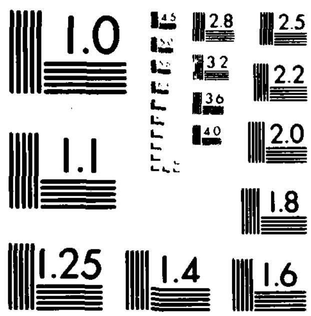
AD-A085 769   MINNESOTA UNIV MINNEAPOLIS DEPT OF PSYCHOLOGY   F/G 5/10
INTERACTIVE COMPUTER ADMINISTRATION OF A SPATIAL REASONING TEST--ETC(U)
APR 80   A T CHURCH, D J WEISS   N00014-76-C-0243
UNCLASSIFIED   RR-80-2   NL

END



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

(12) LEVEL II

INTERACTIVE COMPUTER ADMINISTRATION OF A SPATIAL REASONING TEST

Austin T. Church
and
David J. Weiss

ADA085769

Research Report 80-2
April 1980

Computerized Adaptive Testing Laboratory
Psychometric Methods Program
Department of Psychology
University of Minnesota
Minneapolis, MN 55455

This research was supported by funds from the Navy Personnel Research and Development Center, and the Office of Naval Research, and monitored by the Office of Naval Research.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: (9) Research Report 80-2
2. GOVT ACCESSION NO.: AD-A085 719
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): (6) Interactive Computer Administration of a Spatial Reasoning Test
5. TYPE OF REPORT & PERIOD COVERED: Technical Report
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): (10) Austin T. Church [redacted] David J. Weiss
8. CONTRACT OR GRANT NUMBER(s): (15) N00014-76-C-0243
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: P.E.:6115N PROJ.: RR042-04; T.A.:RR042-04-01; W.U.:NR150-382
11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs, Office of Naval Research, Arlington, Virginia 22217
12. REPORT DATE: (11) Apr[redacted]80
13. NUMBER OF PAGES: 72
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (12) 83 (14) RR-80-2
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

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

(16) RR04204 (17) RR0420401

18. SUPPLEMENTARY NOTES

This research was supported by funds from the Navy Personnel Research and Development Center, and the Office of Naval Research, and monitored by the Office of Naval Research.

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

adaptive testing, tailored testing, spatial ability, spatial reasoning, problem-solving, problem difficulty, perceived difficulty, response latency, short-term memory, computerized testing, computer-administered testing, psychological effects

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

This report describes a pilot study on the development and administration of a test using a spatial reasoning problem, the 15-puzzle. The test utilized the on-line capabilities of a real-time computer (1) to record an examinee's progress on each problem through a sequence of problem-solving "moves" and (2) to collect additional on-line data that might be of relevance to the evaluation of examinee performance (e.g., number of illegal and repeated moves, response latency trends). The examinees, 61 students in an introductory psy-

chology class, were required to type a sequence of moves that would bring one 4 × 4 array of scrambled numbers (start configuration) into agreement with a second 4 × 4 array (goal configuration), using as few moves as possible. Data analyses emphasized the comparison of several methods of indexing problem difficulty, methods of scoring individual performance, and the relationship between response latency data, performance, and problem-solving strategy.

Subjective ratings of the perceived difficulty of replications of the 15-puzzle were obtained from a separate student sample to investigate (1) the subjective dimensions used by students in evaluating the difficulty of this problem type, (2) how accurately the actual performance difficulty of these problems could be evaluated by students, and (3) whether there were reliable individual differences in difficulty perceptions related to actual performance differences.

Results of the study suggested that four performance indices might be useful in indexing problem difficulty: (1) mean number of moves in the sample, (2) proportion of students solving the problem, (3) proportion of students solving the problem in the optimal number of moves, and (4) a Special Difficulty Index, defined as the sample mean number of moves divided by the minimum number of moves required. Four alternative methods of scoring total test performance and two methods of scoring individual problem performance were studied. The scores that took into account differential numbers of moves between the optimal and maximum number allowed were related somewhat more to performance ratings obtained from independent judges.

Examination of problem performance indices, the Special Difficulty Index, and students' perceptions of the difficulty of the test problems indicated that most of the problems were too easy for most students. However, the possibility of obtaining a more discriminating subset of problems was suggested by item-total score correlations obtained for each problem. The data suggested that better consistency might be obtained using problems of similar difficulty levels, and it was hypothesized that an adaptive test tailoring problems to the ability level of each student would increase the reliability of measurement.

Mean initial and total "move" latencies for each problem were strongly related to some of the performance indices of problem difficulty. At the level of individual performance, only total latency or problem solution time was related to problem performance. Latency data appeared to confound differences in the ability to visualize a sequence of moves and differences in students' work styles. Strong evidence for these work styles was found in student consistency of initial, average, and total response latency measures across all problems.

Perceived difficulty ratings showed reliable individual differences in the level and variability of difficulty perceptions. The data suggested that the individual differences found were related to individual differences in ability to visualize and to maintain a sequence of moves in short-term memory. It was concluded that an adequate selection of problem replications should be able to tap these differences, resulting in reliable solution performance differences.

Improvements in problem selection and design were suggested by the data in this study. Future tests of this type should consist of fewer but more difficult problems, particularly problems not permitting reactive, impulsive solutions. This type of test would seem especially appropriate for adaptive administration: (1) scores on problems tailored to the individual's ability would likely be more highly related to each other, resulting in more highly reliable total scores; (2) the motivational aspects of the tests, which seem more taxing and potentially frustrating than conventional item formats, would likely be improved, and (3) for most testees equally precise measurements could be obtained in shorter periods of time than with conventional test administration.

CONTENTS

Acknowledgments

The authors thank Nancy Strandmark and Richard Bale for their assistance in the data collection and analysis.

Technical Editor: Barbara Leslie Camm

A

Interactive Computer Administration of a Spatial Reasoning Test

Most research on computer-administered testing has emphasized the ability of the computer to adapt item difficulties to the ability level of examinees. Such computerized adaptive tests have been shown to provide more equiprecise measurement across all trait levels (e.g., Vale, 1975; Vale & Weiss, 1975), to provide generally higher test-retest stabilities than conventional tests (e.g., Betz & Weiss, 1973, 1975), and to result in tests of fewer items while achieving the same or higher levels of measurement accuracy (Weiss & Betz, 1973). In addition, research has indicated that immediate knowledge of results administered to testees after each item in computer-administered tests results in enhanced performance (Betz & Weiss, 1976a) and favorable psychological effects for examinees (Betz & Weiss, 1976b). Research with computer administration of a concept attainment task (Johnson & Baker, 1972) indicated that improved standardization could also be obtained with computer test administration; and the results of Johnson and Mihal (1973) and Pine, Church, Gialluca, and Weiss (1979) indicated that differences in mean peformance of racial groups might be reduced or eliminated with computer-administered testing.

Almost all of the research on computer-administered testing has measured intellectual abilities and utilized item types that are conveniently measured by conventional paper-and-pencil tests as well. However, computers would seem to be especially useful in measuring various perceptual, memory, and problem-solving abilities that utilize the computer's capabilities to present novel item formats, modifying item presentation over time in response to the examinee's performance and allowing the computer to interact with the student while working on a task. It is of interest to determine whether the advantages previously found for computer-administered tests, particularly in an adaptive mode, can be extended to tests of new abilities that make fuller use of the unique capabilities of the interactive computer.

Although the use of computers to control the presentation of visual stimuli on a cathode-ray-tube (CRT) is fairly common in psychological research, most of this research has been concerned with the discovery of processes of attention, memory, and perception that apply to all individuals. Recently, however, investigators have begun to explore the potential of computer-administered tests for measuring individual differences in various cognitive abilities. For example, Cory (1977; Cory, Rimland, & Bryson, 1977) has developed tests for five abilities--short-term memory, perceptual speed, perceptual closure, movement detection, and dealing with concepts/ information--and compared scores on these tests to conventional paper-and-pencil tests of comparable abilities. The conclusion was that these tests provided measures of attributes that are different from those measured by paper-and-pencil tests. For example, a "sequential reasoning dimension," which did not appear in the paper-and-pencil tests, was identified in the computerized tests. Computer test administration is also being increasingly used by psychologists interested in measuring individual differences in various basic information processing abilities (e.g., Chiang & Atkinson, 1976; Hunt, Lunneborg, & Lewis, 1975; Rose, 1978).

A common characteristic of such new ability tests is that traditional psychometric indices of individual performance (such as number-correct scores) and item characteristics (such as item difficulty and item discrimination) may no longer be meaningful. To measure individual differences in examinee performance, researchers have used scores derived from reaction time data; slope and intercept parameters relating reaction time to memory set size (Sternberg, 1969); component scores on various stages or subprocesses derived from hypothesized models (e.g., Clark & Chase, 1972); and parameter scores ($\underline{D}'$, beta) derived from signal detection theory. Some, but not all, researchers using such measures of individual differences have attempted to demonstrate the psychometric characteristics (e.g., reliability) of these new performance indices. Such a demonstration is necessary, however, for each new score derived from new types of ability tests before the validity and utility of the scores can be investigated.

Purpose

This report describes a pilot study reporting the development and administration of a spatial reasoning problem, the 15-puzzle, which utilized the on-line capabilities of a real-time computer to record a testee's progress on each problem throughout a sequence of "moves" and to collect additional on-line data that might be of relevance to the evaluation of testee performance. Although spatial ability has been shown to be an important special ability predictive of some job criteria (for a summary of predictive validities for various occupational areas between 1920 and 1971, see Ghiselli, 1973), it was also hoped that this problem type and others to be developed would be able to tap generalized problem-solving and reasoning abilities.

The 15-puzzle problem used in this study involved presentation of the numbers 1 to 15 in a 4 x 4 matrix of scrambled numbers and in a target matrix with the numbers in another configuration. The testee was required to move the numbers in the first configuration, one number at a time, to match the second configuration. This problem type was chosen because it seemed to tap abilities important in problem-solving situations, especially in the spatial domain, while providing the following additional advantages:

1. Utilization of the unique capabilities of interactive computers.
2. The existence of a well-defined optimal solution against which to evaluate a student's performance.
3. The ease of generating large numbers of replications of varying and relatively controllable difficulty levels.

If the advantages of computerized adaptive testing are to be applied to tests of this type, precise indices of individual performance and problem difficulty must be devised. Thus, an important emphasis in this study was on a comparison of alternative methods for quantifying student performance and a comparison of alternative indices of problem difficulty for the 15-puzzle spatial reasoning problem. For example, the number of moves a student requires to solve replications of the 15-puzzle may not be an adequate index of problem performance where the minimum number of moves for various problems differs. Some of the questions studied were, Is the minimum number of moves to solution a meaningful index of problem difficulty, or do other physical aspects of the puzzle configuration influence problem difficulty as well? Can response latencies be used to quantify difficulty and/or individual performance? In addition, to determine whether or not the 15-puzzle task could be used to successfully measure problem solving in the spatial domain, the reliability of individual performance scores across problems of similar and varying difficulty levels was examined.

One further advantage of the problem type studied here may be its interactive game format, which may prove to be more motivating to examinees than the usual separate item format. In addition, the provision of knowledge of results may be a built-in feature of these problems, since the students can tell when they have reached a solution. On the other hand, the need for perseverance and the possibly greater potential for frustration and anxiety with this type of problem must also be considered. Thus, motivational data were collected and examined in this study in an attempt to draw some preliminary conclusions about the psychological effects of working on such problems.

To a large degree, the psychological effects of problems of this type on examinees will depend on the perceived difficulty of replications of the problems. It would seem that problems of this type that are inappropriate for the student's ability level may be even more discouraging than the typical conventional test item because the student cannot merely guess and continue with the next item. In problems of this type, guessing becomes not a response bias to be eliminated but a trial-and-error strategy on the part of the examinee. Thus, eventual adaptation of problems to the student's ability level may be especially important for making the testing experience reasonably pleasant and nonfrustrating.

However, whether an adaptive presentation of problems can actually equalize the psychological effects of such a test will depend largely on whether students can accurately perceive the difficulties of the items administered (Prestwood & Weiss, 1977). Even though some previous research has found agreement between perceived and objective indices of item difficulty (e.g., Bratfish, Dornic, & Borg, 1972; Munz & Jacobs, 1971; Prestwood & Weiss, 1977), it would seem necessary to answer this question anew when item or problem types differ significantly. The present study, therefore, reports some preliminary data relating to the similarity of objective and perceived indices of problem difficulty for replications of the 15-puzzle.

METHOD

Computer-Administered Problems

Tests

Problem description. A series of spatial 15-puzzles, each a reasoning problem, were administered to students on an interactive cathode-ray-tube (CRT) display terminal. The sequence of problem presentations and the simultaneous collection of performance data were controlled by a computer program written for a Hewlett-Packard real-time minicomputer.

Figure 1 shows a sample of the display presented on the CRT screen while the student worked on each problem. As Figure 1 shows, the student was instructed to type a three-character "move" on the terminal keyboard specifying which number in the left pattern he or she wished to move left, right, up, or down one square in an attempt to eventually bring the configuration of numbers in the left pattern into agreement with the pattern of numbers on the right.

Figure 1
Sample 15-Puzzle Problem

Make your "moves" in this pattern

10	9	3	7
4	8		6
12	5	2	14
1	11	15	13

Try to match this pattern

10	2	9	7
12	8		6
5	4	3	14
1	11	15	13

Enter your move by typing three characters and the "RETURN" key.

The first two characters should be the number you want to move. If the number has only one digit, type one space and then the one digit number.

The third character should be:
- L - if you want to move the number one square to the left.
- R - if you want to move the number one square to the right.
- U - if you want to move the number up one square.
- D - if you want to move the number down one square.

After each three-character move was typed, the computer processed the move for legality. If the move was legal, the pattern on the left was updated immediately using a cursor addressing system, which allowed specified screen locations to be manipulated without rewriting the entire screen. If the three-character move was illegal, an explanatory error message was displayed, and in some cases the student was instructed to notify the test proctor for assistance. The testing program detected illegal moves of both a syntactical (e.g., typing errors) and a logical (e.g., trying to move a number into an already occupied square or beyond the outer edge of the pattern) nature. Appendix A contains a complete list of diagnostic error messages utilized by the testing program.

Performance data. While the student worked on the problem, the following data were collected on-line by the computer:

1. Whether the problem was solved or not, i.e., whether the student was able to type a sequence of moves that would make the configuration on the left match the configuration on the right.

2. The number of moves required for solution.
3. The number of illegal moves, including impossible moves of both a syntactical and a configural nature.
4. The number of repeated moves, i.e., how many times the student "backed up," or reversed a possibly incorrect sequence of moves to return to an earlier pattern configuration.
5. Response latencies, i.e., the time in seconds required for each move.
6. The actual sequence of moves utilized.

The performance data were collected for possible use in drawing inferences about several aspects of spatial problem-solving ability. For example, the number of illegal moves, as well as the initial response latencies, might index the student's initial ability to define and to clarify the task situation. The sensitivity of students to the task information provided (in this case, the continually updated left pattern and its relationship to the right pattern) and their ability to plan a sequence of moves might be indexed by the number of nonoptimal moves, the number of repeated moves, and the total number of moves required. A student's inability to recenter (Sweeny, 1953; Wertheimer, 1959) or the presence of a debilitating set might be inferred from a persistent sequence of moves that did not bring the start pattern closer to the goal pattern.

The pattern of response latencies as the student approached the solution might also be useful information in making inferences about a student's problem-solving strategy. For example, in the initial stages of the problem, a planning-ahead strategy might be inferred from longer initial response latencies, and a more impulsive, reactive strategy or problem-solving style would be associated with shorter latencies. If the student was sensitive to the relationship between the two stimulus patterns, a shortening of the response latencies might be expected as the left (start) pattern approached the right (goal) pattern (Hayes, 1965).

Individual differences in the ability to visualize or to maintain sequences of moves of varying lengths in short-term memory might also be reflected in the patterns of response latencies. For example, an individual with a greater ability to maintain a sequence of moves in short-term memory might need longer pauses or study points only once every six or seven moves, as opposed to every three or four moves. Isolation and interpretation of such differences may be difficult, however, since momentary differences in short-term memory capacity may also reflect differences in the allocation of limited cognitive resources (Norman, 1976).

Test Administration

Sixty-one students in an introductory psychology class took the problem-solving test. Of these, tests for five students had to be discarded because of computer problems. After being logged onto the CRT by a test monitor, the student was presented a series of instructional screens by the computer. The text of each instruction screen is in Appendix E. The presentation of instruction screens was student paced, with the student pressing the "SPACE BAR" and "RETURN" key on the terminal keyboard to proceed to the next instruction screen.

As Appendix B shows, the instructions first told the student how to utilize important keyboard characters, such as the "RETURN" key, to enter responses. Next, after describing the 15-puzzle task and providing instructions on entering a three-character move, the instructions told the student how to correct a mistyped move before transmitting the move to the computer. Biographical information, including name, student identification number, age, sex, year in school, major field of study, race, and grade-point average, was then requested from each student. The final instructional screen (see Screen 16 in Appendix B) was intended to standardize the desired motivational set for each student. The student was then presented with a practice problem. This practice problem (Problem 1) was very simple, requiring only three straightforward moves, and was used to allow students to clarify questions and to gain confidence in entering moves under nontesting conditions.

Following the practice problem, students were presented a maximum of 12 problems (Problems 2 to 13). These problems varied in difficulty, which was initially indexed by the minimum number of moves required for solution (solution path length) using a solution algorithm provided by Nilsson (1971). Each of the 12 problems consisted of one problem requiring 4 and 6 moves and two problems for each of the following solution path lengths: 8, 10, 12, 14, and 16. The 13 problems used, along with their solution path length and other physical problem characteristics, are in Appendix C.

Data for all students were not obtained for all the problems for a variety of reasons. Since the students differed in both solution efficiency and in the amount of time they had available to participate in the study, not all students completed all the problems. In addition, after about half the students (the first 33) had completed the tests, it appeared that a test consisting of 12 problems was somewhat too long and that some students did not have enough time in the experimental hour to finish the longer problems. For this reason, two of the easiest problems (Problems 2 and 3), which everyone seemed to be solving in the minimum number of moves, were eliminated to make the test shorter. Finally, in a few cases, data for a single problem were lost for a student due to computer problems.

There was no fixed time limit for each problem. However, in order to prevent a student from spending too much time on a single problem to the exclusion of others, a message advising the student to notify the test proctor was displayed on the terminal screen after the student had been working on a problem for what was thought to be an unduly long time. The maximum time allowed for each problem was a multiplicative function of the minimum number of moves required, up to a maximum of 15 minutes. For example, about 4 minutes were allowed for a problem requiring 3 moves, about 10 minutes for a problem requiring 8 moves, and about 15 minutes for problems requiring 12 to 16 moves. The proctor then had the option of advancing the student to the next problem or resetting the problem timer to allow the student to continue work on that problem. Students were encouraged to discontinue work on a problem unless they felt confident they were near solution and needed only a little more time.

Similarly, the student was stopped when he or she had taken the maximum number of moves allowable for a problem. The maximum number of moves allowed by the computer was also a function of the minimum number of moves (solution path length) required to solve the problem. The maximum number of moves was defined as the solution path length times 3.5; if the maximum number of moves was greater than 28, the maximum move limit was set equal to 28. This maximum was intended to terminate work on a problem the student appeared unable to solve so that he/she would proceed to subsequent problems. The number of moves it would take to recover from nonoptimal moves was taken into consideration in specifying this initial maximum move limit. It was realized, however, that this maximum limit might have to be adjusted once actual performance data were obtained.

The maximum number of moves allowed was increased for about half the students to determine if students could reach solution if they were given more moves. Thirty-three students were limited to 28 moves for the longest problems and the remainder were allowed 43 moves. The larger move limit seemed to allow more students to reach solutions for some of the longer problems.

A student was permitted to voluntarily choose to terminate a problem before the solution was reached by asking the test proctor to advance him/ her to the next problem. In the few instances where this situation arose, students were encouraged to continue work on a problem unless the time limit message had already appeared.

When the student successfully completed a problem by matching the start and goal pattern: the computer displayed the message:

Good. You have succeeded in matching the two patterns. Press the "SPACE" bar and "RETURN" to start the next problem.

Test Reaction Data

Upon completing all the test problems, a message thanked the student for his/her participation. Students then completed a paper-and-pencil questionnaire providing information on prior experience, difficulty perceptions, and other motivational questions that could be used to evaluate student reactions to this type of test.

