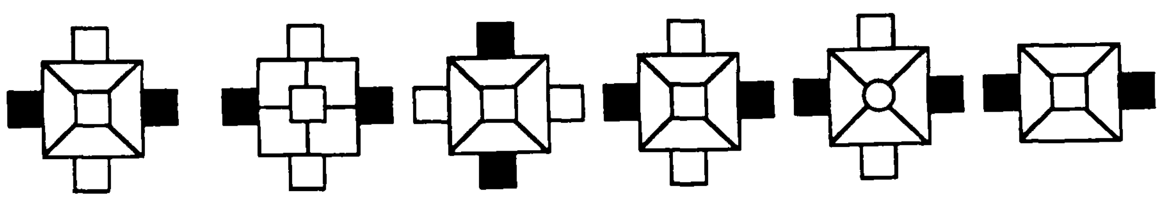


Figure 2. Facsimile problem from the Identical Pictures test.

Shape Memory

This test was also taken from the ETS Reference Kit. The subject viewed a large two-dimensional scene that consisted of a number of black-and-white blobs. After studying the scene, the subject was shown potential subsets from the scene and indicated whether the smaller scene was contained in the larger scene. An example is presented in Figure 3. Two different scenes were tested. The scenes were studied for 4 minutes. Sixteen recognition problems followed each scene.

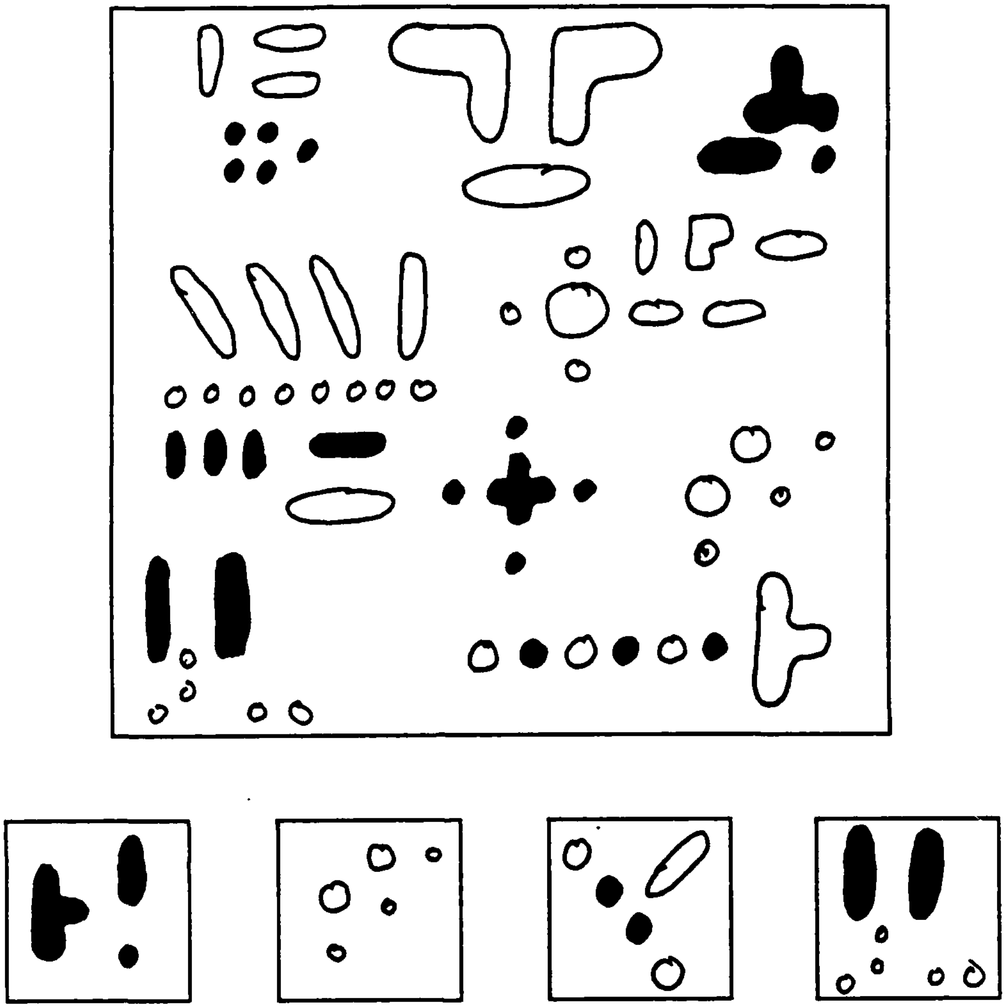


Figure 3. Facsimile problem from the Shape Memory test.

Nelson-Denny Vocabulary

The vocabulary portion of the Nelson-Denny Reading Test (Brown, Bennett & Hanna, 1981) was administered. From a set of five alternatives, the subject selected the word that was the best synonym of the target word.

Spatial Orientation

This test of the spatial orientation factor was taken from the Guilford-Zimmerman Aptitude Survey (Guilford & Zimmerman, 1947). For each problem, the subject viewed two pictures of a shore as seen from a boat, with the prow of the boat in the picture. From the first picture to the second, the prow of the boat might have moved *up* or *down*, the boat might have turned *left* or *right*, and/or the boat might have tilted *left* or *right*. The subject selected which of five alternatives actually occurred.

PMA Space

This test is the spatial relations test from the Primary Mental Abilities Battery (Thurstone, 1965). The subject was shown a target figure, which was a two-dimensional line drawing. Five alternative drawings were also shown. The subject indicated whether or not each alternative was identical to the target figure except for a possible rotation in the plane of the paper. Alternatives that did not match the target figure were mirror images of the target figure.

3-D Mental Rotation

This test was designed by Lansman (1981) to be a paper-and-pencil analog of the mental rotation task developed by Shepard and his colleagues (Shepard & Cooper, 1982). The task was similar to the PMA Space test, except that the objects to be compared were 3-D snake-like strings of cubes. Subjects were told they could consider rotations in all three dimensions (though when two objects were the same they could always be aligned by a rotation in the plane of the paper). The test had four timed sections with 24 problems each. The times for the tests were 2 minutes, 1-½ minutes, 2 minutes, and 2-½ minutes.

Dynamic Computer Tasks

A brief description of each dynamic computer-controlled task follows. Detailed descriptions are presented in Appendix A.

Path Memory

This task tests memory for the paths of moving objects. A trial consisted of a sequential display of three small squares moving across the screen. The subject indicated which path was different. The squares followed a parabolic path, starting at the lower left, and moved upwards. A square completed its path and "disappeared" before the next square appeared. On each trial, either the first and second square followed the same path and the third square followed a different path or the second and third squares followed the same path and the first followed a different path.

Three parameters were used to construct the paths: the starting height of the parabola, the height of the apex of the parabola, and the length of the parabola from the start to the apex. One of these parameters was varied to make one of the paths different from the others.

There were eight levels of difficulty; the easier the item, the larger the difference between the different path and the two same paths. Difficulty was increased or decreased by changing one of the three parameters of the parabola, either decreasing or increasing the difference between the parameter used to construct the two "same" paths and the "different" path. The difficulty was increased when the subject correctly answered 2 trials in a row and decreased when the subject incorrectly answered 1 trial. There were 72 trials. The dependent measure was the average difficulty level of the trials, with the first 24 trials not considered. Three different accuracy measures were calculated, corresponding to the three different ways of changing the parabola.

Arrival Time–One Object

This task was designed to measure the ability to estimate the time of arrival of a moving object at a fixed point. A square moved horizontally from the left side of the screen toward a vertical line on the right. One-quarter to one-half of the way across the screen, the object disappeared from view. The subject pressed a key when he or she thought the object would have crossed the line if it had continued the same course at the same speed. Two performance measures were recorded: *accuracy* (the absolute value of the difference between the correct response and the subject's response), and *bias* (the average difference between the correct response and the subject's response).

Arrival Time–Two Objects

This task introduced a situation involving two moving objects. The subject guessed which of two objects would have arrived at its destination (a vertical or horizontal line) first. Thus, this task measured judgments of relative speed, whereas Arrival Time–One Object measured judgments of absolute speed. The subject saw two different objects, each moving towards its destination at a constant speed. A fifth of the way to the destination, the objects disappeared. The subject then indicated which object would have arrived first. Five variations of this task were used, varying the beginning locations and destinations of the objects. In two variations, the objects moved perpendicularly to each other; in the three others, the objects moved in parallel.

Level of difficulty was established by the time difference between when the two objects would have arrived at their destination. There were eight levels of difficulty. The level of difficulty was increased when the subject answered two trials in a row correctly and was decreased when the subject answered a trial incorrectly. The dependent measure was the average level of difficulty (calculated separately for each variation of the task).

Arrival Time–Four Objects

This task also required judgments of relative speed. The task is something like guessing the winner of a foot race before the race is completed. Four objects moved horizontally from right to left at individual, constant speeds, towards a vertical line on the far left of the screen. The objects started at varying distances from the line. Halfway to the vertical line, the objects disappeared. If the objects had continued traveling at the same speed, three of the objects would have intersected the line at the same time, and one would have arrived earlier. The subject indicated which object would have intersected the line first.

There were eight levels of difficulty, corresponding to the size of the time difference between the arrival times of the first and the other objects. The level of difficulty increased when the subject answered two trials in a row correctly and decreased when the subject answered a trial incorrectly.

Extrapolation

This task measured the ability to extrapolate the location of a trajectory. Note that in a sense, this was a complement of time extrapolation as tested by the Arrival Time–One Object task because in order to intercept a moving object (e.g., catch a pass in football), one must extrapolate both time and point of arrival.

Three types of curves were presented: a straight line, a sine wave, and a parabola. A part of the curve was shown on each trial, starting from the left side of the computer screen and ending somewhere in the center. Figure 4 shows what the screen might look like at this point. The subject used a joystick to move an arrow up or down a vertical line on the right side of the screen, to indicate where the curve would have intersected the line. After the subject responded the curve was displayed. There were 108 trials.

The dependent measure was the difference (in pixel units X 100) between the correct answer and the subject's answer. Three measures were recorded, one for each of the three types of curves. A subject's overall score was the sum of the average score for each curve type divided by the standard deviation of that curve type across subjects.

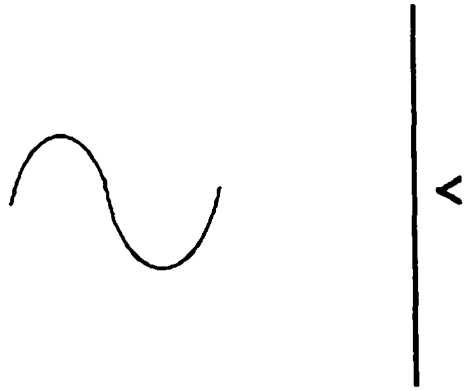


Figure 4. Sample problem from the Extrapolation task.

Intercept

This task measured the ability to combine path and speed extrapolation. The task itself was something like a video game in which a player attempts to "shoot down" a moving object. A small rectangular target moved from left to right at a constant horizontal speed. The target path was either a horizontal straight line, a sine wave, or a parabola. When the subject pressed a key on the keyboard, a triangularly shaped object (called a "missile") began moving upwards at a constant velocity. The subject attempted to launch the missile at the correct time so that it would hit the target. The dependent measure was the vertical distance (in pixel units X 100) between the missile and the target when the target crossed the path of the missile.