Since a general measure of spatial reasoning ability was sought, individual differences in test performance should not be accounted for by specific prior experience with this type of puzzle. Therefore, the first question asked the student how often he/she had worked on this kind of puzzle in the past. In order to evaluate the clarity of the instructions for this new type of test item, the second question asked students how much difficulty they had in understanding the instructions. Because this was the first time this problem type had been used on this student population, it was not known before data collection how difficult puzzle replications would have to be to challenge the students. Thus, the third question obtained information on how difficult the students thought the puzzles in the test were.

It was felt that the student's motivation level during testing would be especially important for performance on problems of this type, which require more concentration and within-problem perseverance than more typical single item formats. Consequently, Question 4 asked students how hard they tried to solve each puzzle in the optimal number of moves, and Question 5 asked whether the length of the test affected their motivation. Students indicated how nervous or uncomfortable they were while working on the puzzles in Question 6. Overall evaluations of how well they thought they had performed and how well they enjoyed working on the puzzles were provided by students in Questions 7 and 8, respectively. Any further comments the students had were elicited by Question 9. Since all Puzzle Reaction questions referred to different content, no scores were derived across items.

Data Analysis

Indices of problem difficulty. Data collected for each problem were used to describe problem difficulty in several ways. For each of the 13 problems (12 problems plus 1 practice problem), the frequency and proportion of students requiring various numbers of moves to solve or to fail to solve the problem was calculated. The following were also computed for each of the 13 problems as potential indices of problem difficulty:

1. The mean number of moves taken. This was the average number of legal moves used by the student to solve the problem or the number of moves at which the problem was terminated due to using too many moves or too much time. Since the move limit was extended from 28 to 43 for about one-third of the students, the mean number of moves was slightly lower for the longer puzzles than it would have been had all students been allowed the larger maximum number of moves.
2. The proportion of students solving the problem within the original maximum number of moves (i.e., for the longer puzzles, 28 moves.)
3. The proportion of students solving the problem in the minimum or optimal number of moves.
4. The mean number of illegal moves.
5. The mean number of repeated moves.

In addition, for each problem a Special Difficulty Index was computed, defined as the mean number of moves used, divided by the minimum number of moves required (solution path length). This index was designed to provide a possible difficulty index that was corrected for differences in minimum solution path lengths for each problem. For example, a problem requiring 16 moves may not be more difficult (in the sense that nearly everyone could solve it in the minimum number of moves) than a problem requiring only 10 moves.

A possible advantage of the relatively formal nature of the 15-puzzle is the availability of potentially objective physical problem characteristics, which could function as potential indices of task difficulty. One such index, solution path length (i.e., the minimum number of moves required for solution), has already been mentioned. Several other indices relating the start pattern to the goal pattern were computed to determine if they related empirically to the actual difficulty in solving each problem as indexed by student performance. If such a relationship was found, these physical indices could be used in selecting problem replications for inclusion in a test on the basis of their predicted difficulty.

The following physical problem characteristics of each pair of patterns were considered as potential difficulty indices:

1. Path length: the minimum number of moves required to solve the problem.
2. The number of squares not matching in the start and goal patterns at the start of the problem (maximum = 16).
3. The number of rows disrupted or not matching in the two patterns (maximum = 4).
4. The number of columns disrupted or not matching in the two patterns (maximum = 4).
5. Euclidean distance function: the sum of the distances of each number's position in the start pattern from its position in the goal pattern using the Pythagorean theorem (i.e., diagonal distances allowable).
6. City-block distance function: the sum of the distances of each number's position in the start pattern from its position in the goal pattern with only vertical and horizontal (not diagonal) displacements calculated.

Appendix C shows each of these physical problem characteristics for each of the 13 problems.

<u>Assessment of student performance</u>. Deriving scores for a student on a single problem, and on this type of test as a whole, is complicated by several factors. For example, some students were not able to work on the test as long as others; some students naturally worked faster than others; and in a few cases, data on isolated problems was lost because of computer failure. In addition, half the students did not work on Problems 2 and 3, since these were eliminated to shorten the test.

As a result, scoring a student's peformance merely by the number of problems solved was not only undesirable from a theoretical point of view but it was also impractical due to the above confounding factors. For this reason, and also from the point of view of using these problems in future adaptive testing, it was desirable to develop scoring methods that did not depend on the particular problem replications on which the student worked. This suggested using such measures as the proportion of problems worked on by the student that he or she was able to solve or the proportion of problems attempted that the student solved in the optimal number of moves. However, these measures do not take into account the differential difficulty of different problems or individual differences in the number of moves used between the optimal and maximum allowed number. Using the number of moves a student made on a problem would not take into account the differential solution path lengths and the difficulty of problems. Potential measures that would take into account the difficulty of various problems, such as the mean difficulty of problems solved or the highest difficulty problem solved in the optimal number of moves, would not be comparable for students who did not receive problems of the same difficulty level.

Taking into consideration all these problems, two methods of scoring student performance on individual problems were devised:

1. Score 1 = $\frac{\text{the number of moves the student used}}{\text{the minimum number of moves actually required}}$

   For example, if a student took 15 moves to solve Problem 6, which required 10 moves, his/her score was 15/10 = 1.5. Since a perfect score would be 1.0, this student required 50% more moves than were necessary. Note that although this score corrected for different solution path lengths of various problems, it did not take into account the difficulty of the problem as indexed by the total group's performance on the problem.

2. Score 2. This score was Score 1 adjusted by the Special Difficulty Index. Thus,

   Score 2 = (Score 1)/(Special Difficulty Index)

   This score reduces to

   Score 2 = $\frac{\text{the number of moves the student used}}{\text{the mean number of moves required by the total group}}$

   Thus, if Score 2 = 1.0, the student's performance was equal to the group average. If Score 2 was less than 1.0, the student solved the problem in fewer moves than the average student; conversely, if Score 2 was greater than 1.0, the student solved the problem in more moves than the average student.

To determine whether these specially defined scores were any more meaningful than more direct scores, such as the proportion of problems solved, the relationships between the following four scores for the test as a whole were examined:

1. PROPS = the proportion of problems that the student attempted (worked on) and solved within the maximum number of moves (28).
2. PROPM = the proportion of problems that the student attempted and solved in the minimum (optimal) number of moves.
3. Total 1 = the average Score 1 obtained on the problems the student attempted.
4. Total 2 = the average Score 2 obtained on the problems the student attempted.

It was hypothesized that the Total 2 score would prove to be the most meaningful score, since it took into account both the solution path length and the difficulty of the problems the student attempted and did not depend on the number of problems attempted. By adjusting for problem difficulty, a student was penalized more by Total 2 for less than optimal solutions on easier problems than on more difficult problems.

<u>Consistency of performance across problems</u>. An important question for determining the usefulness of this problem type in assessing spatial problem-solving ability was whether reliable individual differences on various performance criteria could be identified across prob-

lem replications of similar and varying difficulty levels. To examine this question, the consistency of the various performance scores was examined across all 13 problems using Pearson product-moment correlations. Since both of the individual problem scores (Score 1, Score 2) were linear transformations of the optimal number of moves, the consistency of these scores across problems in terms of Pearson product-moment correlations would be the same as the stability of the number of moves used. Thus, the stability of the following performance indices were examined:

1. The total number of legal moves used for each problem,
2. The number of illegal moves, and
3. The number of repeated moves.

The relationship between individual problem scores and total scores on the problem set as a whole was investigated by examining the correlations between individual problem scores (Score 1, Score 2) and total test scores (PROPS, PROPM, Total 1, Total 2) with and without the particular problem being excluded from the total score. In addition, the relationships between the total number of legal moves used (or, equivalently, Score 1 and Score 2), the number of illegal moves, and the number of repeated moves for each problem were examined by computing the Pearson product-moment correlations between pairs of these performance indices across students for all pairings of problem replications.

Response latencies. During testing the time in seconds taken by a student for every move was recorded by the computer. This allowed latency trends across moves to be plotted and studied for each problem. Three indices were used to quantitatively characterize a student's response latencies for a problem:

1. Initial move latency, i.e., how long the student studied the initial problem configuration before making the first move;
2. The average move latency, i.e., the average time taken for a move across the particular problem; and
3. Total problem latency, i.e., the total time in seconds taken by the student on a particular problem.

In order to compare the time taken on various problems with the problem difficulty as indexed by various performance measures, the mean of the above three latency measures was computed across all subjects for each problem.

Although the tendency for various performance measures (e.g., the number of moves needed) to correlate across problems indexes the reliability of problem-solving performance, the tendency for a student's response latencies to show consistency across problems may indicate a cognitive style, e.g., reflectivity versus impulsiveness (Kagan, 1965; Kagan et al., 1964) or a strategy of planning-ahead versus trial and error or impulsive responding. To study this possibility, the consistency of the initial, average, and total response latency measures across problems was examined using Pearson product-moment correlations. For example, by correlating the initial move latency across students for each pair of problems, it could be determined whether some students consistently studied each problem for longer or shorter times than other students. Similarly, by correlating the total problem latency over students for each possible problem pair, it could be determined whether the same students who took longer or shorter times to solve one problem also did so on the other problems.

It was also of interest to examine the response latency trends as the student progressed throughout each problem. Such trends may indicate the degree of initial planning, the number of moves a student made between study points, and the point at which the sequence of moves to solution had been detected. For this purpose latency graphs for individual students showing the response latency for each move from start to solution were plotted and inspected visually. Latency plots were examined for students who had performed well on the test and those who had performed poorly and for problems solved and problems unsolved, in order to detect any systematic differences in latency trends.

Relationship between performance and response latencies. In order to determine if any relationship existed between students' performance on the problem and the way they allocated their time on each problem, Pearson product-moment correlations were computed for each problem between the initial, average, and total move latencies and the number of moves each student used. In addition, correlations were also computed between total test score, which better indexed the student's performance on the test as a whole, and the initial, average, and total latencies for each problem. For these correlations the total test scores used were Total 2 and a mean judges' performance rating, described below.

Judges' ratings of performance. Because reliable external criteria against which the student performance scores could be validated were not available, each student's performance on each problem was studied independently by three judges and each student's overall test performance was rated on a 10-point scale, with 5 being anchored to average or mean performance, considering the sample as a whole. The mean of the ratings of the three judges (MRATE) was used as another index of student total test performance.

Since the judges were familiar with the difficulty of each problem and could carefully examine the student's performance on each problem, it was felt that these ratings would provide a more complete assessment and rank ordering of student performance. Although less subjective, the performance scoring methods described above were not equally able to take into account all the information that the judges could in their ratings. Thus, one way to compare the adequacy or refinement of the various scoring methods was to compare the rank ordering of students by each method with the rank ordering assigned by the judges' ratings. This was done using Spearman rank-order correlation coefficients.

To determine how well independent judges could agree on the ratings of student performance, interrater reliability as estimated by the following form of the intraclass correlation was used:

$$\rho_I = \frac{MS_{students} - MS_{error}}{MS_{students} + (K-1)\, MS_{error}} \quad [1]$$

where the various mean squares (MS) were derived from a standard two-way analysis of variance and the mean square for error term represented variation due to the interaction of students and judges. Note that since only the reliability of the rank ordering of students, and not mean level of differences of judges' ratings, was of interest (i.e., interrater reliability versus interrater agreement), the error term did not include variation due to judges (Tinsley & Weiss, 1975).

Motivational and biographical data. The frequency and percentage of students endorsing various response alternatives to questions in the Puzzle Reaction Questionnaire, completed at the end of testing, were tabulated in order to determine students' prior experience with this problem type, the perceived difficulty of the instructions of the test, and the motivation and anxiety level of the students during the test. Completed posttest questionnaires were obtained from 50 students. Although the responses to the Puzzle Reaction Questionnaire were analyzed and provided useful information on the motivational characteristics of the total group, the small number of students distributing themselves over various response categories made group performance comparisons between students in different response categories inappropriate for many of the questions.

One exception was Question 2, which was especially important because it involved whether previous practice with problems of this type would affect test performance. The relationship between a student's prior experience with this problem type and his/her test performance was determined by performing t tests on the differences in mean total score (Total 2, MRATE) for those students reporting little or no prior experience with this problem type versus students reporting much experience.

Since problems of the type used in this study may require higher levels of motivation than more traditional psychometric measures, it was also important to investigate the effect of motivation level on performance with the limited data available. For this purpose t tests were performed on the performance means of students reporting different levels of motivation in Question 4.

In addition, since males as a group have generally been found to score higher than females as a group on tests of spatial abilities (Garai & Scheinfeld, 1968; MacCoby & Jacklin, 1974), it was of interest to determine whether sex differences existed for this test. Thus, a t test was used to compare the male and female group mean total scores.

Perceived Difficulty Ratings

Subjective ratings of the perceived difficulty of replications of the 15-puzzle were obtained from a separate sample of students in order to investigate the following questions:

1. What subjective dimensions do students use in evaluating the difficulty of this problem type?
2. How accurately can students evaluate the actual difficulty of these problems? That is, do difficulty ratings agree with actual performance data? How finely can discriminations be made between problems of similar difficulty levels?
3. Are there reliable individual differences in the perceived

difficulty of these problems and in the ability to make finer discriminations?

The latter two questions, in particular, address indirectly the question of whether students' perceptions of task difficulty can be related to their performance. For example, to the extent that reliable individual differences in the ability to visualize a sequence of moves in short-term memory exist, this might be expected to result in reliable differences in both perceived task difficulty and in actual task performance.

To maximally associate perceived difficulty with actual performance, the same students would ideally make the ratings and solve the problems. Due to limitations in student time, this was not possible in the present study; instead, a second sample from the same population was utilized. Using separate samples for the two tasks has the advantage that a student's rating of problem difficulty would not influence or be influenced by actual performance on the problem.

Procedure

Subjects. A total of 47 students from an introductory level psychology course rated the difficulty of 67 stimuli. Each stimulus consisted of a typed start-and-goal configuration for one 15-puzzle on an index card. To shorten the length of the rating task for each student, the 67 puzzles were divided into 4 sets of 16 or 17 puzzles each and the 47 subjects were randomly assigned to one of the 4 puzzle sets. Since the students were divided into groups merely to shorten the task, analyses were generally carried out for the sample as a whole; thus, the results will not be discussed separately for each group.

Data for three students were not included in the analysis because they failed either to perform or to record their ratings in accordance with instructions. Students took an average of about 40 to 45 minutes to complete the rating task.

Puzzle stimuli. Selection of the 67 puzzles used in this study was done with care because they were to be used in several ways. For example, in order to be able to trace the perceived difficulty trend within a single puzzle (which might require 16 moves from start to goal), ratings were obtained for several puzzles with the same goal configuration but with start configurations that converged on the goal. As a result, it was possible to detect how many moves from the goal a student would have to be before the problem would begin to look somewhat easy, then easy, and so on.

Since one hypothesized difficulty dimension was that of path length (or number of moves required), puzzles utilized a relatively uniform continuum of path lengths from 1 to 26. Of the 12 problems used in the problem-solving performance portion of the study, 9 were included among the stimuli rated in the rating task. Of these 9, 4 were divided into subpuzzles of varying lengths, as described above, in order to examine the perceived difficulty trend within the individual problems.

Rating procedure. Appendix D contains a copy of the self-administered instruction and recording booklet that each student received.

Students were told how this type of problem was solved so that they could rate how difficult they thought it would be if they had to solve it. Students first sorted the puzzles into six categories labeled Very Difficult, Difficult, Somewhat Difficult, Somewhat Easy, Easy, and Very Easy. It was made clear to the students that there were no required number of puzzles to be sorted into any of the piles but that they should put each puzzle into the category that had a label best describing how difficult they thought the puzzle would be to solve. In each puzzle set four of the puzzles were specially selected ahead of time to range from Very Easy to Very Difficult, in terms of path length. These four puzzles had a special message on the index card instructing the student to provide reasons, or a basis, for sorting the stimuli into a specific category. These reasons, along with the posttask questions (see Appendix D) regarding what rules or criteria they used for sorting into each of the six categories, constituted the protocols that were later analyzed to determine the dimensions on which the students thought they were sorting.

After recording the puzzles that were sorted into the original six categories, students were asked to attempt to break down each category into subcategories based on finer difficulty discriminations. The students were encouraged to subdivide into as many subcategories as they could but only to do so if they felt they could differentiate the difficulty of the puzzles in the same category. No re-sorting across the original six categories was allowed. After recording the stimuli in each of the final subdivided categories, students responded to a questionnaire that gathered information about their prior experience with this kind of puzzle, whether they had difficulty understanding the task, and their motivation level during the study. More importantly, students provided their own rules or criteria for sorting into each of the categories, for example, how they distinguished a Very Easy from a Somewhat Easy puzzle.

On the last page of the booklet, and after the students had already volunteered their own rating dimensions or rules, a list of nine dimensions was provided, which were hypothesized to be related to students' ratings. Students were asked to indicate for each of the nine dimensions whether they considered it in all, most, some, or none of the puzzle ratings. These nine dimensions also included two dimensions that were supposed to serve as validity dimensions (see Questions 8c and 8f in Appendix D). It was felt that these dimensions (particularly 8f) would be irrelevant to perceived difficulty and would therefore serve to detect students who were randomly responding or feeling that they should have used every dimension suggested by the experimenter.

Analysis

Reported dimensions of difficulty. Self-reported dimensions of perceived difficulty were thus of two types in this study. First, students voluntarily provided the basis for their difficulty judgments during the sorting task. During this portion of the task, students were provided no information as to the dimensions to be used in making their judgments. After sorting the puzzles into piles representing different perceived difficulty levels, an experimenter-provided list of possible rating dimensions was provided and students indicated whether they used each dimension on all, most, some, or none of the problems.

For each type of self-report (the voluntary protocols and the experimenter-provided dimensions), the proportion of students reporting use of each dimension was calculated and a determination was made of the most frequently used or important rating dimensions. Judgments of which dimensions were being reported during the sorting task were made by one graduate and one undergraduate research assistant and involved studying the students' written responses to the "Provide your reasons" section of the rating booklet (see Appendix D, Step 1) and Questions 5, 6, and 7 in the postrating task questionnaire (see Step 4 in Appendix D). Representative protocols provided by the students to indicate use of each reported rating dimension are contained in Appendix E.

Perceived difficulty mean ratings. Scale values representing mean perceived difficulty were obtained from the final subdivided category sorting of the puzzles. The center point of each of the original six categories was assigned the number 5, 15, 25, 35, 45, or 55 for the respective categories Very Easy, Easy, Somewhat Easy, Somewhat Difficult, Difficult, and Very Difficult. When puzzles within one of these six categories were subdivided into subcategories, the five integer intervals on each side of the center point were prorated or divided to assign differential rating values to each puzzle. The mean rating across all students was then computed to obtain the subdivided scale values. These subdivided scale values were then divided by 10 to scale them from 1 to 6, thus making them comparable to the original category labels. Thus, a puzzle felt to be Very Easy by the average student would have a scale value in the range of about .5 to 1.5, an Easy puzzle's scale value would range from 1.5 to 2.5, and so on.

These scale values were then used to determine the range of problems (e.g., problems requiring three to six moves) perceived to be in each of the categories (e.g., Very Easy, Easy) by plotting the scale values versus the solution path lengths of the puzzles. Finally, the relationship between perceived difficulty and actual performance on the set of puzzles administered to the first group of students was investigated by correlating mean difficulty ratings with the performance and response latency measures obtained for the nine puzzles that were included in both the performance and difficulty rating portions of this study.

Relationships Between Objective and Subjective Difficulty Indices

Each of the performance measures, response latency measures, physical problem characteristics, and the perceived difficulty mean ratings can be considered potential problem difficulty indices. For example, the difficulty of a problem could be indexed in several ways: (1) by the proportion of persons solving it, (2) by the average response latency used in working on the problem, or (3) by the number of squares needing to be moved large distances in the pattern. The similarity of the rank orders of various objective indices will likely vary.

In addition, the rank orderings of the problem difficulties by performance or physical indices obtained in the first part of the study can be compared with the rank ordering of subjective (perceived) difficulty obtained in the second part of the study. For this purpose, the Spearman rank-order correlation coefficient was computed between the rank orders of problem difficulty provided by all performance, latency, physical, and perceived difficulty indices. Some of the questions addressed through examination of these correlations were as follows:

1. Do the performance criteria used in this study (mean number of moves, proportion solving the problem, proportion solving it in the minimum number of moves) similarly index problem difficulty?
2. Do problems that take the most total time to solve or that require longer average move latencies also involve longer initial study times or latencies?
3. Is there a relationship between the difficulty of a problem as indexed by performance criteria and the initial move latency, average move latency, or total time taken in solving the problem?
4. How well does the perceived difficulty of the problems compare with the actual difficulty as indexed by performance and latency data and various physical attributes of the problem?
5. Which physical characteristics of the problem (e.g., path length, number of squares out of order) are most predictive of various performance and latency measures?