Static Computer Tasks

Perceptual Comparisons

The subject saw two objects and decided as quickly as possible whether they were the same or different. The objects were irregularly shaped polygons with from 6 to 14 randomly positioned points, following Cooper (1976). The objects varied in complexity (the number of points on the objects varied from 6 to 14) and degree of difference (number of noncorresponding pairs of points). The dependent measures were reaction time and accuracy.

Mental Rotation

This task was analogous to the two-dimensional paper-and-pencil test (PMA Space), except that speed of rotation could be measured directly. Two objects were presented on the computer screen. The objects were either the same or mirror images of one another. The objects appeared at different rotation angles, where rotation angle was defined by the angle of intersection between the principal axes of the objects.

The rotation angles ranged from 0 to 180 degrees, in 20 degree increments. Reaction time was measured on each trial. Based on previous research, it was expected to be a linear function of the value of the rotation angle. A variety of derived measures (e.g., rate of mental rotation) were obtained from the reaction time measure. Errors were also recorded.

Adding Detail

This task measured the ability to integrate details into an image (Kosslyn, 1980; Poltrock & Brown, 1984). A six-pointed star was presented, with a dot on either the inside or the outside of the star, at either an inner or outer vertex. Each time the subject pressed a key, the preceding dot disappeared and the next dot appeared. When four to seven dots had been presented, the test stimulus appeared instead of another dot. The test stimulus was a star with several dots. The presentation sequence is shown in Figure 5. The subject indicated whether the dots on the test stimulus were in the same positions as the dots that had been presented one by one. The dependent measures were accuracy, latency of viewing each dot, and latency of the response to the test stimulus.

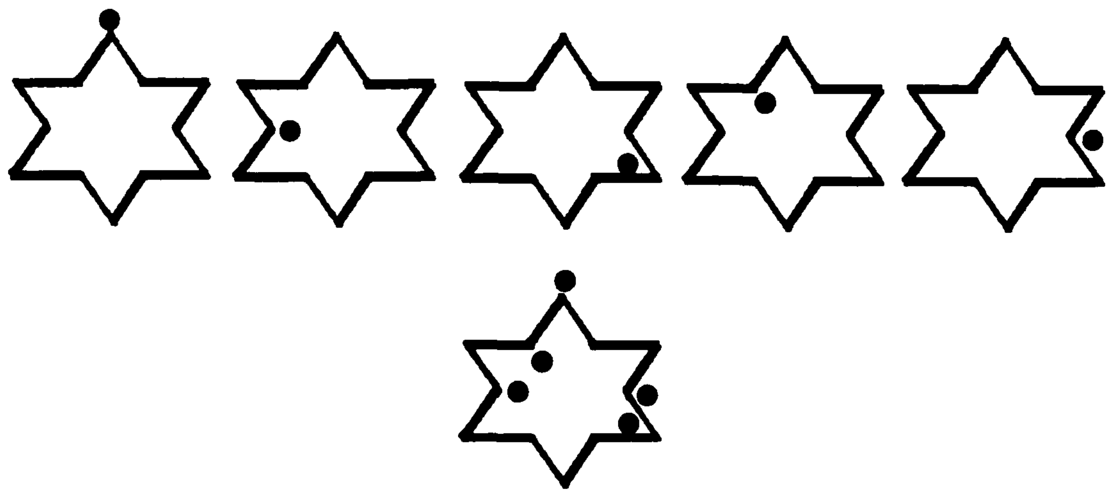


Figure 5. Sample problem from the Adding Detail task.

Integrating Details

This task measured the ability to imagine pieces combining to form a larger object. The subject saw a number of two-dimensional figures. The edges were marked, such that the edge of one piece corresponded to the edge of another. The examinee was supposed to imagine what form would be created if the pieces were joined at the indicated edges. When the examinee had done this, he or she pressed a key, the pieces disappeared, and a whole figure appeared. The subject indicated whether the whole object was the same as the one that would have been made from those parts. The number of pieces to be integrated varied from three to six.

Latency and accuracy scores were recorded. Latencies were recorded separately for the time an examinee spent viewing the separated figure and the time he or she spent deciding whether the whole figure was the correct figure. Previous research had shown that both of these latencies vary with the number of pieces to be integrated.

Surface Development

This task was a computer analog of the DAT Space test previously described. On one side of the screen, a two-dimensional figure that represents an unfolded cube was presented. The base of the cube was labeled, and two or three surfaces of the cube had a dot placed on them, positioned either in the center of the side or off-center. On the other side of the computer screen, a three-dimensional folded cube was presented with an appropriate number of surfaces containing dots. The subject indicated whether the unfolded cube matched the folded cube. The dependent measures were accuracy and response latency. Different folding patterns were used that systematically varied the number of mental folds that needed to be made and the number of surfaces that had to be mentally "carried along" during such folds. Decision time and accuracy were expected to be a monotonic function of the number of folds and surfaces carried along.

RESULTS

The results will be presented in two major sections. The first, "Univariate Analyses," describes the correlations between the various measures taken within each task. Three minor sections within the major section discuss the statistics for the computer-controlled dynamic and static tasks and the paper-and-pencil tasks. The second major section, "Multivariate Analyses," describes analyses of the correlations between tasks. It is divided into a section considering the relation between computer-controlled tasks and paper-and-pencil tests and a section describing the relation between static and dynamic tasks.

Univariate Analyses

Throughout, correlations with an absolute value greater than .16 were statistically significant at p less than .05, using a two-tailed test. Correlations greater than .21 were statistically significant at the .01 level (two-tailed).

Path Memory

Average difficulty of level tested (using the staircase procedure) was the only informative measure for this task. The split-half reliability of this measure was .50.

Arrival Time–One Object

Four measures were calculated. The first was the mean absolute error. The second was the mean error, which is a measure of response bias. The third measure was another measure of absolute error based on proportions: the correct answer divided by subject's response when the correct answer was larger than the subject's response, and the subject's response divided by the correct answer when the subject's response was larger. The fourth measure was an estimate of bias, again using proportions.

The means are presented in Table 2. The negative bias measures demonstrated that people tended to respond too quickly. The reliabilities and correlations between these measures are presented in Table 3.

Table 2

Means and Standard Deviations for the Arrival Time–One Object Task

Variable	Mean	SD
Absolute Error (msec)	977.0	434.0
Bias (msec)	-207.0	907.0
Percent Error	22.9	12.7
Percent Bias	-3.5	16.9

Table 3

Correlations and Split-half Reliabilities for the Arrival Time–One Object Task

	Absolute Error	Bias	Percent Error	Percent Bias
Absolute Error	(.985)	.00	.92	.00
Bias		(.994)	-.30	.9996
Percent Error			(.983)	-.30
Percent Bias				(.994)

The percentage bias and the absolute bias were highly correlated and extremely reliable, so only one (the absolute bias) was used in the subsequent analyses. The absolute error and percentage error were also correlated highly enough that only one of them (absolute error) was used in the subsequent analyses.

Arrival Time–Two Objects

A staircasing method was used for this task. The first trials represented the adjustment period and the initial starting level, so they were discarded. The data from the last two-thirds of the trials were used in the analysis.

There were five subtasks, corresponding to the different configurations of the moving objects. The means for the overall score (the average of the subtasks) and each subtask are presented in Table 4. The reliabilities and subtask correlations are presented in Table 5.

Table 4

Means and Standard Deviations for the Level Tested in Arrival Time–Two Objects Task

Variable	Mean	SD
Summed score	4.9	0.6
Subtask 1	4.1	1.0
Subtask 2	4.0	0.9
Subtask 3	5.8	0.9
Subtask 4	5.0	1.0
Subtask 5	5.4	0.8

Table 5

Correlations and Split-Half Reliabilities for the Arrival Time–Two Objects Task

	Summed Score	S1	S2	S3	S4	S5
Summed Score	(.63)	.63	.52	.62	.67	.59
S1		(.26)	.16	.18	.29	.25
S2			(.08)	.18	.05	.17
S3				(.60)	.32	.21
S4					(.41)	.29
S5						(.24)

The test-retest reliability of the second subtask was not statistically significant ($p = .29$); but its correlation with the third subtask was significant at p less than .01 and its correlation with the fifth subtask was significant at p less than .02. No subfactors were apparent in the correlational matrix.

A weighted score was calculated; but the subtasks had approximately equal variances, so the weighted score correlated with the summed score at $r = .9986$ and the weighted mean was not more reliable than the summed score. Therefore, the summed score was used in the subsequent analyses.

Arrival Time–Four Objects

The only dependent measure was the average difficulty of the level tested (using the staircase procedure). This measure had a split-half reliability of .30, which was disappointingly low.

Extrapolation

The dependent measure was the amount of error observed for each of the three different curves to be extrapolated. The means of the errors are presented in Table 6. The sine wave had more variance than the other two types of curves. This meant that performance on the sine waves had a larger effect on the overall score than the performance on the other two curves. To correct this, a weighted score was calculated for each subject by summing standardized scores within a curve type. The reliabilities and intermeasure correlations are presented in Table 7. The weighted score was used in subsequent analyses.

Table 6

Means and Standard Deviations for the Extrapolation Task

Variable	Mean	SD
Overall	10.87	2.06
Line	5.60	1.73
Sine	20.81	4.96
Parabola	6.21	1.53

Table 7

Correlations and Split-Half Reliabilities for the Extrapolation Task

	Overall	Weighted	Line	Sine	Parabola
Overall	(.66)	.92	.54	.91	.48
Weighted		(.81)	.74	.65	.72
Line			(.74)	.22	.35
Sine				(.77)	.17
Parabola					(.69)

Intercept

Three different curves were used in this task. Following the logic used in the analysis of the extrapolation task, a weighted score was calculated for each subject, combining the scores obtained on each curve. Means are presented in Table 8, and correlations and reliabilities are presented in Table 9. The weighted score was used in subsequent analyses.

Table 8

Means and Standard Deviations for the Intercept Task

Variable	Mean	SD
Overall	22.23	4.61
Line	16.14	5.38
Sine	19.09	5.18
Parabola	31.47	7.37

Table 9

Correlations and Split-Half Reliabilities for the Intercept Task

	Overall	Weighted	Line	Sine	Parabola
Overall	(.74)	.995	.77	.72	.81
Weighted		(.79)	.80	.77	.74
Line			(.76)	.45	.40
Sine				(.65)	.32
Parabola					(.57)

Perceptual Comparison

The objects to be compared were polygons with randomly located vertices. In previous research with highly practiced subjects (Cooper, 1976), the latency of the comparison did not increase linearly with the number of vertices to be compared. However, the present experiment used unpracticed subjects. Given the possibility of a linear relationship for unpracticed subjects, the appropriate linear function was determined individually for each subject. Separate functions were computed for "same" and "different" trials. Error rates were also recorded separately for same and different trials. This yielded six measures per subject: the errors rates and the slope and intercept parameters for same and different trials. The means are presented in Table 10 and the correlations in Table 11. The data from three subjects could not be analyzed.

In this task, as in all of the other static tasks, slopes and intercepts were calculated. To avoid spurious negative correlations arising from measurement error, half of the trials (selected randomly) were used to calculate the slope; and the other half were used to calculate the intercept.