RESULTS

Computer-Administered Problems

Problem Characteristics

Indices of problem difficulty. Table 1 shows the number of students who attempted each problem (including the practice problem, Problem 1), the optimal or minimal number of moves required to solve each problem (path length), and the frequency and percentage of students who used various numbers of moves before solving or giving up working on the problem. These data suggest that most of the problems were too easy, with from 70.4% to 98.2% of the students solving 9 of the 13 problems in the optimal number of moves. Problems 10, 12, 13, and, to a lesser extent, Problem 9 were more challenging, with from 14.6% to 45.7% of the students solving the problems in the optimal number of moves. The data in Table 1 also show that the optimal number of moves was not a perfect indicator of difficulty as indexed by student performance. Problems 4 and 5, which could optimally be solved in 8 moves, were solved in the optimal number of moves less frequently (75.9% and 77.8%) than Problem 6 (87.0%), for which the optimal number of moves was 10. Similarly, Problems 10 and 11 could both be solved optimally in 14 moves; but only 29.5% of the students solved Problem 10 in that number of moves, whereas 79.6% of the students solved Problem 11 in the optimal number of moves.

Additional data on student performance characteristics of the problems are shown in Table 2. With the exception of Problems 9, 10, 12, and 13, the mean number of moves used on each problem (row 1 of Table 2) were quite close to the minimum number of moves required for its solution (row 9). Row 2 of Table 2 shows that all students solved the first five problems in the allowed maximum number of moves (for the longer problems the maximum number of moves allowed was 28), and only for Problems 12 (66.6% solving) and 13 (66.4% solving) were there sub-

Table 1
Optimum Number of Moves and Distributions of
Observed Number of Moves for Each Problem

Problem Number	Optimum No. of Moves	N	Observed No. of Moves	Frequency N	Frequency %
1	3	55	3	54	98.2
			5	1	1.8
2	4	33	4	31	93.9
			12	1	3.0
			13	1	3.0
3	6	33	6	32	97.0
			8	1	3.0
4	8	54	8	41	75.9
			10	8	14.8
			18	1	1.9
			21	1	1.9
			25	2	3.7
			26	1	1.9
5	8	54	8	42	77.8
			12	4	7.4
			13	1	1.9
			14	3	5.6
			18	1	1.9
			24	1	1.9
			26	2	3.7
6	10	54	10	47	87.0
			12	3	5.6
			16	1	1.9
			27	1	1.9
			32	1	1.9
			34	1	1.9
7	10	54	10	38	70.4
			12	2	3.7
			16	2	3.7
			18	1	1.9
			20	8	14.8
			22	1	1.9
			26	1	1.9
			30	1	1.9
8	12	50	12	39	78.0
			14	1	2.0
			20	1	2.0
			24	1	2.0
			25	1	2.0
			27	1	2.0
			28	3	6.0
			32	1	2.0
			33	1	2.0
			35	1	2.0
9	12	46	12	21	45.7
			14	5	10.9
			16	3	6.5
			18	1	2.2
			20	1	2.2
			22	2	4.3
			24	2	4.3
			25	1	2.2
			26	2	4.3
			27	1	2.2
			28	2	4.3
			32	2	4.3
			33	1	2.2
			35	1	2.2
			36	1	2.2
10	14	44	14	13	29.5
			16	3	6.8
			18	7	15.9
			20	2	4.5
			21	1	2.3
			22	3	6.8
			24	1	2.3
			26	3	6.8
			27	4	9.1
			28	1	2.3
			32	1	2.3
			34	2	4.5
			35	3	6.8
11	14	49	14	39	79.6
			16	4	8.2
			20	1	2.0
			22	1	2.0
			26	1	2.0
			27	1	2.0
			28	1	2.0
			32	1	2.0
12	16	48	16	7	14.6
			18	3	6.3
			20	5	10.4
			23	1	2.1
			24	3	6.3
			25	3	6.3
			26	4	8.3
			27	3	6.3
			28	3	6.3
			30	5	10.4
			32	4	8.3
			33	2	4.2
			34	2	4.2
			35	2	4.2
			39	1	2.1
13	16	49	16	10	20.4
			18	1	2.0
			19	1	2.0
			20	2	4.1
			22	2	4.1
			25	4	8.2
			26	2	4.1
			27	3	6.1
			28	8	16.3
			30	4	8.2
			32	1	2.0
			33	1	2.0
			34	7	14.3
			35	3	6.1

stantial numbers of students failing to solve the problems. Row 3 reports the proportion of students solving each problem in the optimal number of moves. With the exception of Problems 9, 10, 12, and 13, 70% or more of the students were able to solve the rest of the problems in the optimal number of moves.

Row 4 of Table 2 contains the Special Difficulty Index, which adjusts for the differing path lengths (minimum number of moves required) of the problems. For example, for Problem 4 this index equaled 1.21, indicating that the average student required 21% more than the minimum number of moves to solve the problem. The difficulty of each problem as indicated by this index agreed quite well with the performance data in rows 1 through 3. Again, only Problems 9, 10, 12, and 13, with special indexes of 1.45, 1.44, 1.50, and 1.51, required substantial numbers of moves over the minimum number required for solution.

A comparison of the performance index and the Special Difficulty Index with the minimum number of moves required (row 9) indicates that although the difficulty of the problems tended to increase with solution path length (minimum number of moves required), the relationship was not strictly monotonic. For example, although Problem 11 required at least 14 moves for solution, this problem was much easier for students than some of the problems requiring fewer moves. Thus, minimal solution path was not the sole determinant of a problem's difficulty.

Mean problem latencies. Rows 5 through 7 of Table 2 show the mean initial, average, and total latencies of students for each of the 13 problems. The data on average amount of time spent by students prior to their first move (mean initial latency) indicates a strong, though not perfect, relationship with the difficulty of the problem as indexed by the other performance criteria. This relationship appears even stronger for the total time in seconds used by the average student (row 7) to solve problems of varying difficulty. For example, the mean initial and total move latencies were smallest for two of the problems with the shortest path lengths (Problems 2 and 3) and were longest for the four problems with the longest path lengths (Problems 9 through 13). The trend for the remaining seven problems was less consistent, except that students seemed, not surprisingly, to use more time to study and to complete the practice problem than would be predicted on the basis of its short path length. Students usually took about 20 to 60 seconds to make their first move, whereas total time working on a single problem ranged from about 67 to 361 seconds. Most problems were solved in about 2.5 minutes (150 seconds) or less.

There appeared to be no consistent relationship between path lengths of the problems and the average latency for the moves within a single problem (row 6 in Table 2). Students generally took from 8 to 15 seconds to make a single move, although again more time was taken on the practice problem (Problem 1).

Perceived scale values. Row 8 of Table 2 shows the mean perceived difficulty scale values for the nine test problems that were included in the perceived difficulty rating portion of the study. Given the assignment of the numbers 1, 2, 3, 4, 5, and 6, respectively, to the categories Very Easy, Easy, Somewhat Easy, Somewhat Difficult, Difficult, and Very Difficult, row 8 shows that none of the problems was

Table 2
Indices of Problem Difficulty:
Problem Characteristics, Student Performance Data, and Perceived Scale Values for Each Problem

Difficulty Index	Problem Number 1	2	3	4	5	6	7	8	9	10	11	12	13
Mean No. of Moves	3.04	4.52	6.06	9.68	9.90	11.39	12.81	15.24	18.00	21.04	15.61	25.54	25.84
Proportion Solving	100.00	100.00	100.00	100.00	100.00	94.50	98.20	88.00	84.80	86.40	98.00	66.60	66.40
Proportion Solving Optimally	98.20	93.90	97.00	75.90	77.80	87.00	70.40	78.00	45.70	29.50	79.60	14.60	20.40
Special Difficulty Index	1.01	1.13	1.01	1.21	1.23	1.12	1.28	1.24	1.45	1.44	1.11	1.50	1.51
Mean Initial Latency	53.50	27.40	21.80	43.90	45.60	39.90	33.10	43.10	61.70	62.20	45.10	63.60	59.50
Mean Average Latency	25.80	15.30	10.70	14.00	15.10	12.80	8.90	9.90	14.00	13.10	9.30	11.60	14.50
Mean Total Latency	94.30	82.30	67.40	151.50	160.90	150.20	117.90	151.30	244.90	278.30	150.60	290.10	361.60
Perceived Difficulty				2.20		3.00	2.90	3.60	3.91	4.45	3.08	4.31	3.00
Solution Path Length	3	4	6	8	8	10	10	12	12	14	14	16	16
Mean Illegal Moves	.49	.82	.21	.87	.54	.50	.63	.58	.74	1.05	.61	1.56	.92
Mean Repeated Moves	.04	.15	.03	.37	.17	.32	.04	.18	.56	1.09	.22	1.02	1.40

considered Difficult or Very Difficult by the average student. Four problems (9, 10, 12, and, to a lesser extent, 8) were considered Somewhat Difficult, and the remaining problems (4, 6, 7, 11, and 13) were perceived as Easy or Somewhat Easy. The difficulty perceptions generally indicated agreement with actual performance indices of difficulty, but there were some marked exceptions. In particular, Problems 8 and 11 were perceived as being more difficult than was indicated by the performance data, whereas the most difficult problem--Problem 13 with Special Difficulty Index of 1.51 (row 4)--was perceived as being Somewhat Easy by the average student. These data indicate that students' initial difficulty perceptions of these problems are fallible, particularly for problems with longer solution paths.

<u>Illegal and repeated moves</u>. Rows 10 and 11 of Table 2 contain the mean number of illegal and repeated moves made on each problem. These data indicate that students made few illegal or repeated moves (means less than 1.0) on most of the problems and that, with the exception of Problems 10, 12, and 13, there seemed to be little if any relationship between the difficulty or the minimum number of moves required and the number of illegal or repeated moves. For Problems 10, 12, and 13, however, the average student made approximately one or more illegal and one or more repeated moves. This is to be expected for the more difficult problems, since the students worked longer on them and thus had a greater chance of making typing errors and other illegal moves. It would be difficult to unconfound this tendency with any tendency to be more careless on the more difficult problems. The slightly increased number of repeated move configurations for Problems 10, 12, and 13 may be more meaningful, indicating a greater likelihood of students needing to back up in their solutions to the more difficult problems. Because of the small number of illegal and repeated moves made by the average student on these problems, these measures were not considered further as potential indices of problem difficulty (e.g., they do not appear in Table 3).

<u>Relationships among indices of problem difficulty</u>. Table 3 shows rank-order correlations among the potential indices of problem difficulty--performance indices, latency measures, perceived difficulty, and various physical problem characteristics. Data for Variables 1 through 9 are in Table 2; data for Variables 10 through 14 are in Appendix C for each of the problems.

The correlations in rows 2 through 4 of Table 3 show that the difficulty indices based on group performance data rank ordered the difficulty of the problems quite similarly, with the strongest agreement between the Special Difficulty Index and the proportion of students solving the problem in the optimal number of moves (ρ=-.95) and between the mean number of moves used and the proportion of students solving the problem in the maximum allowed moves (ρ=-.94). The utility of the Special Difficulty Index over the other performance indices of difficulty is suggested by its lower correlation with solution path length (ρ=.77). For example, using the mean number of moves required by the sample to solve different problems is less adequate as an indicator of problem difficulty because it labeled all puzzles with long solution paths as difficult (ρ=.98) when, in fact, not all long puzzles were difficult (e.g., Problem 11).

Table 3

Rank-Order Correlations of Problem Difficulty Indices

Difficulty Index	Performance Indices				Latencies			Perceived Difficulty		Physical Problem Characteristics				
	1	2	3	4	5	6	7	8		9	10	11	12	13
1. Mean Moves														
2. Proportion Solving	-.94													
3. Proportion Solving Optimally	-.86	.75												
4. Special Difficulty Index	.85	-.81	-.95											
5. Mean Initial Latency	.65	-.63	-.67	.63										
6. Mean Average Latency	-.34	.24	.15	-.07	.25									
7. Mean Total Latency	.84	-.76	-.89	.86	.84	.08								
8. Perceived Difficulty	.64	-.63	-.43	.40	.75	.00	.48							
9. Solution Path Length (Optimal Moves)	.98	-.91	-.80	.77	.61	-.42	.79	.63						
10. No. of Squares Not matching	.42	-.37	-.11	.04	.04	-.67	.19	.32	.51					
11. No. of Rows Not Matching	.25	-.26	-.04	.00	.46	-.21	.17	.70	.34	.37				
12. No. of Columns Not matching	.23	-.20	.10	-.08	-.10	-.35	.13	-.03	.30	.81	.06			
13. Euclidean Distance	.62	-.60	-.36	.31	.22	-.68	.35	.38	.70	.85	.34	.60		
14. City-Block Distance	.67	-.62	-.43	.35	.27	-.71	.37	.37	.74	.79	.35	.48	.98	

Note. For Row 8 only, correlations greater than .64 are statistically significant; for all other rows correlations greater than .51 are statistically significant.

The intercorrelations of the latency variables in rows (and columns) 5 through 7 of Table 3 indicate that only the mean initial and total latency measures rank ordered the problem difficulties similarly. That is, problems which took longer times to solve were also studied longer initially (ρ=.84), but the average time for moves within a problem was not significantly related to either the initial move (ρ=.25) or the total problem latency (ρ=.08).

The correlations between the latency variables (rows 5 through 7) and the performance variables (columns 1 to 4) show that the total time spent on a problem (row 7) by the average student was highly predictive (ρ's=.84, -.76, -.89, .86) of difficulty as indexed by performance indices, and the amount of initial study time spent by the average student (row 5) was also strongly related to the four performance indices (ρ's=.65, -.63, -.67, .63). That is, not surprisingly, more difficult problems were studied longer initially and took longer to solve. The correlations in columns 5 to 7 of row 9 also show a strong relationship between mean initial and total latency and solution path length (ρ's= .61 and .79, respectively), indicating that the problems with longer solution paths were studied longer initially and worked on longer.

The correlations in columns 1 to 4 of row 8 show that students' perceptions of problem difficulties agreed somewhat, but not as much as might be expected, with the actual performance measures (ρ's=.64, -.63, -.43, .40). Although all these correlations were in the appropriate direction, only the first two approached statistical significance due to the small number of problems (nine) for which both performance and perceived difficulty indices were available. The perceived difficulty scale values in row 8 of Table 2 suggest that this lower-than-expected relationship was due to the students' inability to differentiate the relative difficulties of problems with longer solution paths (such as those used in Problems 9 through 13). The correlation between perceived difficulty and solution path length (ρ=.63) was not as high for the problems solved on the computer as for the larger stimulus set used in the rating study (r=.88), probably because the range of path lengths used in the computer test was more restricted.

The only significant correlation between perceived difficulty and latencies (columns 5 to 7) was with the mean initial latency measure (ρ=.75). In fact, this represented the highest correlation in the matrix for both variables. This relationship suggests that the problems that were studied longest before a move was made were the ones perceived as being most difficult (even more than whether or not these problems actually were the most difficult).

Examination of the correlations in column 8 shows that perceived difficulty of the problems in the test was significantly related to only two physical problem characteristics--solution path length (ρ=.63) and number of rows not matching in the two patterns (ρ=.70). Correlations with some of the other physical problem characteristics, e.g., the number of squares not matching and the Euclidean and City-Block distance functions, were probably restricted by the reduced range of values in the computerized test as opposed to the rating study (see section below on dimensions of perceived difficulty).

Examination of rows 9 to 14, columns 1 to 4, shows that only the solution path length (row 9), and to a lesser extent the Euclidean and City-Block distances (rows 13 and 14), were useful in predicting difficulty as indexed by the four performance measures of difficulty. Solution path length rank ordered problem difficulty quite similarly to the four performance measures (ρ's=.98, -.91, -.80, and .77), being most independent of the Special Difficulty Index (ρ=.77). The two distance functions moderately predicted mean number of moves (ρ's=-.36 and -.43, neither significant) and the Special Difficulty Index (ρ's=.31 and .35, neither significant).

Solution path length (row 9) was the only physical problem characteristic to predict mean initial (ρ=.61) or mean total (ρ=.79) problem latency. Interestingly, while average move latency (column 6) was not related to any of the performance criteria, it was inversely related to three physical problem characteristics--the number of squares not matching (ρ=-.67), the Euclidean distance function (ρ=-.68), and the City-Block distance function (ρ=-.71). These negative correlations suggest the possibility that students worked faster and made moves more quickly when they could see that many numbers would need to be moved, especially if these numbers had to be moved long distances in the puzzle.

The intercorrelations of the physical problem characteristics in rows and columns 9 to 14 show that the more highly related problem characteristics were solution path length, the number of squares not matching, and the two distance functions. For this set of problems, the Euclidean and City-Block distances were virtually identical (ρ=.98). Although the number of rows not matching did not relate to other physical problem characteristics, the number of columns not matching did correlate with the number of squares not matching (ρ=.81) and the Euclidean distance measure (ρ=.60). Whether the number of rows or columns not matching was more or equally related to other physical indices, however, is strictly dependent on the particular set of problem replications used.

Assessment of Individual Student Performance

Scoring methods. For each individual problem two scores were computed--Score 1, defined as the number of moves the student required divided by the minimum number required, and Score 2, defined as Score 1 divided by (corrected for) the Special Difficulty Index. Four total scores were also derived--Total 1 and Total 2 were the averages over the problems attempted of Score 1 and Score 2, respectively, and PROPS and PROPM were the proportion of problems attempted that were solved within the maximum allowed moves (PROPS) and in the minimum number of moves (PROPM). Table 4 shows the means, standard deviations, and range of all these scores for the present sample.

Note that although not all students worked on each individual problem, thus not having a score (Score 1, Score 2) for each problem, the four total scores were obtainable for all students (N = 55) as a result of the way these scores were defined. PROPS and PROPM can be considered additive scores, which essentially total the number of problems solved or solved optimally; whereas Total 1 and especially Total 2 take into account the pattern of scores across the problems attempted. The latter two scores would appear to be particularly appropriate for

Table 4
Mean, Standard Deviation, and Range of Four Total Scores and Thirteen Individual Problem Scores

Score	Problem	*N*	Mean	Standard Deviation	Best Score Obtained[a]	Poorest Score Obtained[a]
PROPS	—	55	.83	.12	1.00	.50
PROPM	—	55	.66	.16	1.00	.36
Total 1	—	55	1.25	.18	1.00	1.70
Total 2	—	55	1.00	.14	.84	1.38
Score 1						
	1	55	1.01	.09	1.00	1.67
	2	33	1.13	.52	1.00	3.25
	3	33	1.01	.06	1.00	1.33
	4	54	1.21	.56	1.00	3.25
	5	54	1.23	.55	1.00	3.25
	6	54	1.12	.41	1.00	2.80
	7	54	1.28	.49	1.00	2.80
	8	50	1.24	.49	1.00	2.33
	9	46	1.45	.54	1.00	2.33
	10	44	1.44	.40	1.00	2.00
	11	49	1.11	.27	1.00	2.00
	12	48	1.50	.29	1.00	1.75
	13	49	1.51	.31	1.00	1.75
Score 2						
	1	55	1.00	.09	.99	1.65
	2	33	1.00	.08	.89	2.88
	3	33	1.00	.06	.99	1.32
	4	54	1.00	.46	.83	2.69
	5	54	1.00	.44	.81	2.64
	6	54	1.00	.37	.89	2.50
	7	54	1.00	.38	.78	2.19
	8	50	1.00	.40	.81	1.88
	9	46	1.00	.37	.69	1.61
	10	44	1.00	.28	.69	1.39
	11	49	1.00	.25	.90	1.80
	12	48	1.00	.19	.67	1.17
	13	49	1.00	.20	.66	1.16

[a]Note that higher numbers represent better scores for the PROPS and PROPM scores and lower numbers reflect better scores for the Total 1 and Total 2 scores.

adaptive testing where not all students work on the same problems.

From the mean PROPS score it can be seen that the average student solved 83% of the problems attempted in the maximum allowable moves. At least one student solved all the problems attempted (best score = 1.00), and the student with the poorest score (.50) solved only half of the problems attempted. The PROPM data indicate that the average student solved 66% of the problems attempted in the optimal number of moves, with proportions ranging from 100% to 36% solved optimally. The Total 1 mean score shows that the average student required 25% (mean = 1.25) more moves than optimally required to solve the average problem.

At least one student averaged 70% more moves than required (poorest score = 1.70), and one solved all problems attempted in the minimum number of moves (best score = 1.00).

The Total 1 score represents the proportion of moves beyond the minimum number possible, and the Total 2 score represents the proportion of moves greater or less than the mean number required by the group as a whole. This is also true for the difference between the two individual problem scores, Score 1 and Score 2. Thus, by definition, the mean Total 2 score and mean Score 2 equal 1.00. The best Total 2 score was .84, indicating an average problem solution of 16% fewer moves than the group norm; whereas the poorest Total 2 score was 1.38, indicating that one student required 38% more moves on the average than did the average student in the group.

By definition, the mean Score 1 for each problem will be equal to the Special Difficulty Index (i.e., mean number of moves required by the sample divided by optimal number of moves). However, the data in

Table 5
Independent Judges' Ratings and Mean Rating (MRATE) of Total Test Performance for Each Student

Student	Judge 1	Judge 2	Judge 3	MRATE	Student	Judge 1	Judge 2	Judge 3	MRATE
1	6	6	7	6.3	30	6	4	4	4.7
2	7	5	6	6.0	31	6	5	7	6.0
3	6	6	7	6.3	32	3	3	3	3.0
4	4	4	5	4.3	33	2	3	3	2.7
5	5	5	5	5.0	34	7	8	6	7.0
6	7	7	8	7.3	35	8	6	7	7.0
7	6	6	8	6.7	36	5	4	7	5.3
8	4	3	6	4.3	37	2	3	6	3.7
9	7	5	7	6.3	38	7	6	2	5.0
10	8	6	8	7.7	39	6	4	5	5.0
11	4	3	6	4.3	40	5	5	5	5.0
12	8	8	8	8.0	41	5	5	6	5.3
13	5	5	6	5.3	42	4	5	6	5.0
14	4	4	5	4.3	43	9	8	9	8.7
15	2	2	3	2.0	44	4	3	2	3.0
16	3	3	4	3.3	45	5	4	6	5.0
17	8	8	8	8.0	46	7	7	8	7.3
18	7	5	6	6.0	47	5	5	5	5.0
19	6	5	5	5.3	48	5	5	6	5.3
20	5	5	5	5.0	49	3	3	3	3.0
21	4	5	5	4.7	50	8	8	8	8.0
22	3	3	3	3.0	51	7	6	7	6.7
23	8	6	7	7.0	52	2	2	1	1.7
24	6	4	5	5.0	53	7	5	8	6.7
25	6	5	5	5.3	54	2	3	2	2.3
26	3	3	3	3.0	55	1	2	2	1.7
27	8	8	7	7.7	Mean	5.3	4.8	5.4	5.2
28	3	3	3	3.0	*SD*	2.0	1.7	2.0	1.8
29	6	5	5	5.3					

Table 4 show differing levels of difficulty for the 13 problems as indexed by Score 1. For example, Problem 9 (mean Score 1 = 1.45) was more difficult for the sample than Problem 4 (mean Score 1 = 1.21), since Problem 9 required an average of 45% more moves than the optimal number versus 21% more moves than the optimal number for Problem 4.

Score 2, like Total 2, indexes performance relative to the mean student. As a result, the mean Score 2 across all students is 1.00 for each problem by definition. Values of Score 2 below 1.00 indicate fewer moves than the average student, and scores greater than 1.00 reflect more moves than the average student. Examination of the best and poorest values of Score 2 indicate considerable variability in student performance on the problems. The best student on Problem 13 completed the problem in two-thirds of the average number of moves required by the average student, and the poorest student on Problem 2 required 2.88 times the average number of moves.

Judges' ratings of performance. Table 5 contains the ratings on a 10-point scale of each student's overall test performance by three independent judges and the resulting mean rating (MRATE) used as a criterion in this study against which to compare the alternative scoring methods. The mean and standard deviation of each judge's ratings and the overall mean ratings are also shown. The means of each column were all close to 5.0, which is appropriate, since the judges were instructed to assign a rating of 5.0 to students with average performance. The similar standard deviations indicate a comparable spread of judgments by each judge. For only 6 of 55 students did any two judges differ by more than 2 in their assigned ratings; of these 6 students 4 were inconsistent in that they performed either very well on most problems and very poorly on a few (Students 8 and 11) or well on some difficult problems but less well on easier ones (Students 37 and 53). One of the students (Student 36) did not have data for three problems on an important part of the test, making it difficult to evaluate that student's overall performance on the test.

Table 6 shows the results of the interrater reliability analysis. As Table 6 shows, most of the variance in ratings was due to individual differences in student performance, and substantial interrater reliability ($\rho_I = .80$) was obtained.

Table 6
Sources of Variance in Performance Ratings

Sources of Variance	*df*	*SS*	*MS*
Students (*s*)	54	502.5	9.3
Judges (*j*)	2	10.6	5.3
Error (*s* × *j*)	108	75.4	.7

Relationship between judges' ratings and scoring methods. Table 7 shows the Spearman rank-order coefficients between each of the individual total performance scores (PROPS, PROPM, Total 1, and Total 2) and MRATE. In terms of its relationship with the other scoring methods and MRATE, PROPS was clearly the least adequate total score. This is not surprising, since this scoring method does not use important information on the differential number of moves that are less than the maximum allowed. The highest relationship between scores was between Total 1 and Total 2 (r = .96); these scores undoubtedly are so similar in this study because the test was not adaptive. Most students attempted the same problems, so that the Total 2 adjustment for the difficulty level of the problems attempted did not differentiate between students. In an adaptive test where students converged on problems of varying difficulty levels, performance as indexed by Total 1 and Total 2 would be expected to differ appreciably.

Table 7
Spearman Rank-Order Correlations Between Individual Total Performance Scores and Mean Performance Ratings

Score	PROPS	PROPM	Total 1	Total 2
PROPS				
PROPM	.71			
Total 1	-.79	-.88		
Total 2	-.74	-.81	.96	
MRATE	.68	.87	-.85	-.89

Note. All Spearman coefficients significant at $p \leq .001$.

Although the correlation of these two scores was high, examination of the students who were classified as the best performers by each score showed that they did evaluate performance differently. The top 10 students on each score were essentially the same group, with the exception of three students who had the top three Total 1 scores but ranked 14 through 16 on Total 2 scores. All three of these students worked only on the easier problems and solved them all in close to the optimal number of moves; as a result, their Total 1 scores were high. However, many students who did well on the more difficult problems received higher Total 2 scores as well, because such scores take into account the difficulty level of problems attempted.

If the judges' ratings, which examined each protocol in a more comprehensive way, were used as a criterion against which to evaluate the different scoring methods, Total 2 was slightly but not significantly better than the PROPM and Total 1 scores. The judges, in describing how they made their ratings, were clearly taking into account not only the number of moves beyond the optimal number (Total 1) but also the relative difficulty of the problems attempted by each student; therefore, if students had worked on problems of more varied difficulty levels, Total 2, which takes both these factors into account, would seem to be even more superior to PROPM and Total 1.

Consistency of performance across problems. Important for the usefulness of this problem type in assessing spatial problem-solving ability is whether reliable individual differences on various performance criteria can be identified across problem replications of similar and varying difficulty levels. Table 8 shows the intercorrelations of the total number of moves used by students (lower triangle) and the intercorrelations of the number of illegal moves made (upper triangle) across the 13 problem replications. The correlations in the lower half of Table 8 fail to demonstrate strong consistency of the Number of Moves performance measure across problems. That is, there was not a consistent tendency for students to rank order themselves similarly across problems on this performance score. Some small clusters of statistically significant and moderate size correlations existed between Problems 2 through 4, Problems 5 through 10, and to a lesser extent

Table 8
Intercorrelations of Number of Illegal Moves (Upper Triangle), and
Total Number of Moves (Lower Triangle) for 13 Problems

Problem		1	2	3	4	5	6	7	8	9	10	11	12	13
1	*N*		*33*	*33*	*54*	*54*	*54*	*54*	*50*	*46*	*44*	*49*	*48*	*49*
	r		.41	.13	.14	.07	-.23	-.04	-.08	-.17	-.20	.03	-.15	-.03
2	*N*	*33*		*33*	*33*	*32*	*32*	*32*	*30*	*28*	*28*	*28*	*28*	*29*
	r	-.04		.08	.35	.48	.13	.48	.20	.11	-.10	.29	.02	.42
3	*N*	*33*	*33*		*33*	*32*	*32*	*32*	*30*	*28*	*28*	*28*	*28*	*29*
	r	-.03	.74		-.15	-.20	.20	.16	-.19	.06	.16	-.08	-.10	-.20
4	*N*	*54*	*33*	*33*		*53*	*53*	*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	-.05	.41	.60		.24	.06	.41	.12	.03	-.06	.29	.39	.39
5	*N*	*54*	*32*	*32*	*53*		*54*	*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	-.06	.02	-.08	.25		.24	.21	.19	.02	-.07	.24	.36	.57
6	*N*	*54*	*32*	*32*	*53*	*54*		*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	-.04	-.09	-.06	-.04	.21		.20	.04	-.00	.23	.33	.17	.07
7	*N*	*54*	*32*	*32*	*53*	*53*	*53*		*50*	*46*	*44*	*49*	*48*	*49*
	r	.21	.10	-.11	.13	.34	.32		.23	.07	.12	.14	.33	.32
8	*N*	*50*	*30*	*30*	*49*	*49*	*49*	*50*		*46*	*44*	*48*	*47*	*48*
	r	-.07	-.12	-.08	.18	.38	.10	.06		.27	.17	.04	.03	.22
9	*N*	*46*	*28*	*28*	*45*	*45*	*45*	*46*	*46*		*44*	*45*	*44*	*45*
	r	-.12	-.22	-.16	-.05	.10	.35	.12	.13		.14	.07	.20	.20
10	*N*	*44*	*28*	*28*	*43*	*43*	*43*	*44*	*44*	*44*		*43*	*42*	*44*
	r	.22	-.05	-.11	.20	.03	.30	.35	-.20	.14		.29	.02	.18
11	*N*	*49*	*28*	*28*	*48*	*48*	*48*	*49*	*48*	*45*	*43*		*47*	*48*
	r	.02	.18	-.10	.17	.18	-.10	.18	.19	-.12	.08		.35	.12
12	*N*	*48*	*28*	*28*	*47*	*47*	*47*	*48*	*47*	*44*	*42*	*47*		*47*
	r	.13	.17	.14	-.03	-.07	.08	.01	.00	.41	.26	-.11		.41
13	*N*	*49*	*29*	*29*	*49*	*49*	*49*	*49*	*48*	*45*	*44*	*48*	*47*	
	r	.11	.04	.17	.12	-.14	.15	.02	-.06	.26	-.02	-.03	.05	

between Problems 9, 10, 12, and 13. These moderate positive correlations, which tend to be located near the diagonal, suggest that although individual differences as indexed by total number of moves were not very consistent for the particular set of problems used here, consistency of performance was more likely to be obtained across problems of more similar difficulty levels.

A probable reason for the lack of consistent performance across problems is the small variation in performance for most of the problems due to the overall easiness of the test. With the majority of students solving many problems in the minimal or close to minimal number of moves, the low variability of the performance scores across problems would greatly decrease correlations.

Similarly, there was not a strong tendency for the same students to make more illegal moves across problems, as indicated in the upper half of Table 8. However, many more moderate and statistically significant correlations existed than would be expected by chance. It was originally expected that the number of illegal moves might relate to difficulty in understanding the instructions and problem task. The small number of illegal moves made by students on most problems (see Table 2), however, not only decreased the likelihood of large correlations across problems but also suggested that the moderate correlations that did appear were due more to carelessness on the part of some students in entering their responses on the CRT.

From Table 2 it was also seen that there were very few repeated moves made by students, indicative of backing up in the problem solution. Not surprisingly, then, no strong consistency across problems was found for this performance index (see correlation matrix in Appendix Table F-1).

To examine the relationship between the number of legal moves used, the number of illegal moves, and the number of repeated moves within a single problem and across problems, the intercorrelation matrices between these performance indices were computed (see Appendix Tables F-2, F-3, and F-4). If all three indices were related to ability to solve these problems, they should be related to each other within and across problems. Examination of the intercorrelation matrices demonstrated that the number of total, illegal, or repeated moves on the same or on a different problem were not highly correlated, with the exception that within the same problem the number of repeated moves correlated moderately highly (average r = .49) with the number of total moves (see Appendix Table F-3). This latter relationship is not surprising, since it is a part-whole correlation, with the number of repeated moves being included in the total number of moves.

Another way to examine consistency of performance is to relate performance on individual problems with performance on the test as a whole, as indexed by various total scores. These "item-total" correlations, shown in Table 9, can assist in selecting the problems that are most discriminating. In Table 9 the five or six highest correlations in each row are underlined. These data indicate that generally problems in the middle range of difficulty (Problems 4 to 10) were most discriminating. Since correlations between individual problem scores and the four alternative total scores are to varying degrees part-whole correlations, the last two rows of Table 9 show the correlations between a problem score and the total score on the remaining problems using the two total scores discussed earlier as being the most promising (Total 1, Total 2). Considering that the problem-excluded total scores consist of only 12 "items" and that the easiest and most difficult problems were not very discriminating, some of the correlations are encouraging. The data suggest that if several problems can be tailored to the same difficulty level (see discussion of Table 8 above), one appropriate for each individual student, improved reliability may be obtained.

Table 9

Product-Moment Correlations Between Individual Problem Scores (Score 1, Score 2[a]) and Several Total Test Scores, by Problem

	Problem												
Total Score	1	2	3	4	5	6	7	8	9	10	11	12	13
PROPS	-.07	-.21	-.20	-.41	-.36	-.47	-.24	-.31	-.47	-.38	-.25	-.36	-.42
PROPM	-.11	-.32	-.18	-.44	-.35	-.36	-.58	-.29	-.46	-.40	-.21	-.37	-.38
Total 1	.06	.35	.26	.53	.55	.49	.60	.41	.49	.43	.26	.33	.27
Total 2	.06	.36	.27	.56	.61	.49	.61	.43	.44	.43	.32	.28	.22
Problem Excluded													
Total 1	.04	.14	.24	.30	.30	.31	.39	.18	.26	.27	.15	.18	.10
Total 2	.02	.11	.23	.32	.39	.29	.42	.21	.24	.28	.19	.16	.10

Note. If $|r| > .36$, $p \leq .01$; if $|r| > .27$, $p \leq .05$.

[a]Since Score 1 and Score 2 were linear transformations of each other, correlations with total scores were identical.

Response Latencies

Consistency of latency measures across problems. Table 10 shows the intercorrelations of initial response latencies (lower triangle) and average response latencies (upper triangle) across all 13 problems. The initial latency correlations showed a moderate to strong tendency for individuals to be consistent in the amount of time they spent in initial study of a problem prior to their first move. There was an even stronger tendency for the average time per move to be consistent across problems, with most of the correlations in the .30 to .50 range and many in the .70 to .80 range.

Table 11 shows the intercorrelations of the total time spent on each problem. These data indicate a moderate relationship across most problems.

Thus, there seemed to be a substantial degree of consistency in the initial, average, and total time taken by individuals in working on these problems. The response latency measures may tap differences in the cognitive style of reflectivity versus impulsiveness (Kagan, 1965; Kagan et al., 1964) or the degree of planning by the student. Since all three correlation matrices (initial, average, and total latencies) showed a slight tendency for the correlations to be largest near the diagonal, the work strategy or style of each student may vary somewhat at different points in the test, being more consistent for problems that are worked on closer to each other in time.

The response latency measures may also reflect individual differ-

Table 10
Intercorrelations of Initial Response Latencies (Lower Triangle), and Average Response Latencies (Upper Triangle) for 13 Problems

Problem		1	2	3	4	5	6	7	8	9	10	11	12	13
1	*N*		*33*	*33*	*54*	*54*	*54*	*54*	*50*	*46*	*44*	*49*	*48*	*49*
	r		.74	.65	.50	.45	.45	.27	.19	.20	.20	.17	.32	.20
2	*N*	*33*		*33*	*33*	*32*	*32*	*32*	*30*	*28*	*28*	*28*	*28*	*29*
	r	.35		.83	.85	.77	.84	.67	.48	.25	.26	.31	.45	.50
3	*N*	*32*	*32*		*33*	*32*	*32*	*32*	*30*	*28*	*28*	*28*	*28*	*29*
	r	.29	.14		.84	.78	.81	.34	.11	.32	.37	.30	.33	.41
4	*N*	*54*	*33*	*32*		*53*	*53*	*53*	*49*	*45*	*43*	*46*	*47*	*49*
	r	.31	.26	.34		.71	.81	.55	.47	.30	.28	.35	.22	.43
5	*N*	*54*	*32*	*31*	*53*		*54*	*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	.21	.16	.04	.54		.66	.59	.46	.41	.59	.35	.32	.56
6	*N*	*54*	*32*	*31*	*53*	*54*		*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	.27	.34	.08	.63	.52		.47	.49	.43	.53	.29	.24	.47
7	*N*	*54*	*32*	*32*	*53*	*53*	*53*		*50*	*46*	*44*	*49*	*48*	*49*
	r	.19	-.21	.24	.51	.39	.25		.58	.48	.48	.44	.32	.57
8	*N*	*50*	*30*	*30*	*49*	*49*	*49*	*50*		*46*	*44*	*48*	*47*	*48*
	r	.16	.17	.07	.11	.08	.29	.37		.48	.45	.50	.45	.49
9	*N*	*46*	*28*	*28*	*45*	*45*	*45*	*46*	*46*		*44*	*45*	*44*	*45*
	r	-.06	-.05	.26	.17	.28	.34	.29	.52		.71	.60	.67	.58
10	*N*	*44*	*28*	*28*	*43*	*43*	*43*	*44*	*44*	*44*		*43*	*42*	*44*
	r	.13	.07	.20	.46	.71	.46	.33	.16	.42		.52	.67	.67
11	*N*	*49*	*28*	*28*	*48*	*48*	*48*	*49*	*48*	*45*	*43*		*47*	*48*
	r	.07	.19	.05	.25	.38	.28	.31	.49	.63	.59		.39	.39
12	*N*	*47*	*27*	*27*	*46*	*46*	*46*	*47*	*46*	*43*	*41*	*46*		*47*
	r	.23	.35	.31	.03	.24	.24	.02	.31	.23	.23	.19		.62
13	*N*	*49*	*29*	*29*	*49*	*49*	*49*	*49*	*48*	*45*	*44*	*48*	*46*	
	r	.08	.19	.31	.26	.34	.34	.37	.33	.63	.43	.35	.58	

Table 11
Intercorrelations of Total Response Latency for 13 Problems

Problem		Problem 1	2	3	4	5	6	7	8	9	10	11	12
2	*N*	*33*											
	r	.48											
3	*N*	*33*	*33*										
	r	.51	.55										
4	*N*	*54*	*33*	*33*									
	r	.17	.46	.52									
5	*N*	*54*	*32*	*32*	*53*								
	r	.20	.30	.45	.58								
6	*N*	*54*	*32*	*32*	*53*	*54*							
	r	.25	.36	.59	.30	.35							
7	*N*	*54*	*32*	*32*	*53*	*53*	*53*						
	r	.22	.29	.25	.28	.23	.33						
8	*N*	*50*	*30*	*30*	*49*	*49*	*49*	*50*					
	r	-.07	-.02	-.12	.46	.61	.23	.20					
9	*N*	*46*	*28*	*28*	*45*	*45*	*45*	*46*	*46*				
	r	-.01	-.27	.18	.04	.22	.45	.19	.11				
10	*N*	*44*	*28*	*28*	*43*	*43*	*43*	*44*	*44*	*44*			
	r	.20	.05	.30	.08	.40	.26	.35	.13	.42			
11	*N*	*49*	*28*	*28*	*48*	*48*	*48*	*49*	*48*	*45*	*43*		
	r	.30	.46	.27	.46	.47	.04	.42	.19	.18	.41		
12	*N*	*43*	*28*	*28*	*47*	*47*	*47*	*48*	*47*	*44*	*42*	*47*	
	r	.11	.13	.31	-.01	.13	.20	.04	-.00	.42	.65	.13	
13	*N*	*49*	*29*	*29*	*49*	*49*	*49*	*49*	*48*	*45*	*44*	*48*	*47*
	r	.11	.16	.33	.26	.32	.23	.18	.14	.41	.47	.18	.51