Table 10

Means and Standard Deviations for the Perceptual Comparison Task

Label	Variable Description	Mean	SD
P1	Overall Decision Latency (sec)	2.3	0.7
P2	Latency "Same" Trials (sec)	2.8	2.0
P3	Latency "Different" Trials (sec)	1.7	0.5
P4	Slope "Same" Trials (msec)	137.0	78.0
P5	Intercept "Same" Trials (msec)	596.0	723.0
P6	Slope "Different" Trials (msec)	97.0	65.0
P7	Intercept "Different" Trials (msec)	834.0	246.0
P8	Overall Error Rate (percent)	6.1	5.2
P9	Error Rate "Same" Trials (percent)	4.5	5.2
P10	Error Rate "Different" Trials (percent)	7.6	6.5

Table 11

Correlations and Split-Half Reliabilities for the Perceptual Comparison Task

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	(.98)	.97	.80	.65	.38	.60	.31	-.51	-.27	-.52
P2		(.99)	.63	.65	.42	.55	.17	-.56	-.25	-.60
P3			(.93)	.46	.19	.57	.59	-.27	-.25	-.18
P4				(.85)	-.40	.40	.10	-.32	-.09	-.40
P5					(.82)	.19	.08	-.26	-.17	-.24
P6						(.67)	-.27	-.32	-.23	-.27
P7							(.78)	.02	-.07	.08
P8								(.93)	.75	.83
P9									(.93)	.24
P10										(.92)

Note. Labels are explained in Table 10; N = 167.

The data in Table 11 indicate that the slower responders tended to be more accurate. Virtually all the correlations between latency and error rate measures are negative and, although generally low in magnitude, many of the correlations are reliably different from zero. This suggests a moderate speed-accuracy tradeoff across subjects and can be considered an argument for the need to make an analysis of errors controlled for latency or vice versa. Note that such analyses are difficult to accomplish using paper-and-pencil tests.

The correlation between error rates on "same" and "different" trials is surprisingly low (.24). This could be the result of response biases. A person who was biased to respond "same" would tend to have *high accuracy on "same" trials* and lower accuracy on "different" trials.

Mental Rotation

The two objects to be compared differed in how much one had to be rotated to have the same alignment as the other. The decision latency could be broken down into a slope (increase in response latency per degree of rotation) and an intercept (response time when no rotation is necessary). Error rates were also recorded. The analysis parallels that for the perceptual comparisons task. The means and standard deviations are presented in Table 12 and the correlations in Table 13. The data from three subjects could not be analyzed.

Table 12

Means and Standard Deviations for the Rotation Task

Label	Variable Description	Mean	SD
R1	Response Latency (sec)	2.66	0.90
R2	Latency "Same" Trials (sec)	2.60	0.90
R3	Latency Different Trials (sec)	2.72	0.90
R4	Slope Same Trials (msec)	8.00	6.00
R5	Intercept Same Trials (msec)	1165.00	452.00
R6	Slope Different Trials (msec)	7.00	6.00
R7	Intercept Different Trials (msec)	1312.00	503.00
R8	Overall Error Rate (percent)	4.00	6.50
R9	Error Rate Same Trials (percent)	4.40	6.50
R10	Error Rate Different Trials (percent)	3.60	9.00

Table 13

Correlations and Split-Half Reliabilities for the Rotation Task

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
R1	(.99)	.97	.98	.75	.70	.68	.77	.04	-.05	.09
R2		(.98)	.90	.76	.72	.66	.72	.01	-.01	.02
R3			(.97)	.70	.64	.67	.77	.07	-.09	.15
R4				(.85)	.17	.73	.40	-.22	-.21	-.15
R5					(.86)	.28	.72	.23	.20	.15
R6						(.78)	.22	-.24	-.11	-.23
R7							(.88)	.02	.02	.02
R8								(.97)	.63	.84
R9									(.95)	.12
R10										(.98)

Note. Labels are explained in Table 12; N = 167.

The correlations between latency and error measures were all low and tended to be negative, again suggesting that there was an across-subjects speed-accuracy tradeoff. The low correlation between error rates on "same" and "different" trials suggests the presence of a response bias.

Adding Detail

The subject controlled the rate of presentation of the dots and the time to respond, so there was both a presentation and a response latency. Reaction times were analyzed as a linear function of the number of dots presented. On a trial, the answer could be either "same" or "different."

The means and standard deviations are presented in Table 14 and the correlations in Table 15. The data from three subjects could not be analyzed.

Table 14

Means and Standard Deviations for the Adding Detail Task

Label	Variable Description	Mean	SD
A1	Overall Latency Stimulus Presentation (sec)	1.3	0.8
A2	Slope Presentation Latency (msec)	17.0	129.0
A3	Intercept Presentation Latency (msec)	958.0	659.0
A4	Overall Response Latency (sec)	2.8	0.9
A5	Response Latency "Same" Trials (sec)	2.9	0.9
A6	Response Latency "Different" Trials (sec)	2.8	1.0
A7	Overall Error Rate (percent)	24.0	9.0
A8	Error Rate "Same" Trials (percent)	31.0	15.0
A9	Error Rate "Different" Trials (percent)	14.0	8.0

Table 15

Correlations and Split-Half Reliabilities for the Adding Detail Task

	A1	A2	A3	A4	A5	A6	A7	A8	A9
A1	(.99)	.29	.78	.64	.62	.59	-.35	-.38	-.10
A2		(.95)	-.33	.14	.09	.18	-.24	-.23	-.13
A3			(.97)	.48	.50	.41	-.20	-.23	-.05
A4				(.91)	.94	.95	-.26	-.31	-.03
A5					(.84)	.78	-.20	-.23	-.04
A6						(.87)	-.29	-.35	-.01
A7							(.60)	.90	.62
A8								(.90)	.21
A9									(.90)

Note. Labels are explained in Table 14; N = 167.

Integrating Details

In this task, the problems differed in how many pieces had to be integrated. Therefore, both the presentation latency (which the subject controlled), and the response latency could be broken down into a slope (time per additional piece) and an intercept.

The means for this task are presented in Table 16, and the correlations are presented in Table 17. The data from one subject could not be analyzed.

Table 16

Means and Standard Deviations for the Integrating Details Task

Label	Variable Description	Mean	SD
I1	Overall Response Latency (msec)	3826.0	1807.0
I2	Slope Response Latency (msec)	72.0	378.0
I3	Intercept Response Latency (msec)	3254.0	1846.0
I4	Overall Presentation Latency (sec)	19.8	8.4
I5	Slope Presentation Latency (sec)	4.8	2.7
I6	Intercept Presentation Latency (sec)	-3.4	7.8
I7	Overall Error Rate (percent)	29.0	13.0
I8	Error Rate "Same" Trials (percent)	27.0	14.0
I9	Error Rate "Different" Trials (percent)	31.0	16.0

Table 17

Correlations and Split-Half Reliabilities for the Integrating Details Task

	I1	I2	I3	I4	I5	I6	I7	I8	I9
I1	(.94)	.10	.70	.45	.26	.02	-.01	.05	-.06
I2		(.09)	.50	-.11	.10	-.24	-.14	-.03	-.19
I3			(.40)	.24	.15	-.02	.07	.18	-.05
I4				(.92)	.75	-.23	-.26	-.22	-.22
I5					(.67)	-.80	-.43	-.37	-.36
I6						(.26)	.40	.34	.34
I7							(.75)	.83	.87
I8								(.64)	.45
I9									(.69)

Note. The labels are explained in Table 16; N = 169.

The correlations between latency and error rates were again negative, although generally much lower than was the case for the previous two tasks. The correlation between error rates in positive and negative trials was substantially higher than before, especially considering the lowered reliability of the error rate measures. Evidently, speed-accuracy tradeoffs and response biases were less of a factor in this task.

Surface Development

The cubes to be folded varied in complexity, depending upon the number of surfaces that had to be carried along when folding. The means are presented in Table 18 and the correlations in Table 19. The variety of trial types approached the number of trials, making the calculation of and odd-even reliability difficult for several of the measures. These odd-even reliabilities are omitted from Table 19. The data from eight subjects could not be analyzed. The error rate and response latency were used in further analyses.

Table 18

Means and Standard Deviations for the Surface Development Task

Label	Variable Description	Mean	SD
S1	Latency (sec)	7.8	2.5
S2	Slope of Latency (msec)	438.0	267.0
S3	Intercept of Latency (sec)	3.2	1.8
S4	Latency Same (sec)	8.1	2.3
S5	Latency Slope Same (msec)	555.0	382.0
S6	Latency Intercept Same (sec)	2.8	2.3
S7	Latency Different (sec)	7.6	3.2
S8	Latency Slope Different (msec)	333.0	242.0
S9	Latency Intercept Different (sec)	3.8	2.2
S10	Error Rate (percent)	14.0	10.0
S11	Error Rate Same (percent)	14.0	10.0
S12	Error Rate Different (percent)	13.0	13.0

Table 19

Correlations and Split-half Reliabilities for the Surface Development Task

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	(.95)	.55	.39	.89	.43	.15	.93	.52	.63	.08	.06	.08
S2		(.73)	-.52	.57	.84	-.57	.45	.72	-.11	-.44	-.36	-.42
S3			(.72)	.24	-.51	.83	.45	-.25	.79	.53	.43	.50
S4					.57	.03	.66	.39	.41	-.08	.00	-.13
S5						-.78	.25	.43	-.13	-.42	-.28	-.46
S6							.21	-.21	.49	.46	.35	.45
S7								.54	.70	.19	.09	.24
S8									-.18	-.24	-.27	-.16
S9										.45	.34	.45
S10										(.95)	.86	.91
S11												.58

Note. Labels are explained in Table 18; N = 162.

Paper-and-Pencil Tests

Descriptive statistics for the scores on the paper-and-pencil tasks are presented in Table 20. These measures are the ones normally used for each of the tests.

Table 20

Descriptive Statistics for the Pencil-and-Paper Tasks

Label	Test Name	Mean	SD	Min.	Max.
PP1	Spatial Test from DAT	40.8	10.4	13.0	60.0
PP2	Raven's Matrices	12.1	3.3	0.5	18.0
PP3	Identical Pictures	75.4	12.7	42.0	96.0
PP4	Spatial Memory	20.2	7.2	-4.0	32.0
PP5	Vocabulary	47.2	17.7	-0.5	90.0
PP6	Spatial Orientation	20.5	10.8	-3.0	54.5
PP7	2-D Rotation	45.1	10.7	19.0	67.0
PP8	3-D Rotation	44.7	17.2	-8.0	88.0

Correlations Between Tasks

The correlations between selected measures for each task are presented in Appendix B.

Multivariate Analyses

General Comments

To reduce the variables to be analyzed to a manageable number, separate factor analyses were conducted on all variables within each task, using an orthogonal factor analysis followed by varimax rotation (Mulaik, 1972). At most, three factors were extracted from each of the within-task measures. Only the best marker for each of these factors was retained for further analysis. The measures retained are shown in Table 21. This table also shows a variable acronym that will be used in later tables to refer to the variable in question. In the following analyses, measures of error and latency were reflected. Thus, a high score reflects better performance. The only two exceptions were two neutral measures, the measure of bias in the Arrival Time--One Object task and sex of subject.

Table 21

Variables Used in the Multivariate Analyses and Their Acronyms

Label	Source Task	Variable Description
Dynamic Tasks		
PTHMEM	Path Memory	Difficulty Level[a]
ARV1A	Arrival Time–One Object	Accuracy, Absolute Value
ARV1B	Arrival Time–One Object	Bias, Signed Accuracy
ARV2	Arrival Time–Two Objects	Accuracy
ARV4	Arrival Time–Four Objects	Difficulty Level[a]
EXTRAP	Extrapolation	Weighted Accuracy
INTCPT	Intercept	Weighted Accuracy
Static Tasks		
PCOMRL	Perceptual Comparison	Response Latency
PCOMRA	Perceptual Comparison	Accuracy
MRLAT	Mental Rotation	Response Latency
MRACC	Mental Rotation	Accuracy
ADDPL	Adding Detail	Dot Viewing Latency
ADDRL	Adding Detail	Decision Latency
ADDACC	Adding Detail	Accuracy
INTGPL	Integrating Detail	Puzzle Integration Time
INTGDL	Integrating Detail	Decision Latency
INTGAC	Integrating Detail	Accuracy
SDLAT	Surface Development	Response Latency
SDACC	Surface Development	Accuracy
Paper-and-Pencil Measures		
DAT	DAT Space	Number Correct
RAVENS	Raven's Matrices	Number Correct
IDPICT	Identical Pictures	Number Correct
SHMEM	Shape Memory	Number Correct
VOCABL	Vocabulary Test, Nelson-Denny	Number Correct
SPAORT	Spatial Orientation	Number Correct
PMA	PMA Space	Number Correct
3DROT	3-D Mental Rotation	Number Correct
SEX	Sex Of Subject	Female (0) and Male (1)

[a] Estimated on final two-thirds of the items.