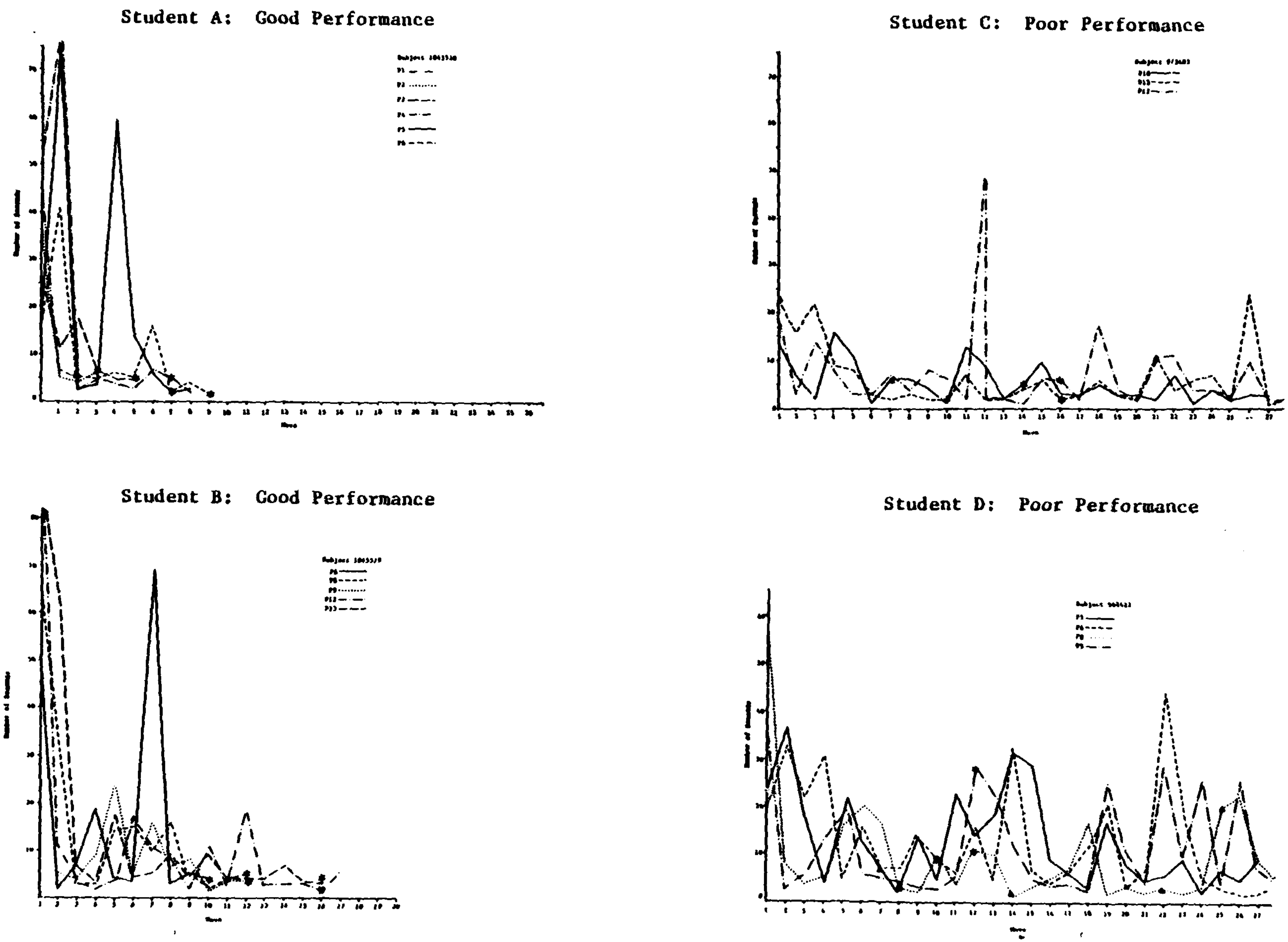
ences in the speed of spatial information processing, which in this case represents the efficiency with which a sequence of moves can be traced out visually and maintained in memory. Such differences may or may not show up in the performance measures, since students may compensate for slower information processing speeds with more care and slower response latencies.

<u>Latency trends</u>. Figure 2 shows plots of response latencies in seconds (vertical axis) versus the numbered moves (horizontal axis) for sampled problems for two students who performed very well on the test as a whole and for two students who performed poorly, based on MRATE. In each graph an * indicates where the plot would have ended had the problem been solved in the optimal number of moves. Graphs which continue beyond the 27th move at the right end of the horizontal axis were not solved by the student.

The graphs shown here suggest that good problem solvers (Students A and B) had larger initial study times for Move 1. Although this seemed to be the case for some of the good problem solvers, typical initial study times for other good problem solvers indicated that this was not a consistent trend. Most of the latency graphs examined did seem to be characterized as follows:

1. Generally, initial latencies were longer than the latencies for subsequent moves.
2. "Spikes" in the graphs frequently occurred every several moves, indicating that the student was restudying the problem

Figure 2
Latency Trends for Two Good Test Performers and Two Poor Test Performers
Based on Judges' Performance Ratings (MRATE)



and/or evaluating his or her progress. Although not analyzed systematically, some student's graphs (e.g., Student A in Figure 2) seem to be characterized by higher spikes than others.

3. For problems that were solved, latencies typically dropped to 2 to 4 seconds for the last 3 to 6 moves, indicating that the solution path had been discovered. This finding may be consistent with short-term memory capacity research, which indicates that somewhat fewer than seven "chunks" of information can be maintained in short-term memory while other cognitive resources are being allocated simultaneously (Kintsch, 1977, p. 199).
4. Poorly solved problems often showed a conspicuous absence of spikes or restudy points. In Figure 2 Problems 10 and 12 for Student C and Problem 8 for Student D exemplified this point. On the other hand, there were problems solved poorly which did contain spikes or restudy points (e.g., Problem 13 for Student C), indicating that the student was trying to get back on the right track.

Overall, some trends were suggestive, but they were by no means universal. Although perhaps providing clues to the work styles of some students (e.g., impulsive responding with few, if any, study points), the latency trends appeared to be too idiosyncratic to be very useful from a psychometric point of view.

Relationship between Performance and Response Latencies

The correlations between the number of moves students used and the initial and average move latencies for each problem indicated no relationship between these latency measures and performance with a single problem. Similarly, when initial and mean latencies for each problem were correlated with total scores (Total 2) and MRATE, no significant correlations were found (see Appendix Tables F-5 and F-6).

Not surprisingly, problems that were not solved well took longer than problems solved well, as indicated by the first row of Table 12, which shows the correlation of total time spent on each problem with the number of moves needed (and, hence, the individual problem scores Score 1 and Score 2). This relationship held for all problems except Problems 1, 3, and 12; comparison with the difficulty index in Table 2

Table 12
Product-Moment Correlations Between Total Time Spent on Each Problem and Performance Measures, by Problem

Performance Measure	Problem 1	2	3	4	5	6	7	8	9	10	11	12	13
Individual Problem Score	05	61**	09	73**	62**	42**	75**	67**	49**	39**	63**	27	40**
Total 2	-12	24	-09	35**	34**	09	41**	35**	13	02	28**	-14	-17
MRATE	12	-28	03	-31	-33**	-14	-37**	-31*	-30*	-08	-27	-04	00

Note. Decimal points are omitted.
*Statistically different from zero at $p \leq .05$.
**Statistically different from zero at $p < .01$.

shows that the relationship was strongest for problems of middle difficulty levels (Problems 4, 5, 7, 8, and 11). When total problem time for each problem was correlated with students' total test performance, as indexed by Total 2 and MRATE, these same problems related most highly.

These data indicate that with the exception of the total latency, or time, spent on a problem, the response latencies did not show any consistent meaningful relationship to performance.

Motivational and Biographical Data

Table 13 shows the frequency and percentage of students endorsing various response alternatives to questions about prior experience, perceived difficulty, motivation level, and self-evaluation. Regarding prior experience with this problem type, Question 1 indicates that 40% of the students had never worked on this problem type, 58% had done so a few times, and only 2% had worked such problems many times.

Describing how students are to solve these problems and enter their moves in a sequence of computerized instructions has certain difficulties, but the responses to Question 2 indicate that nearly all students had little or no difficulty understanding the instructions. Most students thought half or more than half of the problems were rather easy (Question 3), were not at all or only slightly nervous (Question 6), and either enjoyed working on the problems or were neutral about it (Question 8). Responses to Question 4 suggest that the instructional sequence and experimental conditions did not succeed in motivating most students to try hard to solve all of the puzzles in as few moves as possible. This less than optimal motivation under conditions where the test has no particular importance to the student is probably more of a problem for tests of this type than for more traditional psychometric measures, since each item or problem requires more perseverance.

It is difficult to say how much the scores in this study were affected by some students being less concerned about optimal performance. However, to examine this question with the data available, the mean total score (Total 2) and MRATE of students responding to Question 4 with "a" (mean Total 2 = .96, mean MRATE = 5.59), "b" (mean Total 2 = 1.02, mean MRATE = 4.93), and "c" (mean Total 2 = 1.03, mean MRATE = 4.99) were compared and no significant differences found.

Question 5 indicates that about half of the students thought the length of the test affected their motivation. Finally, 56% of the students thought they did fairly well on the test, 30% thought they did not do very well, and 10% had no idea how well they had done (Question 7). For future research with this type of test, it would be of interest to have the computer ask some of these questions during actual testing so that students' motivation, anxiety, difficulty perception, and confidence could be related to the simultaneous quality of their solutions.

It is important to know to what extent a test measures prior experience with the assigned tasks. Differences in test performance due to prior experience may be desirable or undesirable depending on the

Table 13
Distributions of Responses to Puzzle Reaction Questions (*N*=50)

Question	*N*	%
1. Before today, how often have you worked on this kind of puzzle?		
a. never	20	40
b. a few times	29	58
c. many times	1	2
2. How much difficulty did you have understanding the instructions before starting the puzzles?		
a. no difficulty	39	78
b. a little difficulty	10	20
c. much difficulty	1	2
3. Which of the following best describes how difficult you thought the puzzles were?		
a. All of the puzzles were easy	3	6
b. A few puzzles were difficult, the rest were rather easy	27	54
c. About half the puzzles were easy and half were difficult	15	30
d. A few puzzles were easy, the rest were rather difficult	5	10
e. All of the puzzles were difficult	0	0
4. Which of the following best describes your attitude towards completing the puzzles?		
a. I tried hard to solve all puzzles in as few moves as possible	18	36.7
b. I tried hard to solve most but not all of the puzzles in as few moves as possible	19	38.8
c. I tried to solve the puzzles, but was not very concerned about using as few moves as possible	12	24.5
d. I didn't care whether I solved the puzzles or not	0	0
5. Did the length of the test affect your motivation?		
a. not at all	19	38
b. somewhat	26	52
c. quite a bit	5	10
6. Were you nervous or uncomfortable while working on the puzzles?		
a. not at all	33	66
b. somewhat	17	34
c. very much so	0	0
7. How well do you think you did on the puzzles?		
a. very well	2	4
b. fairly well	28	56
c. not very well	15	30
d. I don't really know	5	10
8. How did you feel about working on the puzzles?		
a. I disliked it a lot	3	6
b. I disliked it somewhat	4	8
c. I felt neutral about it	11	22
d. I enjoyed it somewhat	26	52
e. I enjoyed it a lot	6	12

test application. In this study, a general measure of spatial reasoning ability was sought so that performance scores would not be significantly determined by prior experience with any specific spatial task. A comparison of the mean Total 2 score (.96) and performance ratings (5.55) for the 20 students who reported no prior experience with this problem type and of the mean Total 2 score (1.04) and performance ratings (4.73) for the 30 students who reported having worked such problems a few or many times (see Question 1) showed no significant performance differences based on stated prior experience. Similarly, a comparison of male and female mean Total 2 scores (1.00 versus 1.00) and mean ratings (5.09 versus 5.25) also showed no statistically significant differences.

Perceived Difficulty Ratings

Dimensions of Perceived Difficulty

Table 14 shows the proportion of students reporting voluntary use of various rating dimensions in their protocols while sorting the stimuli and the proportion of students selecting each dimension from a prepared list of dimensions provided by the experimenter after the sorting was completed (see Appendix D for the rating booklet). The last column in Table 14 shows the percentage distribution of frequencies with which each of the dimensions in the prepared list was used. Table 14 shows that all dimensions were reported less frequently in the free response voluntary protocol situation than when the prepared list was used. This would be expected, since some students might not have thought to report a dimension they might recall using when prompted later. However, the large discrepancy between these two columns for Dimensions h (number of columns not matching) and i (number of rows not matching) would suggest that these two dimensions were not very salient, despite the high proportion of students endorsing these dimensions post hoc. The number of students endorsing the supposedly irrelevant Dimensions j and l under the prepared list conditions, compared to the near absence of these dimensions in the volunteered protocols, further suggests that something like social desirability responding was occuring in the prepared list condition.

An examination of the percentage distribution data in the last column indicates these less relevant dimensions were most often reported as being used in Some or None of the problems. It seems likely that if students endorsed prepared dimensions that had not actually been used or that were not the most salient, they would endorse the Some category rather than the All or Most categories. On the other hand, the dimensions reported as being used most often in the voluntary protocols were, with the exception of Dimension c, endorsed most heavily in the All or Most categories in the prepared list. Thus, the data from the voluntary protocols, in conjunction with the All and Most categories in the frequency ratings, would seem to be the best indicators of the most salient rating dimensions that students thought they were using.

From Table 14 it is clear that the most salient rating dimension was Dimension a, the number of moves required to solve the puzzle (i.e., the solution path length). Ninety-three percent of the students voluntarily reported this dimension in some form in their protocols,

Table 14
Dimensions Used in Rating Perceived Difficulty

	Percentage of Students Reporting Using the Dimension on at Least *Some* of the Problems		Percentage of Time Students Reported Using the Dimension in Rating the Problems			
Dimension	Voluntary Protocols	Prepared List	All	Most	Some	None
a. The number or explication of moves	93	98	35	35	28	2
b. Whether can "see" solution	68	100	58	28	14	0
c. Number of squares not matching	58	91	26	30	35	9
d. Amount of time to solve	50	93	53	26	14	7
e. Types of moves required	50	-	-	-	-	-
f. How far apart certain numbers were	43	86	14	30	42	14
g. How much thought required	32	-	-	-	-	-
h. The number of columns not matching	18	72	19	23	30	28
i. Number of rows not matching	11	81	21	26	35	19
j. Location of empty space in left pattern	7	63	16	19	28	37
k. Similarity to already solved puzzle	2	-	-	-	-	-
l. Whether one pattern was in numeric order from 1 to 15	0	39	2	2	35	60

Note. Missing entries are for dimensions not included in the prepared list but reported by some students in their voluntary protocols.

and virtually all students (98%) selected the dimension in the prepared list condition. The other most salient dimensions were Dimension b, whether the student could "see" the solution or not; Dimension c, the number of squares not matching in the two patterns; Dimension d, the time the student felt it would take to solve the puzzle; Dimension e, the type or nature of moves required; and Dimension f, how far apart certain numbers were in the two patterns. The relative rank ordering or salience of these dimensions would be difficult to justify, since they are not independent, and a student reporting the number of squares not matching in his or her protocol could have been taking the distance between squares into account as well, without explicitly reporting this dimension.

A further question can be raised as to whether some of these reported dimensions are really rating dimensions underlying difficulty judgments or are actually synonymous with difficulty itself. This would seem to be the case with Dimensions b and d in Table 14. If students had been asked to rate "whether they could readily see the solution" or "how much time it would take to solve the puzzle," the rating task might be equivalent to rating the difficulty; and such physical problem characteristics of each puzzle as path length and the distance between various numbers would probably underly these judgments as well. It would seem, then, that the dimensions most important for students in evaluating the difficulty of these problems were the solution path length, the number of squares not matching in the two patterns, and the distance dimension of how far apart certain squares were in the two patterns. Since no dimension was used for all problems, it seems likely that the relative importance of each dimension varied somewhat for each problem, depending on the particular pattern configurations to be compared.

Individual Differences in Mean Perceived Difficulty

Table 15 summarizes the mean difficulty ratings for each of the four 15-puzzle problem sets separately. These data show that there were substantial individual differences in the level and variation of difficulty perceptions, even for the same problems. For example, for Stimulus Set 1, although the average student thought the problems were Easy or Somewhat Easy, one student thought the average problem in the set was Very Easy and another thought the average problem was Somewhat Difficult. Individual differences in perceived difficulty of the problems within stimulus sets was also evidenced, since about two-thirds of the students utilized all six rating categories, but about one-third utilized only the four easiest categories, and one student rated all stimuli with the two easiest categories. Without data for the same students on an independent rating task irrelevant to the difficulties rated here, it is not possible to determine to what extent these individual differences reflect response biases in the use of category rating scales; but it seems reasonable to assume that the differences found do indicate some true perceptual differences in perceived difficulty. Presumably, these differences reflect individual differences in the ability to visualize and to maintain a sequence of moves in short-term memory.

Table 15
Individual Differences in Mean Difficulty Perception

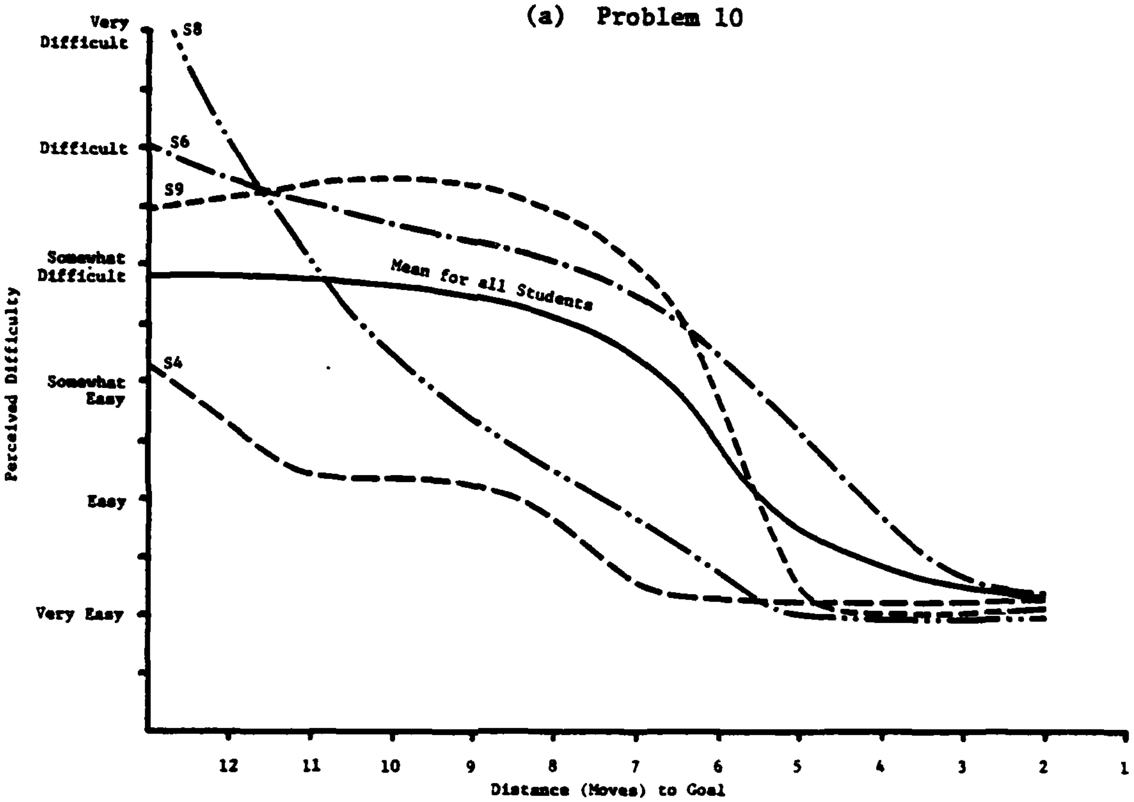
Stimulus Set	Individual Mean Ratings: Lowest	Mean	Highest
1	1.12 Very Easy	2.57 Easy/Somewhat Easy	3.77 Somewhat Difficult
2	1.94 Easy	3.13 Somewhat Easy	3.82 Somewhat Difficult
3	1.69 Easy	2.63 Somewhat Easy	3.44 Somewhat Easy/ Somewhat Difficult
4	1.76 Easy	2.86 Somewhat Easy	4.12 Somewhat Difficult