Multivariate analyses were conducted to answer three questions. First, to determine the level of uniformity of the population, comparisons were made between the data obtained from the University of Washington and the University of California, Santa Barbara, subsamples. Second, factor analyses were conducted within the three subgroups of tasks: static, dynamic, and paper-and-pencil measures. The purpose of these analyses was to determine the dimensionality of each of the three classes of spatial-visual tests. A canonical correlation analysis (Cohen & Cohen, 1975) was conducted to determine whether or not the information about individual differences contained within the paper-and-pencil tests was identical to the information captured by the computer-administered static tasks. A confirmatory factor analysis (Joreskog & Sorbom, 1979) was conducted to test the hypothesis that the factor(s) captured by the static tasks were not identical to those captured by the dynamic tasks.

Comparison of the UW and UCSB Subsamples

Discriminant analysis (Cohen & Cohen, 1975) was used to compare the mean performances of the UW and UCSB samples. The (single) discriminant function extracted was highly reliable (Wilkes' λ = .673, χ^2 (28) = 54.23, $p < .01$). Of the 170 available cases, 78 percent were correctly classified by the function.

Table 22

Means for Both Subsamples for Each Variable and the Standardized Coefficients on the Discriminant Function Separating the Two Subsamples

Variable	Mean UCSB	Mean UW	Coefficient
Dynamic Tasks			
Path Memory	4.4	4.7	.12
Arrival Time–One Object, Accuracy	988.0	970.0	.14
Arrival Time–One Object, Bias	-210.0	-208.0	.21
Arrival Time–Two Objects	4.9	4.9	-.05
Arrival Time–Four Objects	4.9	5.0	.19
Extrapolation	11.4	11.5	-.08
Intercept	33.2	34.3	.01
Static Tasks			
Perceptual Comparisons, Latency	2.2	2.3	-.17
Perceptual Comparisons, Accuracy	94.3	93.5	-.27
Mental Rotation Latency	2.6	2.8	-.43
Mental Rotation Accuracy	96.1	95.9	.04
Adding Detail, Dot Viewing Latency	1.4	1.3	-.31
Adding Detail, Decision Latency	3.1	2.7	.64
Adding Detail, Accuarcy	77.3	77.2	.12
Integrating Detail, Integration Time	22.0	17.5	.76
Integrating Detail, Decision Latency	3.9	3.7	-.12
Integrating Detail, Accuracy	72.0	70.1	-.02
Surface Development, Latency	7.6	8.0	-.22
Surface Development, Accuracy	87.1	85.6	-.26
Paper-and-Pencil Tests			
DAT Space	40.8	40.8	.18
Raven's Matrices	12.0	12.3	.28
Identical Pictures	78.3	72.4	-.58
Shape Memory	20.1	20.4	.21
Nelson-Denny Vocabulary	45.9	48.5	.10
Spatial Orientation	20.4	20.5	.23
PMA Space	46.1	44.0	-.08
3-D Mental Rotation	45.8	43.6	.05
Sex of Subject	.55	.55	.05

Table 22 shows the means of the two subsamples for each of the variables in Table 21. Table 22 also shows the weights of each variable in the standardized discriminant function (reflected such that a positive coefficient represents superior performance by the UW sample). Neither sample has any clear-cut advantage in performance, although the relative difficulty of tasks varied slightly (and reliably) across samples.

The total set of variables (Table 21) includes two measures, Raven's Advanced Progressive Matrices and a vocabulary test, that normally would not be considered measures of spatial-visual ability. The discriminant analysis was repeated with these measures removed, with essentially the same results.

Discriminant analysis provides a measure of the differences in the location of observations from two samples, both embedded in the same multivariate space. The current study is, if anything, more concerned with measures of covariance between the measures defining the space. Therefore it was important to test the assumption that the relation between variables was the same in each sample. A linear structural relations (LISREL) approach was used to investigate this question. The hypothesis of equal correlation matrices in each sample was tested, using the methods outlined in Joreskog and Sorbom (1979). Because the computing time required for this analysis increases roughly as the cube of the number of variables, it was not feasible to test the correlation matrices for all variables simultaneously. Therefore, correlation matrices were compared separately for the measures obtained by paper-and-pencil, static, and dynamic tests. Three measures of fit were obtained: the chi-square values for discrepancies of the observed data from that assumed by the hypothesis, the ratio of the chi-square values to the appropriate value for degrees of freedom (df), the Joreskog goodness-of-fit index (which runs from 0, for random fit, to 1, for perfect agreement of hypothesis and data), and the root mean square deviation of each observed correlation from that predicted under the hypothesis of equality of all correlations. Table 23 presents the appropriate statistics.

Table 23

Goodness-of-Fit Statistics for the Hypothesis of Equal Correlation Matrices Across All Tasks

Task Domain	χ^2	df	p <	χ^2/df	Joreskog Index	RMSD[a]
Paper-and-Pencil	42.71	36	.20	1.19	.92	.08
Static Tasks	101.82	78	.05	1.29	.88	.08
Dynamic Tasks	24.00	28	.50	.86	.96	.06

[a]Root mean square deviation.

These results generally support the conclusion that there is no substantial difference between the correlation matrices obtained for each subsample within any of the domains, even though it might be possible to construct a statistically reliable between-samples discriminator. The fact that reliable discrimination can be obtained is probably due to the extreme sensitivity of the chi-square measure at large degrees of freedom. For samples of the size used in this study, the standard deviation of each estimated correlation coefficient is slightly greater than .10, which is about the size of the mean squared residuals obtained from the models within each domain that assumed equality of the correlational matrices.

Based on these results, it was concluded that while the samples were drawn from slightly different locations, the samples could be combined without committing any major violation of the assumption that we were dealing with a multivariate normal population. This was confirmed by a visual examination of the plots of the distributions of scores from each of the two samples along the discriminant function. The resulting distribution is somewhat platykurtic, but no more so than is often observed in multivariate studies. Therefore the samples were combined for further analyses.

Within-Domain Factor Analyses

Tasks within each domain were analyzed to determine their dimensionality. Recall, from the introductory discussion, that the paper-and-pencil tests would be expected to show either a two- or a three-dimensional factor pattern. The dimensionality of the spaces of the two sets of computer administered tests was an open question.

Table 24 presents the results of an orthogonal factor analysis of the paper-and-pencil tests (variables 21-28 in Table 20), followed by a varimax rotation (Mulaik, 1972). As expected, the analysis identified two factors with eigenvalues greater than 1. The two-dimensional space accounted for 54 percent of the variance between measures. The first factor is closely related to the two rotation tests but is also associated with all other paper-and-pencil measures of spatial-visual ability. The Raven matrix test has only a small loading on this factor, and the vocabulary test is virtually orthogonal to this factor. The second factor is identified by relatively high loadings on the more complex spatial visualization tests and by a very high loading on the Raven matrix test. This test depends both on spatial and abstract reasoning (Hunt, 1974).

Table 24

The Factor Matrix of the Paper-and-Pencil Tests, After Orthogonal Factor Analysis and Varimax Rotation

Test	Factor I	Factor II
DAT Space	.53	.53
Raven's Matrices	.28	.73
Identical Pictures	.59	.06
Shape Memory	.16	.34
Nelson-Denny Vocabulary	-.01	.25
Spatial Orientation	.42	.58
PMA Space	.70	.18
3-D Mental Rotation	.62	.25

These results indicate that the paper-and-pencil test scores obtained in the current study are distributed much as one would expect them to be on the basis of previous research. The pattern of results could be explored further; but there would be relatively little point in doing so because we would be exploring a very well-studied topic, using conventional measures.

A similar factor analysis was conducted for the computer-administered static tasks. The factor matrix is shown in Table 25.

Table 25

Factor Matrix for Computer-Controlled Static Tasks

Variable	Factor I	Factor II	Factor III	Factor IV
Perceptual Comparison, Latency	.06	.20	.04	.80
Perceptual Comparison, Accuracy	.30	-.07	-.23	-.62
Mental Rotation Latency	.27	.38	.10	.41
Mental Rotation Accuracy	.42	-.09	.04	-.03
Adding Detail, Viewing Latency	-.10	.14	.84	.07
Adding Detail, Decision Latency	-.08	.30	.67	.21
Adding Detail, Accuracy	.46	-.07	-.27	-.09
Integrating Detail, Integration Time	-.29	.73	.18	.02
Integrating Detail, Decision Latency	-.13	.66	.16	.14
Integrating Detail, Accuracy	.69	.00	-.05	.04
Surface Development, Latency	.09	.60	.09	.16
Surface Development, Accuracy	.75	.01	-.10	.02
Proportion of variance accounted for by factor	.28	.18	.10	.10

The largest factor shows the highest loadings on measures of accuracy. The second factor generally has positive loadings on latency measures. The third factor is associated with the latency in the adding detail task. Unlike the other tasks, the adding detail task contained an advantage of viewing the stimuli quickly: The more quickly the stimuli were viewed, the shorter the time the stimuli had to be remembered. The fourth factor appears to be primarily associated with the measures taken on the perceptual comparisons task.

To examine these findings further, an oblique (oblimin) factor analysis was conducted. This method was chosen because the psychological processes just described would predict nonindependence between measures of speed and accuracy, both within and across tests. Table 26 presents the results of the oblique factor analysis, and Table 27 shows the correlations between the factors.

Table 26

Factor Pattern Matrix for the Oblimin Solution for Computer-Controlled Static Tasks

Variable	Factor I	Factor II	Factor III	Factor IV
Perceptual Comparisons, Latency	.33	.07	-.82	.17
Perceptual Comparisons, Accuracy	-.21	.31	.65	-.36
Rotation Latency	.44	.27	-.48	.17
Rotation Accuracy	-.11	.41	.05	-.05
Adding Detail, Viewing Latency	.27	-.16	-.21	.86
Adding Detail, Decision Latency	.42	-.13	-.35	.74
Adding Detail, Accuracy	-.14	.49	.14	-.35
Integating Detail, Integration Time	.75	-.31	-.20	.35
Integating Detail, Decision Latency	.69	-.15	-.29	.31
Integating Detail, Accuracy	-.03	.70	-.02	-.15
Surface Development, Latency	.62	.08	-.29	.20
Surface Development, Accuracy	-.04	.76	.00	-.20

Table 27

Correlation Between Oblique Factors for Computer-Controlled Static Tasks

	I	II	III	IV
Factor I	1.00	-.07	-.37	.33
Factor II		1.00	.01	-.23
Factor III			1.00	-.29
Factor IV				1.00

The pattern in Table 26 is similar to that in Table 25, except that Factors I and II are reversed, as are Factors III and IV. Factor I has high loadings on latency measures and only small loadings on accuracy measures. Factor II, conversely, is characterized by high loadings on accuracy measures. The two factors are essentially uncorrelated. Superimposed on this pattern, however, are Factors III and IV. Factor III is a bipolar factor for the perceptual identification task, suggesting a strong speed-accuracy tradeoff across individuals for this task. (This is consistent with the results of experiments studying the task in detail and with the within-task analysis reported above.) Factor IV is a similar, somewhat less strongly defined speed-accuracy tradeoff for the adding detail and integrating details task.

The conclusion that can be drawn from this analysis is that the computer-controlled spatial reasoning tasks offer the potential for distinguishing between speed-accuracy tradeoffs and the spatial reasoning and visualization factors. More research will be required to determine the best measures to use for this purpose.

A third orthogonal factor analysis was computed using the measures obtained with the computer-controlled dynamic tasks. The factor analysis before rotation indicated that a two-factor solution was required, with 44 percent of the common variance between tests located in a two-dimensional space. This solution was made simpler by a varimax rotation, which identified two factors. These are shown in Table 28. The first factor, which accounted for 71 percent of the common (two-space) variance, was marked by high loadings on the two- and four-object arrival time tasks and the intercept task. The second factor was associated primarily with the extrapolation task. Note that the extrapolation task can be solved without considering its dynamic aspects, since at the time the examinee must respond, a static picture is present, and the information in the picture is sufficient to define the correct answer. The communalities of the path memory task and the measures in the arrival time–one object task were very low (less than .1 in all cases), suggesting that these measures tap processes that are not measured by the other tasks.

Table 28

Factor Matrix for the Computer-Controlled Dynamic Tasks, After Varimax Rotation

Variable	Factor I	Factor II
Path Memory	.26	.19
Arrival Time–One Object, Accuracy	.14	-.21
Arrival Time–One Object, Bias	.23	.24
Arrival Time–Two Objects	.59	.05
Arrival Time–Four Objects	.46	.03
Extrapolation	.23	.74
Intercept	.55	.05

Cross-Domain Comparisons

The purpose of the cross-domain comparisons was to determine whether or not the three domains of tests tap the same psychological abilities. Two classes of measurement are of interest, those expressing the relation between the paper-and-pencil and computer-controlled static tasks, and those expressing the relation between the static and dynamic tasks. The first relationship determines whether or not the individual variation captured by the paper-and-pencil measures is embedded in the static tasks. The second relationship of interest was whether or not the dynamic tasks introduced a factor different from those required to summarize individual variation in the static tasks.

A canonical correlation analysis was conducted to make the first comparison. A canonical correlation locates spaces of common variance embedded within each of the two domains (paper-and-pencil and static tasks), and then computes the (maximized) correlations between the dimensions of each of the spaces (Cohen & Cohen, 1975). At most, two canonical correlations should be constructed, because the common variance in the lower order space (the paper-and-pencil tests) is apparently two dimensional (see above, in the discussion of within-domain analyses). As expected, two canonical correlates were extracted. The first canonical correlation was .78 and the second .54, both statistically reliable at p less than .01. The two canonical variates from the static tasks accounted for 50 percent[1] of the (conditional) paper-and-pencil generalized variance. It was concluded that a substantial portion of the common variance in the paper-and-pencil domain is embedded within the common variance of the computer-controlled static tasks.

Canonical analysis was also used to explore the connection between the dynamic tasks and the other two domains. In each case, we would expect at most two canonical correlates because of the low dimensionality of the dynamic tasks. There were two canonical correlates between the static and dynamic tasks, with correlation values of .59 and .46 ($p < .001$ and .05, respectively). The two static task canonical variates account for only 25.3 percent of the (conditional) generalized variance in the dynamic tasks. There was a single significant canonical correlate connecting the paper-and-pencil and dynamic tasks, with a canonical correlation of .60 ($p < .001$). The paper-and-pencil canonical variable accounts for only 19.5 percent of the generalized dynamic task variance. These data indicate that the dynamic tasks tap processes that are correlated with, but not identical to, the processes required for performance in the other two task batteries.

As a further test, a confirmatory factor analysis (Joreskog & Sorbom, 1979) was conducted, analyzing the common covariance of the static and the dynamic computer-controlled tasks. In this analysis, attention was restricted to measures from the intercept and moving object tasks, as they gave the clearest indication of being good measures of a dynamic motion factor. Three principles were used to construct the hypothesized factor structure. They were

---

[1]Stewart and Love (1968) point out that a canonical correlation is the optimized correlation between two linear composites and thus has some interpretive problems. Specifically, whereas a squared multiple correlation represents the proportion of criterion variance accounted for by a predictor set, a squared canonical correlation is the shared variance between linear composites of two sets of variables and may not represent the shared variance of the two sets. An example contrived by J. Brad Sympson (in a personal communication with D. Alderton) makes this point. Assume a four variable example where variables 1 and 2 are correlated 0.99 and variables 3 and 4 are similarly correlated (.99). Furthermore, variables 1 and 3 are intercorrelated 0.02 while all other correlations are 0. If such a matrix is submitted to a canonical analysis the first canonical correlation will be 1.00 suggesting complete redundancy between the two sets, when, in fact, there is only 0.04 percent common variance.

The Stewart and Love (1968) index corrects for this problem. In the original paper the authors suggest using the index for the first m canonical correlates where m is the lesser of the number of variables in either set. However, in the current research the dimensionality of the submatrices (paper-and-pencil, static, dynamic) was separately explored. The results from these analyses showed that only two dimensions could be extracted from the canonical analyses. Given this limitation, the proportion of redundant variance will be conditional on the proportion of criterion variance accounted for in a two dimensional solution. The present statistic, then, is obtained in this way:

$$.268/.540 = .496 \sim .50$$

In words, this measure implies that the first two dimensions of the static tasks account for 50 percent of the variance in the two dimensions of the paper-and-pencil tests. It should be noted that this redundancy index is not symmetric so it cannot be said that the first two paper-and-pencil dimensions account for 50 percent of the static task variance; indeed, they account for less than one-third of the static task (two-dimensional) variance.

1. Three factors were assumed: a latency factor, an accuracy factor, and an "ability to deal with dynamic motion" factor. The first factor was defined by all latency measures on the static tasks, the second by all accuracy measures on the static tasks, and the third by the dynamic tasks.

2. Correlations between the three factors were permitted.

3. The "task specific" (residual) components of each measure taken from the same task were assumed to correlate. The rationale behind this is that these measures are based on different analyses of the same physical response. Any event in time that affects a response but is logically irrelevant to the test situation (e.g., the examinee's attention being momentarily distracted) should affect all measures based on the same response.

The statistics for a fit of this model are shown in Table 29. The chi square value and the chi square divided by degrees of freedom ratios are sufficiently high to be of concern, but the goodness-of-fit indices and the root mean square values are comparable to those obtained for the within domain models. Most importantly, examination of the details of the deviations from a perfect fit indicated that the problems were in the relations between the various *static* test measures, and not in the relation between the static and dynamic tests. Further support for a separate dynamic movement processing factor was provided by a similar confirmatory factor analysis, in which the dynamic process factor was eliminated and the dynamic tasks were related directly to the latency and accuracy factors. Although this model has fewer degrees of freedom than does the model of Table 29, the chi square value was higher. A direct statistical comparison between these two models is not possible, because one model does not fully contain the parameters of the other. However, the second model has more parameters and a worse fit, so it is hardly preferable to the first model, which has fewer parameters and a better fit. We conclude that there is strong evidence for a dynamic movement processing factor that is separate from the spatial-visual reasoning factors previously identified.

Table 29

Indices of Fit for Confirmatory Factor Analysis of the Model Assuming Separate Latency, Accuracy, and Dynamic Processing Factors. (See text for explanation.)

χ^2	dof	$p <$	χ^2/dof	Joreskog Index	RMSD
163.07	78	.001	2.09	.85	.10

CONCLUSIONS

The conclusions of this study can be stated quite simply. The within-domain analysis and the comparison of the paper-and-pencil and static domains indicate that the individual variation in spatial visual reasoning captured by conventional paper-and-pencil tests can also be captured by constructing computer-controlled analogs of these tests. The analogs are preferable to the paper-and-pencil tests because they provide measures of speed and accuracy of spatial visual reasoning within a single task. We have shown that such measures each carry different psychological information. The ability to separate speed from accuracy in the static tasks may be useful in predicting future job performance. In addition, the computer-controlled static tests provide a way of measuring individual differences in speed-accuracy tradeoffs. Very little research has been done to explore speed-accuracy tradeoff measures as predictors, although speed-accuracy tradeoffs have been shown to be related to age and to personality factors. Computer-controlled testing offers a chance to coordinate the measurement of personality and "cognitive" factors (if one cares to make the distinction) within the spatial reasoning domain.

Our results indicate strongly that the ability to deal with moving elements in a spatial display is separate from the ability to deal with static visual displays. Two tasks, Intercept and Arrival Time--Two Objects, are promising measures of a new dimension of spatial-visual ability. Arrival Time--One Object might measure an additional new dimension.

The computer-controlled tasks presenting static displays have all been used extensively, both in psychometrics and experimental psychology. The advantage of computer control lies in the potential for finer analyses of responses. We do not expect to find computer-controlled static tasks that are better than the tasks that we have used simply because these tasks have a long history of development and use. By contrast, the tasks involving dynamic displays are presented here for the first time. There is no reason to believe that these are the best tasks that could be developed. The construction of dynamic visual display tasks would therefore seem to be a promising area for psychometric research.

Our results do not address the question of the utility of either static or dynamic computer-controlled tasks as predictors of performance in situations outside the laboratory. This is an area of considerable interest, but it will have to be the topic of new studies. The work here is a necessary precursor to such research.

FURTHER RESEARCH

Computer-controlled tasks might be measuring some dimensions of spatial-visual ability that are not measured by paper-and-pencil tests. Three questions can be asked. (1) Is the existence of these new dimensions reliable–that is, will these new dimensions appear in different settings with different populations? (2) How should these new dimensions be characterized? (3) Will these new dimensions be useful for predicting job performance? The next logical step, which has the potential for answering all three of these questions, is to determine if performance on these tasks correlates with later job performance. Thus, this is the next step the Navy should take.

RECOMMENDATIONS

The battery of tests should be used in studies of Navy personnel to determine if the abilities measured by these tests predict performance in a variety of jobs.

REFERENCES

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). *Differential Aptitude Test.* New York: The Psychological Corporation.

Brown, J. I., Bennett, J. M., & Hanna, G. (1981). *Nelson-Denny Reading Test.* Riverside, CA: Riverside Publishing Co.

Carroll, J. B. (1982). The measurement of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 29-120). Cambridge: Cambridge University Press.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression and correlational analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cooper, L. A. (1976). Individual differences in visual comparison processes. *Perception & Psychophysics*, 19, 433-444.

Ekstrom, R. B., French, J. W., Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs* (No. 97-2), 1-84.

Guilford, J. P., & Zimmerman, W. S. (1947). *Guilford-Zimmerman aptitude survey.* Orange, CA: Sheridan Psychological Service.

Hunt, E. B. (1974). Quote the Raven? Nevermore! In L. Gregg (Ed.), *Knowledge and Cognition* (pp. 129-157). Hillsdale, NJ: Erlbaum.

Hunt, E. & Pellegrino, J. (1985). Using interactive computing to expand intelligence testing. *Intelligence*, 9, 207-236.

Joreskog, K., & Sorbom, D. (1979). *Advances in factor analysis and structural equation models.* Cambridge, MA: Abt Books.