Perceived Difficulty and Number of Moves

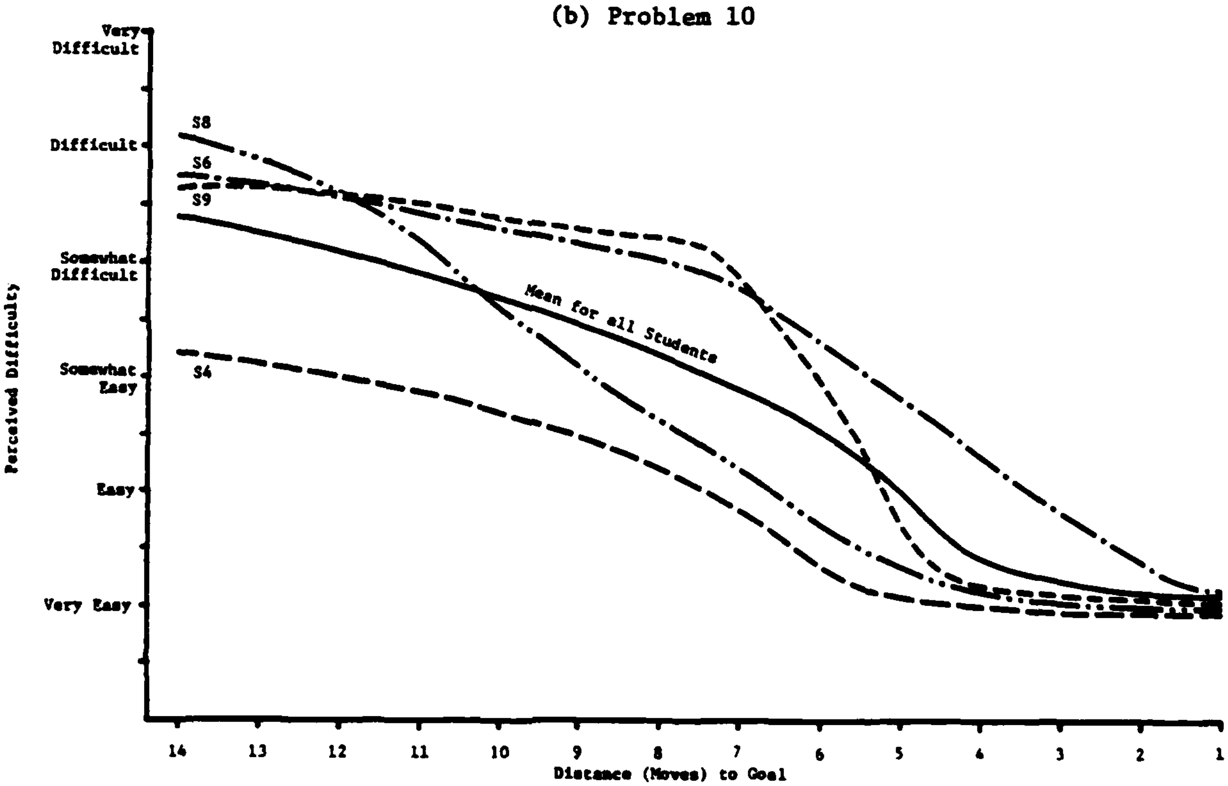
That the obtained individual differences in perception seem to be reliable is suggested by the data in Figure 3, which shows the perceived difficulty ratings of four students within Problems 9 and 10 as the distance in moves from the start puzzle configuration approached the goal puzzle configuration. These graphs were obtained by having students rate the difficulty of reaching the goal, not only from the start configuration, but from various intermediary configurations between the start and goal configuration. Thus, for example, in Figure

Figure 3
Perceived Difficulty Within Problems 9 and 10 for 4 Students

(a) Problem 10



(b) Problem 10



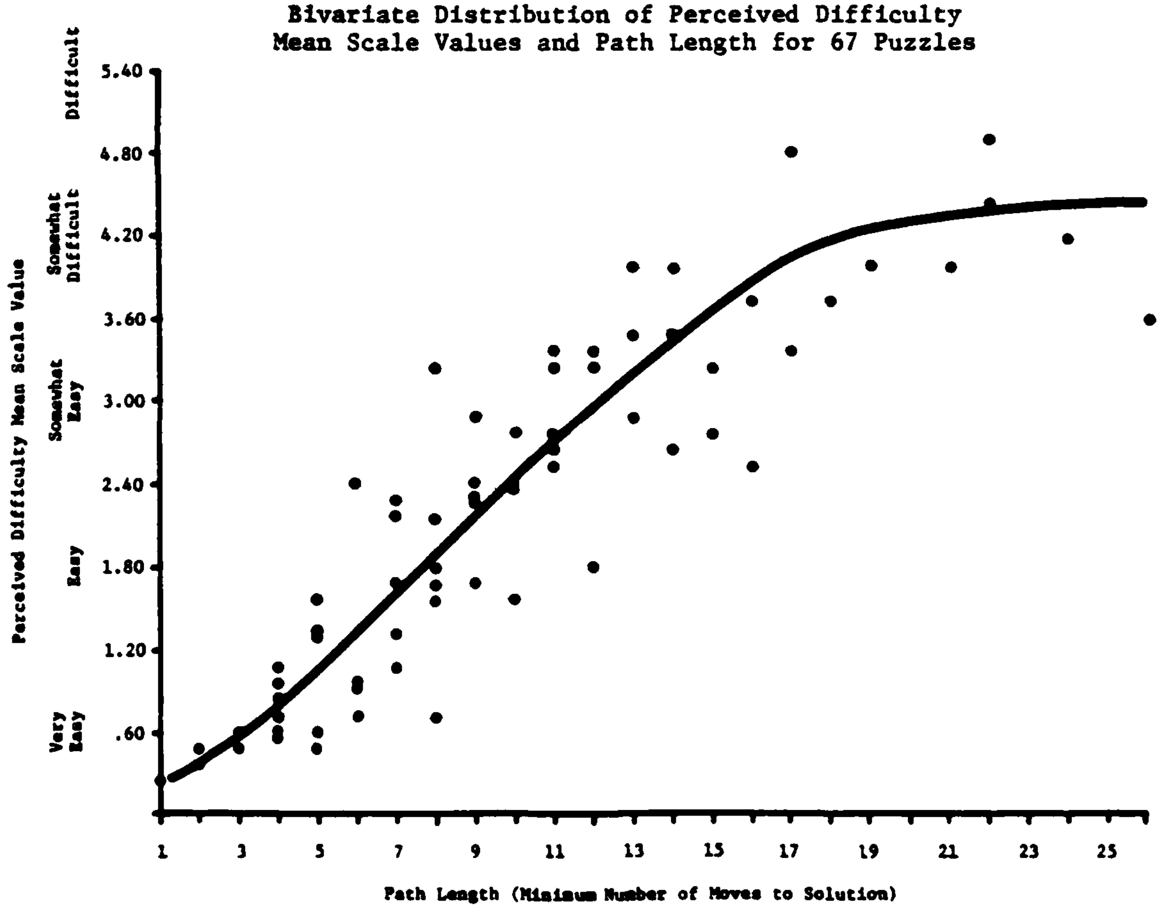
3a it might be presumed that if Student 9 were actually attempting to solve Problem 9, the puzzle would look Somewhat Difficult to him or her until he or she was about 7 moves away from the goal; then difficulty would drop off rapidly until he or she was 4 or 5 moves from the goal, when the puzzle would appear to be Very Easy.

Note that across both the problems shown in Figure 3, the four students show marked consistency in how they perceived the difficulty of different puzzle distances. For example, Student 4 perceived both problems as easier than the mean student at all distances from the goal, whereas Student 6 perceived both problems as more difficult than the mean student did at all distances from the goal. Even though only a few examples of students and problems are shown in Figure 3, this tendency for reliable individual differences in difficulty perceptions was present in nearly all combinations of students and problems examined. These data suggest that if the differences in difficulty perceptions relate to performance, then reliable individual performance differences in solving these problems should be obtainable.

Relationship of Difficulty Perception and Path Length

Since path length seemed to be a dominant dimension in the student protocols, difficulty perception scale values were correlated and plot-

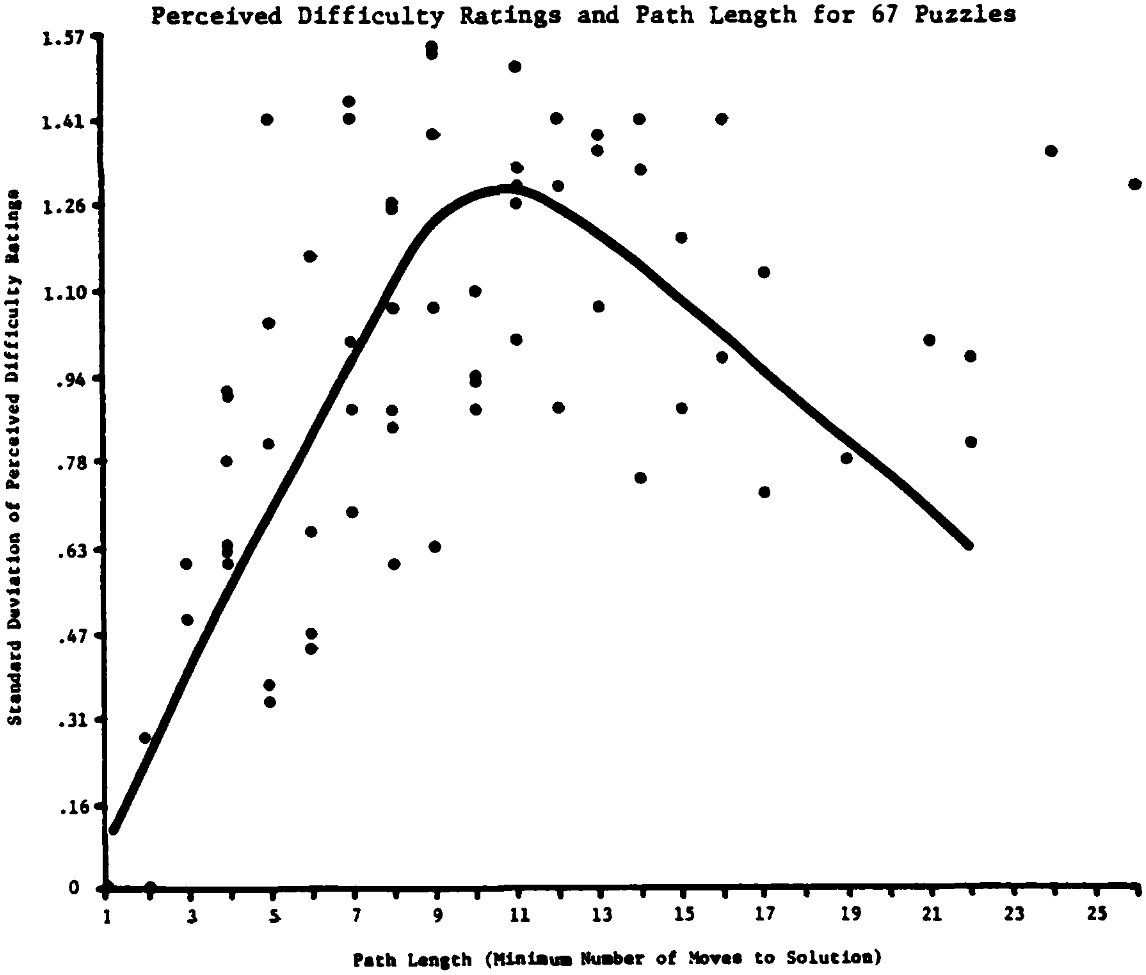
Figure 4
Bivariate Distribution of Perceived Difficulty
Mean Scale Values and Path Length for 67 Puzzles



ted against path length for all 67 puzzles. Figure 4 shows the scatterplot relating solution path length of each puzzle to its mean scale value. Although the correlation between the two variables was .88, the relationship between the two variables was not strictly linear at the right, or high, end of the plot. Although end effects must always be considered in category rating scales, the fact that students could have assigned higher ratings at the high end of the curve would suggest that the flattening of the curve for long path lengths represents a real effect. Students apparently could not discriminate differential path lengths greater than about 16. Perhaps a secondary rating dimension, such as the distance between numbers in the pattern or the number of squares not matching in the two patterns, is important in differentiating problems with longer path lengths.

Figure 4 also provides estimates of how difficult puzzles with different path lengths will appear to the average student when beginning work on a problem. A puzzle perceived to be Very Easy would correspond to a value on the vertical axis in Figure 4 between .5 and 1.5; Easy puzzles would range from 1.5 to 2.5; Somewhat Easy, from 2.5 to 3.5; Somewhat Difficult, from 3.5 to 4.5; Difficult, from 4.5 to 5.5;

Figure 5
Bivariate Distribution of Standard Deviations of Perceived Difficulty Ratings and Path Length for 67 Puzzles



and Very Difficult, from 5.5 to 6.5. Solution path lengths corresponding to the difficulty categories overlapped somewhat with Very Easy ratings corresponding to puzzles requiring 1 to 7 moves each, Easy for puzzles of 4 to 10 moves, Somewhat Easy for puzzles ranging from 6 to 16 moves, Somewhat Difficult for puzzles requiring 8 to 18 moves, and Difficult for puzzles with from 16 to 26 moves. None of the puzzles used here, which ranged from 1 to 26 moves, were rated Very Difficult by the average student.

Figure 5 shows a plot of solution path lengths versus the standard deviations of the students' category ratings. These data demonstrate that although students tended to agree more in their difficulty perceptions for stimuli with short or very long solution paths, there was substantial disagreement in perceived difficulty for puzzles with path lengths in the middle range, with a peak disagreement for solution paths of about 10 moves.

DISCUSSION AND CONCLUSIONS

Problem Characteristics

The data suggested that four performance indices might be useful in indexing problem difficulty: (1) the mean number of moves in the sample, (2) the proportion of students solving a problem, (3) the proportion of students solving a problem in the optimal number of moves, and (4) the Special Difficulty Index. These four indices showed substantial agreement in rank ordering the difficulty of the problems.

Because it adjusts for differences in solution path length while also taking into account the average number of moves required by the sample, the Special Difficulty Index not only appeared to be the best index of problem difficulty but also correlated lower with the solution path length of each problem than the other performance indices used to estimate problem difficulty. This is a desirable situation, since longer puzzles were not always the most difficult. Future research with this problem type should consider use of some short, but less direct or obvious, problems.

The number of illegal and repeated moves were found to be too low and not consistent enough for individuals across problems to be useful performance indices, at least for this problem set and sample.

Examination of problem performance indices, the Special Difficulty Index, and students' perceptions of the difficulty of the test problems indicated that with the exception of Problems 9, 10, 12, and 13, the problems were too easy for most students. For example, except for these four problems, 70% or more of the students solved each of the remaining problems in the minimum number of moves. It seems likely that these highly skewed distributions of number of moves to completion precluded high correlations of individual performance indices across problems, since small absolute differences in scores across problems would be accentuated. Thus, the consistency across problems of the number of moves to completion was generally poor, with indications of only small to moderate consistency for clusters of problems of similar difficulty. It is possible that if a more difficult set of problems that were more similar in difficulty levels were administered, better measures of consistency of performance would be obtained. The item-total score correlations obtained for each problem suggested that it would be possible to obtain a more discriminating subset of problems. Because this was an exploratory study, however, no preselection of problems was possible. Since the data suggest that better consistency may be obtained using problems of similar difficulty levels, an adaptive test, which tailors problems to the ability level of each student, should increase the reliability of measurement.

Scoring Methods

Four alternative methods of scoring total test performance and two methods of scoring individual problem performance were studied. The scores that took into account differential numbers of moves (Total 1, Total 2) between the optimal and maximum number allowed appeared to be the best, on intuitive grounds, and were also related somewhat more to judges' performance ratings. The Total 2 score, which also took into account the difficulty of the problems the student attempted, appeared to be the most meaningful score. Where other methods rank ordered students differently, the rank ordering provided by Total 2 was most highly related to judges' performance ratings. Although Total 2 may appear to be additive in that it averages individual problem scores (Score 2), the pattern or configuration of individual problem performance is taken into consideration, since the individual problem scores (Score 2) are adjusted for the difficulty of each problem, as reflected in the mean performance of the sample on the problem. As a result, students are penalized more for poor performance on easier problems, relative to the group, than they are on more difficult problems. In this way students who solve the same number of problems but have different patterns of performance will obtain different Total 2 scores.

Future research with this problem type will require study of the validity of the various performance scores against relevant external criteria. Since no such reliable criteria were available in this study, the meaningfulness of the scores was tentatively determined by comparing these objective scores with judges' performance ratings of test performance. Strong indications of concurrent validity were found. Those cases in which the objective score ordered students differently than the ratings indicated that whereas the objective score (Total 2) penalized students more than judges' ratings for poor performance on easier problems, the judges penalized students more for not attempting some problems (although this was not always the student's fault) and for doing poorly on more difficult problems. Although it is difficult to determine which measure is more valid without an external criterion, the high correlations between the objective scores and the judges' ratings suggest some validity in both types of data.

Latencies

Mean initial and total latencies for each problem were strongly related to some of the performance indices of problem difficulty. That is, the group as a whole utilized longer initial study times and longer total work times on more difficult problems. Similarly, problems that took longer to solve were initially studied longer. The average latency of moves within a problem did not relate to problem difficulty.

At the level of individual performance, only total latency or problem solution time was related to problem performance. Some good problem solvers were characterized by very long initial latencies, but this tendency was not universal. Many good problem solvers did not intially study the problem longer than did the average poor problem solver. The average problem response latency measure did not relate to individual student performances.

Plots of latency trends across problems were interesting from a descriptive point of view in indicating that most students' trends showed longer initial latencies followed by a few quicker moves, occasional spikes indicating re-evaluation of progress, and finally several very quick final moves indicating that the sequence of moves to solution had been detected. However, no universal trends in response latencies seemed to characterize good problem solvers versus poor problem solvers well enough to be useful in scoring or predicting individual performance. Latencies in this study seemed to confound differences in the ability to visualize a sequence of moves and differences in students' work styles. Strong evidence for such work styles was found in the consistency of initial, average, and total response latency measures across all problems. Students who took longer initial study times, longer average times between moves, and longer total work times on one problem showed a consistent tendency to do so on other problems as well.

Thus, while the response latency measures were predictive of problem difficulty and indicated the existence of consistent styles of problem-solving behavior, they did not appear to be useful in scoring individual performance.

<u>Motivational and Biographical Correlates of Performance</u>

Although the posttest reaction questionnaire indicated that only 40% of the students had never worked problems of this type before, mean performance scores between these students and those who had previously worked such problems were not significantly different.

Only 36.7% of the students reported trying hard to solve all the problems in the minimum number of moves. Slightly more students said they tried hard to solve most, but not all, of the problems. Although mean performance differences between subgroups reporting different levels of motivation were not significantly different, these data plus the fact that 52% of the students felt their motivation was affected by the length of the test indicate that total testing times may need to be shorter for this type of task than for tests with more conventional item formats.

No sex differences in performance were found on this test. That males typically show better spatial ability (Garai & Scheinfeld, 1968; MacCoby & Jacklin, 1974) and restructuring ability (MacCoby, 1966; Sweeney, 1953; Terman & Tyler, 1954; Tyler, 1965) would seem to predict male superiority on this test. On the other hand, females have generally been found to be less impulsive (MacCoby,1966; Terman & Tyler, 1954; Tyler, 1965) and better in perceptual speed and fluency (Garai & Scheinfeld, 1968). The failure to obtain sex differences with this type of task will only be of concern once more reliable measurement is achieved. At that time, hypothesized correlates of these problems should be examined to determine whether scores index spatial reasoning, restructuring, impulsivity, or some other psychological variable.

Dimensions of Perceived Difficulty

The most salient dimensions of perceived difficulty were the number of moves required to solve the puzzle, the number of squares not matching in the two patterns, and the distance dimension of how far apart certain squares were in the two patterns. Since no dimension was reported as having been used for all problems, it seems likely that the relative importance of each dimension varied somewhat for each problem, depending on the particular pattern configurations.

When the actual values of these dimensions were computed for the problems used in the computer-administered test (see Appendix Table C), a hypothesized rank ordering of problems by difficulty was obtained. These three rank orders were quite similar ($.51 < p < .79$) but were not as consistent as the rank orderings for difficulty obtained from performance indices such as mean number of moves, proportion of students solving the problem, and the Special Difficulty Index (see Table 3). Thus, although these physical dimensions may be useful as a tentative index of problem difficulty for use in initial problem selection prior to data collection, the performance measures should provide more precise indices of difficulty once normative data can be obtained.