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Lansman, M. (1981). Ability factors and the speed of information processing. In M. P. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 441-457). New York: Plenum Press.

Likert, R., & Quasha, W. H. (1970). *Revised Minnesota Paper Form Board Test (Series AA).* New York: The Psychological Corporation.

Lohman, D.F. (1979). *Spatial ability: A review and reanalysis of the correlational literature* (Technical Report No. 8, Aptitude Research Project). Palo Alto, CA: Stanford University.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, 86, 889-918.

Mulaik, S. A. (1972). *The foundations of factor analysis.* San Francisco: McGraw-Hill.

Pachella, R. G. (1974). The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41-82). Hillsdale, NJ: Erlbaum.

Pellegrino, J. W., & Kail, R. (1982). Process analyses of spatial aptitude. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 311-366). Hillsdale, NJ: Erlbaum.

Pellegrino, J. W., Mumaw, R. J., & Shute, V. J. (1985). Analyses of spatial aptitude and expertise. In S. E. Embretson (Ed.), *Test design: Developments in psychology and psychometrics* (pp. 45-76). New York: Academic Press.

Poltrock, S. E., & Brown, P. (1984). Individual differences in visual imagery and spatial ability. *Intelligence*, 8, 93-138.

Raven, J. C. (1962) *Advanced Progressive Matrices Set II.* London: H. K. Lewis.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

Stewart, D., & Love, W. (1968). A general canonical correlation index. *Psychological Bulletin*, 70, 160-163.

Swanson, L. (1979, November). *Armed Services Vocational Aptitude Battery, Forms 6 and 7, Validation Against School Performance in Navy Enlisted Schools (July 1976-February 1978).* Supplement to TR-80-01, Navy Personnel Research and Development Center, San Diego, CA.

Thurstone, L. L., & Thurstone, T. G. (1949). *SRA primary abilities.* Chicago: Science Research Associates.

Thurstone, T. G. (1965). *Primary mental abilities.* Chicago: Science Research Associates (SRA 7-234).

APPENDIX A

Detailed Description of Computer-Controlled Tasks

Detailed Description of the Computer-Controlled Tasks

Procedures for the Dynamic Tasks

Path Memory

The subject viewed an object moving in a path three different times. Either the second path was different from the first, or the third path was different from the second. In other words, in the sequence of three presentations, the path changed on either the second or the third presentation.

The path was a parabola, defined by its starting height on the left of the screen, the height of its apex, and how far it travelled from left to right before reaching the apex. These variables were all selected randomly, within some constraints that kept the parabola on the screen until it reached the bottom of the screen.

There were three different conditions, corresponding to the three different changes that could be made (changing the height of the starting point, changing the height of the apex, and changing the distance to the apex). There were eight levels of difficulty, corresponding to the size of the change.

Testing started at Level 4. When a subject answered two trials correctly in a row (at a condition), the condition was moved to a more difficult level. When a subject answered incorrectly, the condition was moved to an easier level. As in other tests, the odd and even trials were independent, to allow calculation of split-half reliabilities.

There were 24 trials per condition, for a total of 72 trials. The subject's response was either ONE or THREE, depending upon whether the change occurs on the second path or the third. The subject had to respond within the time-out period of approximately 7.4 seconds, or else the answer was deemed incorrect. The subject received feedback of either "RIGHT," "WRONG," or "TOO SLOW."

The following instructions were read to the subject before beginning testing: "You will see a small dot move across the screen THREE times. Each time the dot will TRACE an IMAGINARY PATH. One of the three paths will be DIFFERENT from the other two paths: Either the FIRST path will be different from the next two paths, or the THIRD path will be different from the first two paths. You should decide whether the FIRST path is different or the THIRD path is different. You can do this by watching each pair of paths closely. If the SECOND path is different from the FIRST, the answer has to be PATH ONE. If the THIRD path is different from the SECOND, the answer has to be PATH THREE. Wait until all three paths have been shown. The computer will then ask for your choice. Press the key marked ONE on the keyboard if you think the FIRST path was different. Press the key marked THREE if you think the THIRD path was different. You have SEVEN seconds to respond. You will receive a message indicating whether your answer was RIGHT or WRONG."

Arrival Time–One Object

In this task, an object travelled from left to right towards a vertical line. The object was a square. The distance from the object's beginning point to the wall was randomly selected from the range 200 to 260 pixels. The object travelled 70 to 100 pixels, selected randomly. There were five different speeds, selected randomly.

The subject's task was to press a key at the moment that the front edge of the object would have intersected the line on the right edge of the computer screen (had the object continued moving at the same speed in the same direction). The subject initiated each trial by pressing a key. The object first appeared in an enlarged form, then shrunk to normal size and began moving.

There were two blocks of 40 trials each, for a total of 80 trials. The subject had to respond within approximately 7 to 9 seconds after the object disappeared.

The following instructions were read to the subject before beginning testing: "When you are ready to START, you should PRESS the space-bar. You will see a small SQUARE moving from left to right on the screen. The square will be travelling towards a WALL. At some point, the square will DISAPPEAR. Your task is to decide WHEN the FRONT edge of that square would have REACHED the wall, if the square had continued travelling at the SAME speed. Press the SPACE-BAR on the keyboard at the EXACT moment you think the front edge of the object would have hit the wall. The computer will not tell you whether you are right or wrong; just TRY to do your BEST, doing what you think is correct."

Arrival Time–Two Objects

The general pattern of this experiment was that the subject saw two objects moving towards targets. The objects disappeared, and the subject estimated which of the two objects would have hit its target first.

There were five different configurations. In the first configuration, the objects moved perpendicularly towards different targets. In the second configuration, the objects moved perpendicularly towards the same target. In the third, fourth, and fifth configurations, the objects moved in parallel. In the third configuration, the targets were placed adjacently at an edge of the computer screen. In the fourth configuration, the paths of the objects were adjacent as in the third condition; but the objects started at the same location, and one target was closer than the other. In the fifth configuration, the targets were separated, one at the top of the computer screen and one at the bottom; both targets were on the edge of the screen, as in the third configuration.

One object started 220 to 260 pixels away from its target, and the other objects started 130 to 170 pixels away from its target. To determine the speeds of the objects for presentation, the speeds that the objects should travel in order to arrive at the target in a fixed time were calculated; then the speed of one of the objects was slowed by a constant. The constant depended upon the level of difficulty; so the larger this constant was, the easier the trial was. The objects disappeared when the faster object had travelled one-fifth of the way.

There were eight levels of difficulty, corresponding to how much one object was slowed down. The subject started at Level 4, where 8 was the most difficult level and 1 was the easiest level. If the subject answered correctly two trials in a row, testing moved to the next harder level; and if the subject answered incorrectly, testing moved to the next easier trial. The subject's score was the average level that was tested.

The level tested for an odd-numbered trial was determined by previous odd trials at that configuration, and the level for an even-numbered trial was determined by previous even trials at that configuration. Thus, a score and split-half reliability were calculated for each configuration. There were five blocks, with 50 trials per block, for a total of 250 trials. There were 50 trials at each configuration, placed in a random order evenly within blocks.

After each trial, there was a tone signalling that the subject could respond. Before this time, any response was ignored. The subject had to respond within approximately 4.5 seconds, or else the trial was scored as incorrect. The subjects received feedback, "RIGHT", "WRONG", or "TOO SLOW".

The following instructions were read to the subject before beginning testing: "TWO MOVING objects, a ONE and a ZERO, will appear on the screen. They will be moving towards two WALLS. Sometimes, the walls will be CLOSE together, sometimes the walls will be FAR apart, and sometimes the two walls will be at the EXACT SAME location, making the shape of a PLUS SIGN. The objects will DISAPPEAR before they hit the walls. Your job is to say WHICH number (one or zero) would have reached the wall FIRST if they hadn't disappeared. Assume that the objects would NOT have changed speed. Just decide which number would have reached its wall first if they continued to moved in the same manner. Here is how you should tell us which object would have gotten to its wall first. When the objects disappear, a TONE will sound. After the tone sounds, you can indicate your answer by PRESSING the key marked ONE or ZERO on the keyboard. For example, if you think the ONE would have arrived at its wall first, you should press the key marked ONE. You have FOUR seconds to respond. A message on the screen will tell you whether your answer was RIGHT or WRONG. To summarize, you SEE two moving objects, they DISAPPEAR, a TONE sounds, YOU INDICATE which object would have arrived at its wall FIRST, and you find out whether your answer was RIGHT or WRONG."

Arrival Time–Four Objects

In this task, four objects (the numbers one to four) moved from right to left towards a vertical line on the left edge of the computer screen. The vertical line extended from the top to the bottom of the computer screen, 10 pixels from the left. The objects started at 170 to 180, 200 to 210, 230 to 240, and 260 to 270 pixels from the left of the screen. Which object was at which distance was random, and where an object was located within those ranges was random. The heights of the objects were 80, 100, 140, and 170 pixels from the bottom of the screen; which object was at which height was determined randomly.

Three of the objects were moving at a speed such that they would arrive at the left edge simultaneously; but one object was moving faster, such that it would arrive at the left edge before the others. All four objects disappeared when they were halfway to the wall (except for the fast object, which was more than halfway). The faster object ended up 2 to 16 pixels to the right of the above points, depending upon the level of difficulty. The duration of the presentation was approximately 3.8 seconds.

The subject indicated which object would have arrived at the vertical line first (had all objects kept moving at a constant speed). The subjects heard a tone and then made a selection by pressing the appropriate number key on the keyboard. If the subject did not respond within approximately 4.5 seconds, the computer moved to the next trial. The subject received visual feedback of "RIGHT" or "WRONG," or "TOO SLOW."

There were 64 trials. Each digit was the correct answer an equal number of times, placed randomly in the 64 trials. The test required approximately 9 minutes to complete, at the fastest.

The extent to which one object would reach the wall sooner than the others was determined by the level of difficulty. There were eight levels of difficulty. When the subject answered two trials correctly in a row at a given level, the level of difficulty was increased; and when a subject answered incorrectly, the level of difficulty was decreased. Determination of the level of difficulty was interleaved, such that the previous odd trials determined the level of the next odd trial; and previous even trials determined the level of difficulty of the even trials. The interleaving allowed a split-half reliability to be calculated.

The following instructions were read to the subject before beginning testing: "You will see four objects, a ONE, a TWO, a THREE, and a FOUR, TRAVELLING across the screen from right to left. There will be a WALL on the left side of the screen. After travelling about halfway to the wall, the objects will DISAPPEAR. If the objects had CONTINUED travelling all the way to the wall, travelling at the SAME SPEED they HAD BEEN travelling, ONE of the objects would have arrived at the wall BEFORE the others. Your task is to decide WHICH object would have arrived at the wall FIRST. Here is how you will do it. When the objects disappear, a TONE will sound. After the tone sounds, you can indicate your answer by PRESSING the key marked ONE, TWO, THREE, or FOUR on the keyboard. For example, if you thought the TWO would have arrived at the wall first, you should press the key marked TWO. You have FOUR seconds to type in your answer. A MESSAGE on the screen will tell you whether your answer was RIGHT or WRONG."

Extrapolation

A curve was drawn on the screen. The curve began at the left of the screen, and was presented for 110, 140, or 170 pixels in the horizontal direction. The subject's task was to indicate where (vertically) the curve should end on the right of the screen. On the right of the screen was a vertical line, from (270, 7) to (270, 185).

There were three types of curves. All curves began at 1 on the x axis and ended at 270. One curve was a straight line ($y = a*x + b$). The line began at 20 to 160 (8 values) in increments of 20 on the y axis. The line ended on the same range of values.

One curve was a parabola ($y = a*x^2 + b*x + c$). The parabola began at 20, 40, 140, or 160 and ended at 30 to 150, in increments of 30. Therefore, there were 20 possible parabolas. The average ending height was 90, and the parabola reached its maximum (or minimum) at the far right.

One curve was a sine wave of the form $y = a*\sin(.05*x+b) + c$. A had 5 possible values, from 25 to 45, in increments of 5. The factor of .05 determined that the sine wave repeated every 125.7 pixels. B determined the phase the sine wave was in at the start and had 10 possible values, 0 to 90, in increments of 10. C determined the average height and had 5 possible values, from 60 to 140, in increments of 20.

The curve was 3 pixels wide. The error was the difference between the subjects answer and the top of the curve.

The subject indicated his or her response by moving the toggle switch of a joy stick. There was a pointer placed behind the wall. The pointer began each trial pointing at the value 100 and could point to any value from 14 to 184. The pointer was in the shape of an arrowhead, with the point indicating where on the far right wall the subject was placing the response. When the toggle was centered, the pointer remained stationary; but when the toggle was moved up, the pointer moved up; and when the toggle was moved down, the pointer moved down.

The subject indicated that he or she was satisfied with the response by pushing the button on the joy stick (or a key on the keyboard). There was a time-out period that varied from approximately 8 to 30 seconds, proportional to what percentage of the time the subject was adjusting his or her response. As the subject neared this time-out point, there was a beeping sound signalling that time was running out. If the subject did not press the pushbutton by the time-out point, the current location of the pointer was deemed to be the subject's response.

If the subject was within 5 pixels of the correct response, the message "WELL DONE" was flashed following the response.

There were six blocks of 18 trials each, for a total of 108 trials. Each block contains six curves of each type. Each block also contains six curves shown for each of the three possible distances. There were 3 practice trials, showing each of the curves.

The following instructions were read to the subject before beginning testing: "You will see a CURVE on the screen. There are THREE types of curves, a LINE, a SINE WAVE, and a PARABOLA. A SINE WAVE goes up and down; a PARABOLA is like the path a STONE takes when it is thrown. The curve will START at the far left of the screen, and END at the middle of the screen. If the curve had kept going, it would have hit somewhere on a WALL on the right side of the screen. You should decide WHERE the curve would have hit the WALL. Here is how you will do this. There will be a JOYSTICK in front of you. By pulling the joystick FORWARD or pushing it BACKWARD, you can move an ARROW up and down along the wall. You should MOVE the arrow so that it points to WHERE you think the CURVE would have hit the wall. When the arrow points to where you want, PUSH one of the BUTTONS on the joystick. You have a LIMITED amount of TIME to make your answer. When time is about to run out, the computer will begin BEEPING. If time RUNS OUT, the computer will ASSUME that the arrow s pointing to where you thing the curve will go. AFTER you have made your response, the computer will draw EXACTLY how the curve would have continued."

Intercept

The subject saw an object tracing a curve from left to right. There were three curves the object might take: a horizontal line, a sine wave, and a parabola. They all moved at a constant horizontal velocity. The straight line was of the form $y = A$, with A taking on four values, 85 to 115, in increments of 10. The parabola was of the form $y = -.0078*x^2 + 2.2*x + A$, with A again taking on four possible values of -15 to +15. The sine wave was of the form $y = 25*\sin(.05*x) + A$, with A taking on the values 85 to 115, in increments of 10.

The subject pushed a key on the keyboard, starting a second object moving upwards on the screen at the same rate that the target was moving horizontally. The subject's projectile was fired from a figure at the bottom of the screen. The lateral position of the figure varied from 200 to 245, in increments of 3.

If the two objects passed within 8 pixels of each other (on the vertical axis), the trial was deemed a hit: There was a beep, the UFO kind of broke into two pieces and left a trail of debris, and the subject received a congratulatory message ("WELL DONE"). If there was no hit, the objects continued along their paths.

There were three blocks, with 24 trials in each block for a total of 72 trials. There were 8 trials of each type of curve within a block; the order was determined randomly. There were three practice trials, one of each type of curve.

The following instructions were read to the subject before beginning testing: "Welcome to Intercept. This experiment is designed to test how well people can judge the speeds and paths of objects shown on the screen. It's something like a video game. On every trial you will see a 'UFO' which starts on the left of the screen and moves towards the right. The UFO will either move in a straight line or a curved path. Your task is to try to LAUNCH MISSILE to INTERCEPT the UFO. You do this by PRESSING the space bar at the correct time. Your missile will then move UPWARDS at a steady speed. You have to try to estimate the UFO's path and speed in order to intercept it. This task is very DIFFICULT. Try to come as CLOSE a possible. If you are very accurate the UFO will BEEP. Please ask any questions now."

Designs for the Static Tasks

Perceptual Comparisons

Subjects were shown pairs of irregularly shaped polygons that systematically varied in the number of points defining each figure (from 6 to 14, in 2-point increments) and in the relationship between the pair, match or mismatch. The mismatches systematically varied on a quasi-interval scale of increasing similarity (from 1, least similar, to 6, most similar). For each trial, latency and accuracy were recorded. Latency was expected to be a systematic function of the level of stimulus complexity (or points) and degree of stimulus similarity. Decision times should be longer for stimuli with more points and longer still for increasing similarity between the stimuli, with same trials taking the longest time.

There were five standard figures with either 6, 8, 10, 12 or 14 points. Each of the standards had six perturbed versions (d1 to d6) that increased in similarity to the standard (from d1, least, to d6, most similar). A single replication of the design consisted of 30 different trials (each standard, paired with each of its perturbed versions, 5 X 6) and 30 same trials (each standard paired with itself and repeated six times).

The 60 trials were partitioned into five blocks of 12 trials. Each block contained six same and six different pairs. For the same pairs within a block, each level of complexity (five levels; 6- to 14-point figures) was represented once, and one was duplicated. The duplicated "same" trial varied across blocks. For the 6 different trials within a block, there was one instance of each of the six levels of similarity (d1 to d6). Further, each of the complexity levels was represented by a different trial, with one of the complexity levels replicated. This replication was constrained in two ways: (1) The replicated complexity level could not be the one used when replicating the one same trial (i.e., if two 8-point same trials were in the block, then only one 8-point different trial could be included), and (2) the replicated different trial had to be more than one step removed (i.e., if an 8-point d1 was used, then an 8-point d2 could not be used). Under these constraints, each block had 3 trials at two of the complexity levels and 2 trials at the remaining three complexity levels. Within a block, the 12 trials were randomized with two constraints: (1) No more than 2 consecutive trials could involve the same complexity level, and (2) no more than 3 consecutive trials could require the same response (same or different). Each subject was presented with 300 trials; five replications of the five blocks. The five blocks were ordered in a 5 X 5 Latin Square, with each block ordering representing a different starting order. The starting order was randomly determined for each subject; the remaining four block orders were then worked through in sequence.

Mental Rotation

The subject was shown pairs of objects that varied in angular disparity and in the matching relationship between the pair, match or mismatch. The objects were polygons rotated from 0 to 180 degrees, in 20 degree increments. Mismatches involved rotation plus a mirror image reflection. The subject's task was to decide as rapidly and accurately as possible whether the stimuli match. Response latency and accuracy were recorded on each trial. Latency for both match and mismatch trials was expected to be a systematic function of angular disparity. To the extent that subjects make errors, these also should be related to degree of rotation.

Fourteen standard figures were used. Each was asymmetric, so that it could neither be transformed into itself by any reflection or depth plane rotation nor into any other figure in the stimulus set. For each standard, 20 unique problems were created, representing the factorial combination of degree of rotation (0, 20, 40, 60, 80, 100, 120, 140, 160, 180) and correspondence (match or mismatch). Subjects saw the 280 problems divided into four blocks of 70 problems each. The order of problem presentation was random except for the following constraints: (1) The same rotation value could not occur on consecutive trials; (2) at least 2 trials separated the appearance of the same figure; and (3) the 14 stimuli, 10 rotation values, and correspondence between stimuli appeared with approximately equal frequency within a block. Four block orders of the 70-trial blocks were created using a 4 X 4 Latin Square. A subject received one of the block orders, and there was random assignment over subjects.

Adding Detail

A six-pointed star was presented. Four to seven dots were added one at a time in response to the subject's keypress. That is, when the subject pressed a key, a new dot appeared, and the previous dot disappeared. Finally, a composite form was displayed, showing the star with the proper number of dots. The subject had to decide if the composite form displayed the dots in their correct positions. Presentation latencies, response latency, and accuracy were recorded for each trial and were expected to be a systematic function of the number of dots added.

The task design was based on a parsing of the six-pointed star into four quadrants. Dots were placed at either concave or convex vertices of the star and either inside or outside of the star. This provided 24 possible locations for a dot, 6 in each of the four quadrants.

There were 60 trials. The trials were arranged in four blocks of 15, and a Latin Square design was used so that each subject received one of four unique block orders. Trials were constructed such that one dot was placed in each quadrant, with any remaining dots distributed over the quadrants so that no quadrant had more than two dots.

The external-internal and concave-convex dimensions were balanced across trials, and there were an equal number of same and different trials. On different trials, one of the dots in a composite was incorrectly located. The incorrect dot was placed according to one of the following three transformations, which were equally represented: (1) Change an external dot to an internal dot or vice versa; (2) change the position of a dot from concave to convex, or vice versa, within the same quadrant; (3) combine transformations 1 and 2.

Integrating Details

Subjects viewed a display of three to six component shapes. The shapes had labeled edges indicating which edges were to be pieced together to form a complete shape. Following the display of the components, a complete shape was shown; and the subject was asked to decide if the complete shape represented the correct integration of the components. For each trial, presentation latency, response latency, and accuracy were recorded.

To facilitate the display of shapes on a computer screen, only straight lines were used to construct the component shapes. A "vocabulary" of 11 component shapes was constructed as follows:

1: A square, displayed so that the sides were parallel to the sides of the display screen,

2 and 3: Rectangles with one pair of sides equal to the side of the square and the other pair twice as long, displayed horizontally or vertically,

4 to 7: Isosceles triangles with a base twice as long as the square and height equal to the side of the square, displayed in one of four orientations such that the base was parallel to the sides of the display,

8 to 11: Isosceles right triangles with a base equal to the diagonal of the square and sides equal to the sides of the square, displayed in one of four orientations such that the hypotenuse would have coincided with the diagonal of the square.

The complete shapes were formed by adjoining the composite shapes so that there were no overlapping edges. From the numerous possible composite shapes, 48 were selected, 12 each composed of three, four, five, or six components. All of the complete shapes had three, four, five, or six sides, with at least one of each component type having each of the side totals. For presentation, the 48 shapes were arranged in six blocks of 8, with each side total equally represented in each block. The ordering of the shapes was randomized within each block, and a Latin Square design was used so that each subject received one of six unique block orders.

The components were arranged on the screen in the same orientation that they occupied in the complete shape. The horizontal and vertical positions of the components on the screen corresponded roughly to an "exploded" view of the composite shape, with moderate displacement so that the edges did not align exactly. The sides of the components that were to be pieced together to form the composite were labeled with capital letters on the outside of the shape, in the middle of the side. The same letter was used to label the sides of the two components to be joined.

There were an equal number of same and different trials. Different trials were formed by displacing one side of the composite shape, so that the shape had a distinctly different outline, while keeping the total number of sides in the composite the same.