The actual perceived difficulty ratings showed substantial individual differences in the level and variability of difficulty perceptions, even for the same set of problems. Although possible individual biases in the use of category rating scales cannot be discounted, the data suggest that the individual differences found were differences in subjective difficulties relating to individual differences in ability to visualize and to maintain a sequence of moves in short-term memory. Examination of individual difficulty perceptions across problems indicated that these differences were reliable. These data suggest that if the reliable differences in difficulty perceptions do in fact relate to differential ability to visualize successful move sequences, then an adequate selection of problem replications should be able to tap these differences, resulting in reliable performance differences.

Comparison of the easy problems with the problems that challenged students more in the computer-administered test suggested that too many of the problems could be solved in a reactive manner, that is, by responding to the immediate stimulus pattern without trying to visualize or to plan several moves ahead. Such problems would not tap differences in students' ability to visualize a sequence of moves because students would not find themselves in a difficult situation by not planning ahead. The more challenging problems (e.g., Problems 9, 10, 12, and 13) were those in which a student could get "in trouble" by not visualizing several moves in advance (see Appendix C). This implies that future studies should include more problems that prevent reactive solutions, i.e., require more planning ahead.

Comparison of the mean perceived difficulty of the problems included in the computer-administered test indicated less agreement with actual problem difficulty than might be expected from other studies.

This appeared to be due to the inability of students to differentiate the relative difficulties of problems with longer solution paths. Thus, to the extent that increased motivation under adaptive testing depends on correct student perceptions of problem difficulty (Prestwood & Weiss, 1977), adaptive administration of this problem type may not have a motivational advantage. On the other hand, reduced frustration would seem likely to result under adaptive conditions from not requiring students to work on problems much more difficult than their ability levels, even if they cannot accurately perceive the actual difficulty of the problem beforehand.

The perceived difficulty scale values related highly (r = .75) to the mean initial response latency measure for the computer-administered problems. This supports the idea that the students spend time before their first move trying to visualize a sequence of moves, since path length appeared to be a primary rating dimension in determining perceived difficulty.

Conclusions

The results from this pilot study suggest certain improvements in problem selection and design. Future tests of this type should consist of fewer but more difficult problems, particularly those which do not permit reactive, impulsive solutions. If individual differences in the ability to construct an optimal sequence of moves are to be tapped, then more problems must be designed that force the student to plan ahead. More complex problems should overload the memories of students and should induce differences in strategies in manipulating the number patterns.

If reliable performance indices can be obtained, the process of validating the meaning of the scores will be necessary. Do scores reflect individual differences in spatial reasoning and problem-solving ability or in personality variables like perseverance and impulsivity? It might also be of interest to determine what information-processing abilities underly performance on these problems. For example, using Carroll's (1974) provisional coding scheme for cognitive tasks appearing in psychometric tests, the following cognitive operations might be expected to underly performance: (1) mental rotation of spatial configurations in visual short-term memory, Factors S and Vz; (2) performing serial operations in visual short-term memory, Factors S and Vz; and (3) storage in and retrieval from short-term memory, Factor Ms.

The results reported here suggest that reasonable indices of problem difficulty are obtainable given an appropriate norming sample. If reliable and valid ability scores can be obtained in future studies with this item type, this type of test would seem especially appropriate for adaptive administration, since (1) scores on problems tailored to the individual's ability are more apt to be more highly related to each other, resulting in total scores with higher reliability; (2) adaptive administration will likely improve the motivational aspects of the tests, which seem more taxing and potentially frustrating than conventional item formats; and (3) equally precise measurements for most testees can be obtained in shorter periods of time than with conventional test administration. Thus, the data suggest that future development of adaptive problem-solving tests of the type studied here might result in new types of ability tests that should provide ability scores to supplement those available from the paper-and-pencil administration of typical ability measures.

REFERENCES

Betz, N. E., & Weiss, D. J. An empirical study of computer-administered two-stage ability testing (Research Report 73-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1973. (NTIS No. AD 768993)

Betz, N. E., & Weiss, D. J. Empirical and simulation studies of flexilevel ability testing (Research Report 75-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, July 1975. (NTIS No. AD A013185)

Betz, N. E., & Weiss, D. J. Effects of immediate knowledge of results and adaptive testing on ability test performance (Research Report 76-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1976. (NTIS No. AD A027147) (a)

Betz, N. E., & Weiss, D. J. Psychological effects of immediate knowledge of results and adaptive ability testing (Research Report 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1976. (NTIS No. AD A027170) (b)

Bratfisch, O., Borg, G., & Dornic, S. Perceived item difficulty in three tests of intellectual performance capacity. Stockholm: University of Stockholm, Institute of Applied Psychology, 1972 (29).

Chiang, A., & Atkinson, R. C. Individual differences and interrelationships among a select set of cognitive skills. Memory and Cognition, 1976, 4, 661-672.

Clark, H., & Chase, W. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-247.

Cory, C. H. Relative utility of computerized versus paper-and-pencil tests for predicting job performance. Applied Psychological Measurement, 1977, 1, 551-564.

Cory, C. H., Rimland, B., & Bryson, R. A. Using computerized tests to measure new dimensions of abilities: An exploratory study. Applied Psychological Measurement, 1977, 1, 101-110.

Garai, J. E., & Scheinfeld, A. Sex differences in mental and behavioral traits. Genetic Psychology Monographs, 1968, 77, 169-299.

Ghiselli, E. E. The validity of aptitude tests in personnel selection. Personnel Psychology, 1973, 26, 461-477.

Hayes, J. R. Problem typology and the solution process. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 371-379.

Hunt, E., Lunneborg, C., & Lewis, J. What does it mean to be high verbal? *Cognitive Psychology*, 1975, *7*, 194-227.

Johnson, D. I., & Mihal, W. M. Performance of Blacks and Whites in computerized versus manual testing environments. *American Psychologist*, 1973, *28*, 694-699.

Johnson, E. S., & Baker, R. F. *Some effects of computerizing an experiment in human problem solving* (Research Report No. 105). Chapel Hill: University of North Carolina, The L.L. Thurstone Psychometric Laboratory, 1972.

Kagan, J. Reflection--impulsivity and reading ability in primary grade children. *Child Development*, 1965, *36*, 609-628.

Kagan, J., Rosman, B. L., Day, D., Albert, J., & Phillips, W. Information processing and the child: Significance of analytic and reflective attitudes. *Psychological Monographs*, 1964, *78* (Whole No. 58).

Kintsch, W. *Memory and cognition*. New York: John Wiley & Sons, 1977.

Maccoby, E. E. Sex differences in intellectual functioning. In E. E. Maccoby (Ed.), *The development of sex differences*. Stanford, CA: Stanford University Press, 1966.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences*. Stanford, CA: Stanford University Press, 1974.

Munz, D. C., & Jacobs, P. D. An evaluation of perceived item difficulty sequencing in academic testing. *British Journal of Educational Psychology*, 1971, *44*, 195-205.

Nilsson, J. J. *Problem-solving methods in artificial intelligence*. New York: McGraw-Hill, 1971.

Norman, D. A. *Memory and attention*. New York: John Wiley & Sons, 1976.

Pine, S. M., Church, A. T., Gialluca, K. A., & Weiss, D. J. *Effects of computerized adaptive testing on black and white students* (Research Report 79-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1979. (NTIS No. AD A067928)

Prestwood, J. S., & Weiss, D. J. *Accuracy of perceived test-item difficulties* (Research Report 77-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, May 1977. (NTIS No. AD A041084)

Rose, A. M. *An information processing approach to performance assessment: Final report*. Washington, DC: American Institutes for Research, 1978.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, 1969, *30*, 276-315.

Sweeney, E. J. Sex differences in problem-solving (Technical Report 1). Stanford, CA: Stanford University, Department of Psychology, December 1953.

Terman, L. M., & Tyler, L. E. Psychological sex differences. In L. Carmichael (Ed.), Manual of child psychology (2nd ed.). New York: John Wiley & Sons, 1954.

Tinsley, H. E. A., & Weiss, D. J. Interrater reliability and agreement of subjective judgment. Journal of Counseling Psychology, 1975, 22, 358- 376.

Tyler, L. E. The psychology of human differences. New York: Appleton-Century-Crofts, 1965.

Vale, C. D. Strategies of branching through an item pool. In D. J. Weiss (Ed.), Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, November 1975. (NTIS No. AD A018675)

Vale, C. D., & Weiss, D. J. A simulation study of stradaptive ability testing (Research Report 75-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1975. (NTIS No. AD A020961)

Weiss, D. J., & Betz, N. E. Ability measurement: Conventional or adaptive? (Research Report 73-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, February 1973. (NTIS No. AD 757788)

Wertheimer, M. Productive thinking. New York: Harper & Row, 1959. (Originally published, 1945).

APPENDICES

Appendix A:

Diagnostic Error Messages Provided by Testing System

Illegal Moves:

18 IS NOT A NUMBER IN THE PATTERN. REMEMBER TO PUSH THE "SPACE" BAR FIRST IF THE NUMBER TO BE ENTERED CONTAINS ONLY ONE DIGIT.

10P IS NOT A CORRECT MOVE. THE LAST CHARACTER TYPED MUST BE AN L, R, U, OR D.

10 CAN NOT BE MOVED LEFT (RIGHT, UP, DOWN) FROM ITS PRESENT POSITION.

Maximum Move Limit Reached:

YOU HAVE REACHED THE MAXIMUM NUMBER OF MOVES ALLOWED FOR THIS PROBLEM. PLEASE CONTACT THE PROCTOR.

Computer Data File Error:

THE COMPUTER IS HAVING PROBLEMS. PLEASE NOTIFY THE PROCTOR. (ERROR 06 HAS OCCURRED. IERR IS -5).

Maximum Time Limit Reached:

IT MIGHT BE A GOOD IDEA TO GO ON TO THE NEXT PROBLEM. PLEASE CONTACT THE PROCTOR.

Appendix B: Instruction Screens

Screen 1

```
HELLO AND THANK YOU FOR YOUR PARTICIPATION IN THIS STUDY.
THE COMPUTER WILL SOON PRESENT YOU WITH A SERIES OF PUZZLES TO WORK ON,
BUT FIRST SOME INSTRUCTIONS WILL BE GIVEN TO BE SURE YOU UNDERSTAND
HOW TO USE THE TYPEWRITER KEYBOARD TO ENTER YOUR RESPONSES.
FOLLOWING THE INSTRUCTIONS YOU WILL BE GIVEN A PRACTICE PROBLEM TO
CLEAR UP ANY PROBLEMS YOU ARE HAVING. IN ADDITION, IF YOU HAVE
QUESTIONS AT ANY TIME ABOUT THE INSTRUCTIONS OR ANYTHING ELSE PLEASE
FEEL FREE TO CONTACT THE TEST PROCTOR.

YOU MUST REMEMBER TWO THINGS IN ORDER TO TALK TO THE COMPUTER:
   1. ONLY TYPE SOMETHING WHEN A MESSAGE ON THE SCREEN
      IN FRONT OF YOU TELLS YOU TO DO SO AND A QUESTION
      MARK (?) APPEARS.
   2. EACH TIME YOU TYPE A RESPONSE ON THE KEYBOARD
      THE COMPUTER DOES NOT RECEIVE IT UNTIL YOU PRESS
      THE "RETURN" KEY.
NOW THE FIRST THING YOU MUST DO IS FIND THE "RETURN"
KEY. THIS KEY IS THE LARGE RECTANGULAR KEY ON THE
RIGHT END OF THE KEYBOARD. PRESS THE "SPACE" BAR AT THE BOTTOM OF THE
KEYBOARD AND THE "RETURN" KEY TO CONTINUE THE INSTRUCTIONS.
?
*
```

Screen 2

```
       1  2  3  4                  1  2     3
       5  6  7  8                  5  6  7  4
       9 10    11                  9 10 11  8
      12 13 14 15                 12 13 14 15
IN EACH OF THE PUZZLES OF THE TYPE SHOWN HERE YOUR TASK IS TO
TYPE IN A SEQUENCE OF "MOVES" TO CHANGE THE PATTERN OF NUMBERS
ON THE LEFT UNTIL IT MATCHES THE PATTERN ON THE RIGHT. A "MOVE"
CONSISTS OF 3 TYPED CHARACTERS FOLLOWED BY THE "RETURN" KEY. THE
FIRST 2 CHARACTERS TELL THE COMPUTER WHICH "NUMBER" IN THE
PATTERN ON THE LEFT YOU WANT TO MOVE. THE THIRD CHARACTER
(WHICH YOU WILL BE TOLD ABOUT SHORTLY) TELLS THE COMPUTER WHAT
DIRECTION YOU WANT TO MOVE THE NUMBER.

IF THE NUMBER YOU WISH TO MOVE HAS 2 DIGITS YOU SHOULD TYPE
THE 2 DIGITS ON THE KEYBOARD. IF THE NUMBER YOU WISH TO MOVE
HAS ONLY 1 DIGIT YOU SHOULD TYPE THE SPACE BAR ONCE AND THEN THE
DESIRED DIGIT. THUS, THE TWO DIGIT NUMBERS 10 TO 15 CAN BE TYPED
IN DIRECTLY, WHILE THE 'SPACE' BAR MUST BE TYPED FIRST WITH
THE NUMBERS 1 TO 9.
PRESS THE "SPACE" BAR AND "RETURN" TO CONTINUE.
*
```

Screen 3

```
 1  2  3  4          1  2     3
 5  6  7  8          5  6  7  4
 9 10    11          9 10 11  8
12 13 14 15         12 13 14 15
```

AS MENTIONED ABOVE THE THIRD CHARACTER IN YOUR "MOVE" TELLS THE COMPUTER WHAT DIRECTION TO MOVE THE NUMBER IN THE LEFT PATTERN. NUMBERS CAN ONLY BE MOVED INTO THE SPACE IN THE SQUARE PATTERN WHICH IS NOT OCCUPIED BY A NUMBER. YOU TELL THE COMPUTER WHICH DIRECTION TO MOVE THE NUMBER BY TYPING ONE OF THE FOLLOWING 4 LETTERS:

- L - IF YOU WANT TO MOVE A NUMBER TO THE LEFT ONE SPACE
- R - IF YOU WANT TO MOVE A NUMBER TO THE RIGHT ONE SPACE
- U - IF YOU WANT TO MOVE A NUMBER UP ONE SPACE
- D - IF YOU WANT TO MOVE A NUMBER DOWN ONE SPACE

THUS, IN THE PATTERN SHOWN HERE THE FOLLOWING 4 MOVES ARE POSSIBLE: 10R, 11L, 14U, OR <SPACE BAR>7D. ANY OTHER MOVE WOULD BE ILLEGAL AND RESULT IN A REMINDER MESSAGE BEING PRINTED BY THE COMPUTER. FOR EXAMPLE, YOU COULD NOT TRY TO MOVE THE "11" SQUARE TO THE RIGHT ONE SPACE SINCE ALL MOVES MUST STAY WITHIN THE SQUARE PATTERN.
PRESS THE "SPACE" AND "RETURN" TO CONTINUE INSTRUCTIONS.

Screen 4

IF YOU HAVE MADE A LEGAL MOVE THE COMPUTER WILL AUTOMATICALLY AND VERY QUICKLY UPDATE THE PATTERN ON THE LEFT WHERE YOU ARE MAKING YOUR MOVES. IF YOUR MOVE IS NOT LEGAL A MESSAGE WILL BE PRINTED UNDER YOUR MOVE AND YOU SHOULD TRY AGAIN WHEN THE COMPUTER TELLS YOU TO DO SO.
IF YOU ARE HAVING DIFFICULTY UNDERSTANDING THE INSTRUCTIONS SO FAR PLEASE CALL THE PROCTOR. OTHERWISE PRESS THE "SPACE" BAR AND "RETURN" TO CONTINUE THE INSTRUCTIONS.
?
*

Screen 5

SUPPOSE YOU MAKE A MISTAKE TYPING SOMETHING INTO THE COMPUTER. YOU CAN CORRECT A MISTYPED CHARACTER AT ANY TIME BEFORE YOU PRESS THE "RETURN" KEY. BY PRESSING THE "BACKSPACE" KEY WHICH IS LOCATED IN THE TOP RIGHT CORNER OF THE KEYBOARD YOU WILL "ERASE" THE LAST CHARACTER YOU TYPED. TO "ERASE" THE LAST TWO CHARACTERS YOU TYPED PRESS THE "BACKSPACE" KEY TWICE AND SO ON.
AFTER PRESSING "BACKSPACE" THE CORRECT CHARACTER CAN THEN BE TYPED IN. REMEMBER TO PRESS THE "RETURN" KEY TO SEND THE CORRECTED CHARACTERS TO THE COMPUTER.

TO SEE HOW THE "BACKSPACE" WORKS TRY TYPING THE MOVE '14D' ON THE KEYBOARD. THEN CHANGE THE 'D' TO A 'U' BY PUSHING THE "BACKSPACE" KEY ONCE AND THEN THE CORRECT LETTER 'U'. FINALLY, PRESS THE "RETURN" KEY TO SEND YOUR CORRECTED MOVE TO THE COMPUTER.

?

Screen 6

YOU ARE NOW ALMOST READY TO BEGIN WORKING. FIRST, HOWEVER, WE NEED SOME INFORMATION ABOUT YOU.
THE RESULTS OF THE PROBLEMS YOU WILL WORK ON WILL BE STRICTLY CONFIDENTIAL. WE ARE INTERESTED IN YOU AS PART OF A LARGER GROUP, AND AT NO TIME WILL YOUR SCORES BE CONNECTED WITH YOUR NAME.
BUT WE NEED IDENTIFICATION SO THAT WE CAN KEEP YOUR ANSWERS SEPARATE FROM OTHER PEOPLE'S AND SO THAT WE CAN COMPARE THE RESULTS OF THESE SCORES WITH ANY OTHER DATA CONTRIBUTED BY YOU AT AN EARLIER OR LATER TIME.
PLEASE TYPE YOUR FIRST NAME (JUST YOUR FIRST NAME THIS TIME), AND THEN "RETURN".
?
*

Screen 7

PLEASE TYPE YOUR MIDDLE INITIAL (ONE LETTER ONLY).
IF YOU DO NOT HAVE A MIDDLE NAME, TYPE A "?".
DON'T FORGET TO PRESS "RETURN".
?
*

Screen 8

PLEASE TYPE YOUR LAST NAME AND PRESS "RETURN".
?
*

Screen 9

PLEASE TYPE YOUR SIX OR SEVEN DIGIT STUDENT IDENTIFICATION NUMBER AND "RETURN".
IF YOU DO NOT REMEMBER YOUR IDENTIFICATION NUMBER AND DO NOT HAVE IT WITH YOU CALL THE PROCTOR FOR A SUBSTITUTE IDENTIFICATION NUMBER.

?
*

Screen 10

NOW WE WOULD LIKE TO KNOW A FEW THINGS ABOUT YOU. IF THE QUESTION DOES NOT APPLY TO YOU OR YOU DON'T WANT TO RESPOND, TYPE IN A QUESTION MARK AND "RETURN".
PLEASE TYPE YOUR AGE AND PRESS "RETURN".
?
*

Screen 11

WHICH SEX ARE YOU?
    1. FEMALE
    2. MALE
TYPE THE CORRECT NUMBER AND PRESS "RETURN".
?
*

Screen 12

PLEASE TYPE THE NUMBER CORRESPONDING TO YOUR YEAR IN SCHOOL:
1. FRESHMAN
2. SOPHOMORE
3. JUNIOR
4. SENIOR
5. GRADUATE STUDENT
6. OTHER

DON'T FORGET TO PRESS "RETURN".
?
*

Screen 13

LISTED BELOW ARE SEVERAL OF THE COLLEGES WITHIN THE UNIVERSITY.
1. COLLEGE OF LIBERAL ARTS (CLA)
2. COLLEGE OF AGRICULTURE
3. COLLEGE OF BIOLOGICAL SCIENCES
4. COLLEGE OF BUSINESS ADMINISTRATION
5. COLLEGE OF EDUCATION
6. GENERAL COLLEGE
7. COLLEGE OF HOME ECONOMICS
8. INSTITUTE OF TECHNOLOGY
9. SCHOOL OF FORESTRY
10. UNIVERSITY COLLEGE
11. COLLEGE OF VETERINARY MEDICINE
12. GRADUATE SCHOOL
13. LAW SCHOOL
14. OTHER

PRESS THE NUMBER OF THE SCHOOL IN WHICH YOU ARE ENROLLED AND THE "RETURN" KEY.
?
*

Screen 14

WHAT IS YOUR RACE?