Surface Development

The stimuli used in this task were an unfolded template with a marked base and a completed cube presented simultaneously. The items systematically manipulated the number of required folds to completion, the number of surfaces marked, the type of surface marking, and the degree of mismatch between the unfolded and folded versions. For each trial, response latency and accuracy were recorded. Latency was expected to be a linear function of the number of mental folds required for completion. Accuracy was expected to vary inversely with increases in the number of required folds, the number of marked surfaces, and the type of surface marking.

The full set of 192 items reflected five design factors: (1) number of folds and squares carried along, (2) number of shaded surfaces, (3) type of surface marking, (4) response type, and (5) degree of alternative mismatch. Five unfolded cubes or templates were used that allowed almost a complete crossing of these five factors.

Item complexity was defined as the minimum number of folds and surfaces carried along to determine the shading pattern on a completed cube. There were 16 trials at each of 12 levels of item complexity (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 15). Physical constraints on how a template can be folded limited the level of item difficulty a given template could represent. Three of the five templates were used at 5 levels of item complexity, and two were used at 6 levels of item complexity. Each level of item complexity was represented by two templates (except Level 7, which required three). Two of the templates were used on 36 trials, and three were used on 40 trials each.

The second design factor was the number of surfaces marked. This varied from two to three. At Levels 2 and 3 of item complexity, it was not possible to have three surfaces marked. However, for Levels 4 to 15 of item complexity, half of the trials have two surfaces marked; and the other half have three surfaces marked. The templates representing the given level of item complexity were selected because they produced both a two- and three-marked surface item at that level of item complexity using the same base square. The only exception to this was that at Level 7 of item complexity, only one template produced both a two and three-marked surface item at that level. In order to have at least two templates represented at each complexity level, the second set of Complexity Level 7 problems was obtained by matching two other templates: one of which produced a two-square-marked Complexity Level 7 the other produced a three-square-marked Complexity Level 7.

The third design factor was how the surfaces were marked. The surfaces were marked by a simple dot or ball on the surface. Sometimes the dot was placed in the center of the square or surface, and sometimes the dot was placed on one of the four edges of the surface. When only two surfaces were marked, there were two types of problems; the first type had one center-marked surface and one off-center marked surface, the second type had both surfaces off-centered marked. When three surfaces were marked, there were also two types of problems: one with two center-marked surfaces and one off-center-marked surface; the second item type had all markings off-center.

Half of the items required a same response and half required a different response. Response type was fully crossed with the other design factors at each level of item complexity. Nested within response type were two levels of the degree of mismatch. The two levels of different response reflected maximum mismatch (two or three nonmatching surfaces) and minimum mismatch (only one off-center-marked surface mismatched).

Overall, at a given level of item complexity there were 16 trials. These 16 trials were based on two templates. For item Complexity Levels 4 to 15, 8 of these trials were based on items with two marked surfaces; and 8 trials had three marked surfaces. Across all levels of item complexity, 8 trials had some surfaces centrally marked; and 8 trials had all marked surfaces off-center marked. At all levels of item complexity there were 8 trials that required a same response and 8 that required a different response, with 4 of the different trials being maximum mismatches and 4 being minimum mismatches. To avoid a confound between the type of mismatch and the number of marked surfaces at a given level of item complexity, only one of the templates was used to generate the four maximum mismatches; while the other generated the minimum mismatches. This was controlled over the entire item set such that a given template was used equally to generate the two mismatch conditions.

The 192 trials were subdivided into four blocks of 48 trials each. The blocking follows the design factors of the stimulus set such that there were equal numbers of trials at each level of item complexity; equal numbers of two- and three-marked surface items at each level (excluding item Complexity Levels 2 and 3); half with all surfaces off-center marked and half with some center-marked surfaces, half same and half different, with half of the different trials being maximum mismatches and half minimum mismatches. Ordering within a block was random but constrained such that a given template occurred no more often than once every 3 trials, a given level of item complexity did not occur consecutively, and no more than 3 consecutive trials required the identical response (same or different). A 4 X 4 Latin Square for block orders was employed. Since each subject received only one block order, block presentation order was randomly determined across subjects.

APPENDIX B

Correlations Between Selected Measures From All Tasks

Correlations Between Selected Measures From All Tasks

See Table 21 for an explanation of the acronyms used in this appendix. As in the main text, latency and error measures were reflected to reveal a positive manifold.

Table B-1

Correlations Between Dynamic Tasks

	PTHMEM	ARV1A	ARV1B	ARV2	ARV4	EXTRAP
ARV1A	.02					
ARV1B	.11	-.01				
ARV2	.16	.07	.06			
ARV4	.11	.05	.21	.27		
EXTRAP	.22	-.13	.24	.19	.09	
INTCPT	.12	.06	.15	.35	.23	.15

Table B-2

Correlations Between Static Tasks

	1	2	3	4	5	6	7	8	9	10	11
1. PCOMRL											
2. PCOMRA	-.52										
3. MRLAT	.43	-.21									
4. MRLAC	.08	.13	.04								
5. ADDPL	.20	-.31	.12	-.07							
6. ADDRL	.31	-.33	.27	-.09	.65						
7. ADDAC	-.01	.32	.00	.22	-.35	-.26					
8. INTGPL	.15	-.22	.18	-.13	.32	.33	-.19				
9. INTGDL	.32	-.22	.28	-.12	.29	.41	-.18	.45			
10. INTGAC	.08	.19	.20	.26	-.11	-.11	.37	-.26	-.01		
11. SDLAT	.28	-.08	.40	-.03	.18	.25	-.06	.42	.38	.07	
12. SDACC	-.01	.18	.21	.36	-.16	-.15	.40	-.20	.10	.50	.08

Table B-3

Correlations Between Paper-and-Pencil Tests

	SEX	DAT	RAVENS	IDPICT	SHMEM	VOCABL	SPAORT	PMA
DAT	.11							
RAVENS	.05	.54						
IDPICT	.01	.33	.19					
SHMEM	-.14	.22	.36	.14				
VOCABL	.10	.13	.14	.02	.04			
SPAORT	.26	.56	.51	.29	.23	.24		
PMA	.17	.46	.32	.42	.13	.02	.46	
3DROT	.14	.48	.38	.39	.25	.03	.31	.47

Table B-4

Correlations Between Dynamic Tasks and Static Tasks

	PTHMEM	ARV1A	ARV1B	ARV2	ARV4	EXTRAP	INTCPT
PCOMRL	.25	.02	.05	.09	.04	.11	.08
PCOMRA	-.12	.04	.05	.16	.04	.01	.05
MRLAT	.13	.01	.05	.30	.13	.18	.27
MRACC	.06	.14	.15	.18	-.03	.09	.05
ADDPL	.00	.08	-.03	-.06	.14	.07	-.07
ADDRL	.10	.02	.02	.06	.17	-.06	.01
ADDACC	.21	-.08	.10	.28	.03	.29	.05
INTGPL	.10	-.03	-.15	-.01	.03	-.07	.01
INTGDL	.05	-.06	-.04	.06	-.05	-.03	.00
INTGAC	.12	-.05	.21	.30	.12	.30	.20
SDLAT	-.05	.01	.04	.26	.09	-.03	.23
SDACC	.04	.05	.24	.31	.09	.34	.16

Table B-5

Correlations Between the Dynamic Tasks and the Paper-and-Pencil Tests

	PTHMEM	ARV1A	ARV1B	ARV2	ARV4	EXTRAP	INTCPT
SEX	.19	.07	.12	.30	.15	.12	.40
DAT	.20	-.03	.21	.29	.11	.30	.21
RAVENS	.22	-.02	.07	.28	.04	.33	.15
IDPICT	.12	.12	.25	.26	-.01	.24	.24
SHMEM	.04	.06	.07	.11	.07	.24	-.02
VOCABL	-.05	.05	.08	.09	.04	.05	.03
SPAORT	.19	.04	.15	.35	.10	.28	.22
PMA	.11	.05	.16	.30	.15	.17	.32
3DROT	.02	-.08	.14	.23	.05	.28	.24

Table B-6

Correlations Between the Static Tasks and the Paper-and-Pencil Tests

	SEX	DAT	RAVENS	IDPICT	SHMEM	VOCABL	SPAORT	PMA	3DROT
PCOMRL	.00	.23	.08	.42	.09	.07	.24	.28	.25
PCOMRA	.01	.07	.22	-.09	.13	-.02	.05	.05	-.12
MRLAT	.20	.43	.24	.32	.14	-.13	.32	.47	.39
MRACC	.00	.23	.29	.06	.20	.17	.28	.19	.14
ADDPL	.02	-.01	-.11	.13	-.06	-.06	.01	.07	-.12
ADDRL	.06	.05	-.06	.19	-.16	-.06	.05	.19	.04
ADDACC	-.10	.31	.38	.14	.33	.08	.27	.17	.24
INTGPL	.01	-.04	-.15	.11	-.08	-.10	-.11	.02	.06
INTGDL	.10	.18	.03	.20	-.09	.04	.03	.18	.12
INTGAC	.17	.57	.57	.16	.16	.18	.41	.34	.30
SDLAT	.07	.26	.10	.39	-.01	-.08	.23	.29	.27
SDACC	.00	.49	.45	.09	.30	.14	.28	.26	.33

DISTRIBUTION LIST

Assistant Secretary of Defense (Force Management and Personnel)
Deputy Under Secretary of Defense for Research and Engineering (Research and Advanced Technology)
Military Assistant for Training and Personnel Technology (OUSD) (R&AT)
Deputy Assistant Secretary of the Navy (Manpower)
Assistant CNO for Long-Range Planning and MANTRAPERS (OP-01B6)
Assistant for MPT Research and Development and Studies (OP-01B7)
Assistant for Manpower, Personnel, and Training (OP-983D)
Commander, Space and Naval Warfare Systems Command (SPAWAR-00)
Commanding Officer, Naval Training Systems Center
Applications Requirements Branch (Code 6022)
Cognitive and Decision Science (OCNR-1142CS)
Technology Area Manager, Office of Naval Technology (Code 222)
Office of Naval Research, Detachment Pasadena
Office of Naval Research, London
Chief of Naval Education and Training (Code 00A)
Chief of Naval Technical Training (Code 00) (2)
Commandant of the Marine Corps (MPI-20)
Education Center, Quantico, VA (MCDEC)
Commander, U.S. ARI, Behavioral and Social Sciences, Alexandria, VA (PERI-POT-I)
Technical Director, U.S. ARI, Behavioral and Social Sciences, Alexandria, VA (PERI-ZT)
HQ USMEPCOM/MEPCPAT-A, North Chicago, IL
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base, TX
Air Force Resources Laboratory, Brooks Air Force Base, TX, Manpower and Personnel Division (AFHRL/MO)
Air Force Human Resources Laboratory, Brooks Air Force Base, TX, Scientific and Technical Information (STINFO) Office
Air Force Human Resources Laboratory, Brooks Air Force Base, TX, TSRL/Technical Library (FL 2870)
Commander, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base, OH, Logistics and Human Factors Division (AFHRL/LRS-TDC)
Commander, Air Force Human Resources Laboratory, Lowry Air Force Base, CO (AFHRL/ID)
Commanding Officer, U.S. Coast Guard Research and Development Center, Avery Point, Groton, CT
Superintendent, Naval Postgraduate School
Director of Research, U.S. Naval Academy
Center for Naval Analyses
Defense Technical Information Center (DTIC) (2)

DEPARTMENT OF THE NAVY

NAVY PERSONNEL RESEARCH AND
DEVELOPMENT CENTER
(CODE___)
SAN DIEGO, CA 92152-6800

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

U.S. MAIL

Postage and Fees Paid
Department of the Navy
DOD-316

END

10-87

DTIC