1. AFRO-AMERICAN (BLACK)
2. MEXICAN-AMERICAN
3. PUERTO-RICAN
4. OTHER LATIN AMERICAN
5. ORIENTAL OR ASIAN-AMERICAN
6. NATIVE-AMERICAN (INDIAN)
7. WHITE
8. OTHER

TYPE THE NUMBER THAT GIVES YOUR RACE, AND PRESS "RETURN".
?
*

Screen 15

```
IN WHICH CATEGORY IS YOUR CUMULATIVE GRADE-POINT AVERAGE (GPA)?
    1)  3.76 TO 4.00
    2)  3.51 TO 3.75
    3)  3.26 TO 3.50
    4)  3.01 TO 3.25
    5)  2.76 TO 3.00
    6)  2.51 TO 2.75
    7)  2.26 TO 2.50
    8)  2.01 TO 2.25
    9)  2.00 OR LESS
TYPE THE CATEGORY NUMBER ( "1" THROUGH "9" ) AND PRESS "RETURN".
?
*
```

Screen 16

```
YOU ARE NOW READY TO TRY A PRACTICE PROBLEM.
IN THE PRACTICE PROBLEM AND THE ACTUAL PROBLEMS TO FOLLOW
AN IMPORTANT GOAL IN TRYING TO MAKE THE PATTERN ON THE LEFT
MATCH THE PATTERN ON THE RIGHT IS TO DO SO WITH AS
FEW MOVES AS POSSIBLE. YOUR PERFORMANCE WILL BE DETERMINED
NOT ONLY BY WHETHER YOU ARE ABLE TO MATCH THE TWO PATTERNS
BUT ALSO BY HOW FEW MOVES IT TAKES YOU TO DO SO.
THERE IS NO TIME LIMIT ON ANY OF THE PUZZLES BUT TRY TO
USE YOUR TIME WISELY WHILE STILL TRYING TO USE AS FEW MOVES
AS POSSIBLE. TRY TO COMPLETE EACH PROBLEM. IF, HOWEVER, YOU
HAVE WORKED A LONG TIME ON A SINGLE PROBLEM AND FEEL YOU CAN NOT
SOLVE IT CONTACT THE PROCTOR WHO WILL GET THE COMPUTER TO
PRESENT THE NEXT PROBLEM.
A SUMMARY OF HOW TO TYPE IN YOUR THREE CHARACTER MOVE
AS DESCRIBED EARLIER WILL BE PRESENTED WITH EACH PUZZLE
AS A REMINDER.

IF YOU HAVE ANY QUESTIONS ABOUT WHAT YOU ARE SUPPOSED
TO DO CALL THE PROCTOR. OTHERWISE PRESS THE "SPACE"
BAR AND "RETURN" KEY TO BEGIN YOUR PRACTICE PROBLEM.
?
*
```

Appendix C:

Start and Goal Configurations Constituting the Problem-Solving Test, and Values of Various Physical Problem Characteristics

Pattern: Start				Pattern: Goal				Physical Problem Characteristics: Solution Path Length	No. of Squares Not Matching	No. of Rows Not Matching	No. of Columns Not Matching	Euclidean Distance	City-Block Distance
Problem 1													
1	2	3	4	1	2	3	4						
5		6	7	5	9	6	7	3	3	4	2	5.24	6
8	9	10	11	8	13	10	11						
12	13	14	15	12	13	14	15						
Problem 2													
1	2	3	4	1	2		3						
5	6	7	8	5	6	7	4	4	5	3	2	6.00	6
9	10		11	9	10	11	8						
12	13	14	15	12	13	14	15						
Problem 3													
2	11	4	8	2	4	8	12						
7		3	12	7	11	3	13	6	7	3	3	7.41	8
10	1	5	13	10	1		5						
14	6	15	9	14	6	15	9						
Problem 4													
1	2	3	4	1	2	3	4						
5	6	7	11	5	6	7		8	7	3	3	8.83	10
8	13		9	8	9	10	11						
12	14	15	10	12	13	14	15						
Problem 5													
4	2	11	15	4	2	15	13						
12	3	13	7	12	3	11	7	8	6	3	3	7.41	8
1		9	14	1	9	14							
8	6	10	5	8	6	10	5						
Problem 6													
4	1	10	15	4	1	10	15						
	7	3	12	7	3	14	12	10	9	3	4	12.24	14
9	2	14	11	9	13	5	2						
6	13	8	5	6	8		11						

(Continued on next page)

Appendix C (Continued):
Start and Goal Configurations Constituting the Problem-Solving Test, and Values of Various Physical Problem Characteristics

Problem 7													
6	2	10	14	6	2	10	14						
7	15	3	4	7	15	11		10	6	3	2	9.66	12
1	13	5	11	1	13	4	9						
12	8	9	11	12	8	3	5						
Problem 8													
14	10	13	12	15	14	13	12						
15	9	8	5	11	10	9	8	12	11	4	4	12.00	12
11		7	1	7		6	5						
4	3	6	2	4	3	2	1						
Problem 9													
1	2	3	4	1	2	3	4						
5	6	7	8	12	6		8	12	7	3	3	10.24	12
9		10	11	10	5	7	11						
12	13	14	15	13	9	14	15						
Problem 10													
1	2	3	4	1	2	4	10						
5		6	7	5	6	3	7	14	8	4	3	10.06	12
8	9	10	11	8	13		11						
12	13	14	15	12	14	9	15						
Problem 11													
5	9	6	11	12	5	9	6						
12	15	2	1	4	15	2	11	14	13	4	4	14.83	16
4	13	7	14	8	13		7						
8	3	10		3	10	14	1						
Problem 12													
1	2	3	4	9	5	3	4						
5	6	7	8		6	7	8	16	7	4	2	12.06	14
9		10	11	12	1	10	11						
12	13	14	15	13	2	14	15						
Problem 13													
10	9	3	7	10	2	9	7						
4	8		6	12	8		6	16	6	3	3	8.65	10
12	5	2	14	5	4	3	14						
1	11	15	13	1	11	15	13						

Appendix D:

Instruction and Recording Booklet for Perceived Difficulty Rating Study

Directions

Thank you for your participation. In this study, you will be asked to sort certain puzzles into piles based on how difficult they appear to you. Although you will not actually solve the puzzles yourself, you will need to know how they would be solved so that you can estimate how difficult they would be. All the puzzles will be of the type pictured here.

Make your moves in this pattern

1	2	3	4
5		6	7
8	9	10	11
12	13	14	15

Try to match this pattern

1	2	3	4
5	9	6	7
8	13	10	11
	12	14	15

Figure 1

The way to solve these puzzles is to "move" the numbers in the left pattern so that the left pattern will match the pattern on the right. A number may only be moved into the blank square in the left pattern. For example, to solve this particular puzzle (Fig. 1) one must make 3 "moves" as follows:

Move 1

First, by moving the "9" up one square in the left pattern, we obtain the following new pattern:

1	2	3	4
5	9	6	7
8		10	11
12	13	14	15

Figure 2

Move 2

By moving the "13" up one square in this new pattern (Fig. 2) we obtain the following pattern:

1	2	3	4
5	9	6	7
8	13	10	11
12		14	15

Figure 3

Move 3

Finally, by moving the "12" right one square, we obtain the following pattern which solves the puzzle since it matches the original right-hand pattern in Fig. 1.

1	2	3	4
5	9	6	7
8	13	10	11
	12	14	15

Figure 4

If at this point you do not understand how these puzzles are solved, please contact the proctor before reading on.

You will be presented with a number of these puzzles of varying difficulty. Your task is to study each puzzle and, keeping in mind how such puzzles are solved, estimate how difficult each puzzle would be. You should do this using the following steps. You should complete each step before going on to the next step. If you have any questions don't hesitate to contact the proctor.

Step 1 Sort of Puzzles.

First, study each puzzle and place it in one of the six piles provided by the proctor labelled:

Very Difficult, Difficult, Somewhat Difficult, Somewhat Easy, Easy, Very Easy.

There is no requirement that each pile contain a certain number of puzzles. You may feel, for example, that none of the puzzles fits the description "somewhat easy". Just place each puzzle in the pile that you feel provides the best description of how difficult it would be to solve the puzzle. You should try to make your initial placement as accurate as possible but you are free to change the location of any puzzle you wish if you change your mind about its difficulty. Remember that you do not have to actually solve the puzzles. Just study each puzzle long enough to feel reasonably confident about which pile to place it into.

A few of the puzzles contain a puzzle number and the message "Provide your reason(s)" on the top. For these puzzles, you should write down the puzzle number shown, the pile in which you placed it, and the reason(s) for why you are sorting the puzzle into that pile. Use the space provided just below for this purpose.

For example, if you feel the puzzle would be "very easy" to solve then place the card in the "very easy" pile and explain why you think it would be "very easy" to solve next to the puzzle number on the Data Sheet. Do not just write a reason like "Because it is solved very easily or very easily or very quickly." Explain how you decided to would be very easy, that is, on what basis did you decide to sort it into the "very easy" pile.

Provide your reason(s)

Puzzle Number	Assigned Pile	Reason(s) for sorting into the Pile you Did

Step 2 - Record sorting results

Each puzzle card has a number on the back. When you have finished sorting the puzzles into the 6 piles list these numbers under the appropriate label below. There is no required number of puzzles for any category.

Very Difficult	Difficult	Somewhat Difficult	Somewhat Easy	Easy	Very Easy

Step 3 - Subdividing the 6 piles

Examine the puzzles you have sorted into each pile in Step 2. You may feel that not all puzzles in a given pile seem equally difficult to you even though they can all be described as "very difficult", or "somewhat easy" for example. If you feel this is the case, subdivide the puzzles within each of the original piles into as many smaller sub-piles representing different degrees of difficulty as you can. Only create more subpiles if you feel you can distinguish differences in difficulty between the puzzles in a given pile. If you cannot differentiate the difficulty of the puzzles within a given pile then do not subdivide the pile any further. Continue subdividing the piles until you can no longer differentiate the difficulty of the puzzles in each pile. During this step you should only compare and subdivide puzzles within each of the original six piles separately. Do not switch puzzles from one of the original 6 piles to another one, for example, from "Easy" to "Very Easy".

If, when you have completed this step you have been able to subdivide any of the original 6 categories, list the card numbers in each pile in the space provided below. When you list the subpiles always put the hardest puzzles within a category in subpile 1, the second hardest puzzles in subpile 2, and so on.

Very Difficult	Difficult	Somewhat Difficult	Somewhat Easy	Easy	Very Easy
subpiles	subpiles	subpiles	subpiles	sub.	subpiles
1 2 ...	1 2 ...	1 2 ...	1 2 ...	1 ...	1 2 ...

Step 4

Please answer the following questions as completely as possible.

1. Your name ____

2. Your student identification number _ ____________________

3. Before today, how often had you tried to solve the kind of puzzle you were asked to estimate the difficulty of in this study?

    a. never
    b. a few times
    c. many times

4. How much difficulty did you have understanding what you were supposed to do in this study?

    a. no difficulty
    b. a little difficulty
    c. much difficulty

5. When you sorted the puzzles into the original 6 categories, did you use any "rules" or criteria for sorting something into "very difficult", "difficult", "somewhat difficult", "somewhat easy", "easy", and "very easy"?

    YES NO

    If so, what were they?

    Very difficult -

    Difficult -

    Somewhat difficult -

    Somewhat easy -

    Easy -

    Very easy -

6. If you were able to subdivide the original 6 piles into more piles in Step 3, on what basis did you do so? That is, how did you decide which puzzles <u>within</u> a pile were more difficult than others?

7. If you did not subdivide any of the original 6 piles, try to explain why you could not do so.

8. How often did you use each of the following considerations in deciding how difficult a puzzle would be:

a. The number of "moves" required to solve the puzzle	<u>All</u>	<u>Most</u>	<u>Some</u>	<u>None</u>	(of the puzzles)
b. The number of "numbers" which did not match in the two patterns	<u>All</u>	<u>Most</u>	<u>Some</u>	<u>None</u>	"
c. Whether in one of the patterns the numbers were in numeric order from 1 to 15	<u>All</u>	<u>Most</u>	<u>Some</u>	<u>None</u>	"
d. How far apart certain numbers were in the two puzzles	<u>All</u>	<u>Most</u>	<u>Some</u>	<u>None</u>	"
e. The number of rows in the two patterns that did not match	<u>All</u>	<u>Most</u>	<u>Some</u>	<u>None</u>	"
f. The location of the "empty space" in the left pattern	<u>All</u>	<u>Most</u>	<u>Some</u>	<u>None</u>	"
g. The number of columns in the two puzzles that did not match	<u>All</u>	<u>Most</u>	<u>Some</u>	<u>None</u>	"
h. Whether you could "see" the actual sequence of moves that would be needed to solve the problem	<u>All</u>	<u>Most</u>	<u>Some</u>	<u>None</u>	"
i. The amount of time it would take to solve the problem	<u>All</u>	<u>Most</u>	<u>Some</u>	<u>None</u>	"

9. Did the length of this study affect your ability to perform the tasks required?

   a. not at all
   b. somewhat
   c. quite a bit

10. How did you feel about working on this study?

    a. I disliked it a lot
    b. I disliked it somewhat
    c. I felt neutral about it
    d. I enjoyed it somewhat
    e. I enjoyed it a lot

11. Any further comments?

Appendix E:

Sample Subject Protocols of Each Reported Dimension of Perceived Difficulty

a. The number of moves required to solve the puzzle or an explication of the actual moves needed:
   "It only took a few moves"
   "The '12' and '13' will go around corner into place, others look like they will move easily"

b. Whether subject could "see" the actual sequence of moves that would be needed to solve the problem (no number of explication of the moves provided):
   "I can work this out just at a glance--its obvious"
   "I see logical moves"

c. The number of squares ("numbers") which did not match in the two patterns:
   "All numbers--same location, except for '3' in bottom right hand corner"
   "I only had to deal with 5/16 of the digits"

d. The amount of time it would take to solve the problem:
   "Took 10 seconds to solve"
   "Took a while to see the pattern"

e. The type of moves required to solve the puzzle:
   "Some complicated moves must be made"
   "Tricky or misleading moves"
   "Needed a combination of movements of sets of numbers including moving number that was in correct spot to allow for other movements, then replacing at end"

f. How far apart certain numbers were in the two puzzles:
   "Don't move numbers very far"
   "Numbers in some cases move a great distance"

g. How much thought was required to solve the problem:
   "Required lots of thought"
   "I had trouble keeping all the moves in my head"

h. The number of columns not matching in the two patterns:
   "Because you only have to deal with two of the four columns"

i. The number of rows not matching in the two patterns:
   "Two rows match already"

j. The location of the 'empty space' in the left pattern:
   "Will require using the right columns because it contains the open space"

k. Similarity to an already solved or rated puzzle:
   "This puzzle easier since it resembles one already solved"

l. Whether either the left or the right pattern was in numeric order from 1 to 15:
   there were no examples of this dimension in the voluntary protocols

Appendix F: Supplementary Tables

Table F-1
Intercorrelations of Number of Repeated Moves for 13 Problems

Problem		1	2	3	4	5	6	7	8	9	10	11	12
2	*N*	*33*											
	r	-.03											
3	*N*	*33*	*33*										
	r	-.03	1.00										
4	*N*	*54*	*33*	*33*									
	r	-.05	-.07	-.07									
5	*N*	*54*	*32*	*32*	*53*								
	r	-.04	-.04	-.04	.23								
6	*N*	*54*	*32*	*32*	*53*	*54*							
	r	-.04	-.06	-.06	-.09	.00							
7	*N*	*54*	*32*	*32*	*53*	*53*	*53*						
	r	-.03	a	a	-.05	-.05	-.05						
8	*N*	*50*	*30*	*30*	*49*	*49*	*49*	*50*					
	r	-.04	-.05	-.05	.09	-.05	-.08	-.06					
9	*N*	*46*	*28*	*28*	*45*	*45*	*45*	*46*	*46*				
	r	-.07	-.08	-.08	.13	.03	.04	-.10	.52				
10	*N*	*44*	*28*	*28*	*43*	*43*	*43*	*44*	*44*	*44*			
	r	.08	.12	.12	.27	.40	.09	.05	-.18	-.14			
11	*N*	*49*	*28*	*28*	*48*	*48*	*48*	*49*	*48*	*45*	*43*		
	r	.18	-.08	-.08	.21	-.07	-.09	-.08	-.00	-.04	-.08		
12	*N*	*48*	*28*	*28*	*47*	*47*	*47*	*48*	*47*	*44*	*42*	*47*	
	r	-.11	-.12	-.12	.22	-.04	.45	.08	-.03	.03	.22	.09	
13	*N*	*49*	*29*	*29*	*49*	*49*	*49*	*49*	*48*	*45*	*44*	*48*	*47*
	r	-.03	.07	.07	.37	-.12	.11	.21	.34	.33	-.04	.00	.07

[a] Correlations not computed due to near zero standard deviations.

Table F-2
Cross-Correlations of Number of Legal Moves (Vertical) and
Number of Illegal Moves (Horizontal) for 13 Problems

Problem		1	2	3	4	5	6	7	8	9	10	11	12	13
1	*N*	*55*	*33*	*33*	*54*	*54*	*54*	*54*	*50*	*46*	*44*	*49*	*48*	*49*
	r	.07	-.14	-.08	.01	-.08	-.09	-.08	-.10	-.10	-.14	-.09	-.14	.01
2	*N*	*33*	*33*	*33*	*33*	*32*	*32*	*32*	*30*	*28*	*28*	*28*	*28*	*29*
	r	-.13	-.09	.17	.04	.01	.20	-.03	-.17	-.11	.54	.35	.01	-.07
3	*N*	*33*	*33*	*33*	*33*	*32*	*32*	*32*	*30*	*28*	*28*	*28*	*28*	*29*
	r	-.09	-.14	.29	.16	.11	-.10	-.09	-.12	.00	.15	.06	-.05	-.11
4	*N*	*54*	*33*	*33*	*54*	*53*	*53*	*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	-.16	.14	.10	.30	.50	.21	.31	.00	.01	.02	.13	.13	.25
5	*N*	*54*	*32*	*32*	*53*	*54*	*54*	*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	-.15	.30	-.08	.07	.20	.24	.26	.14	.15	.10	-.06	-.08	.15
6	*N*	*54*	*32*	*32*	*53*	*54*	*54*	*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	-.14	.16	-.03	-.06	-.08	.14	.17	.38	.30	.03	-.15	.02	.01
7	*N*	*54*	*32*	*32*	*53*	*53*	*53*	*54*	*50*	*46*	*44*	*49*	*48*	*49*
	r	-.10	.25	.17	.08	-.01	.23	.68	.13	-.11	.13	.08	.03	.06
8	*N*	*50*	*30*	*30*	*49*	*49*	*49*	*50*	*50*	*46*	*44*	*48*	*47*	*48*
	r	-.08	.39	.01	-.09	.21	-.12	-.03	.27	.10	-.03	-.06	.06	.15
9	*N*	*46*	*28*	*28*	*45*	*45*	*45*	*46*	*46*	*46*	*44*	*45*	*44*	*45*
	r	-.15	-.06	.00	-.29	-.08	.03	.08	-.04	.30	-.01	-.13	.12	-.21
10	*N*	*44*	*28*	*28*	*43*	*43*	*43*	*44*	*44*	*44*	*44*	*43*	*42*	*44*
	r	.17	.32	.01	.25	.08	.17	.49	-.09	-.06	.06	.03	-.07	-.01
11	*N*	*49*	*28*	*28*	*48*	*48*	*48*	*49*	*48*	*45*	*43*	*49*	*47*	*48*
	r	.07	.21	.20	-.11	-.01	.01	-.04	.05	-.06	.32	.33	-.13	-.03
12	*N*	*48*	*28*	*28*	*47*	*47*	*47*	*48*	*47*	*44*	*42*	*47*	*48*	*47*
	r	-.09	-.08	.32	-.28	.08	.16	-.01	-.13	.06	-.14	-.15	.11	-.33
13	*N*	*49*	*29*	*29*	*49*	*49*	*49*	*49*	*48*	*45*	*44*	*48*	*47*	*49*
	r	.17	.10	.26	-.14	.11	-.05	.04	-.08	.01	.02	-.04	-.01	.20

Table F-3
Cross-Correlations of Number of Legal Moves (Horizontal) and Number of Repeated Moves (Vertical) for 13 Problems

Problem		Problem 1	2	3	4	5	6	7	8	9	10	11	12	13
1	*N*	55	33	33	54	54	54	54	50	46	44	49	48	49
	r	1.00	-.04	-.03	-.05	-.06	-.04	.21	-.07	-.13	.22	.02	.13	.11
2	*N*	33	33	33	33	32	32	32	30	28	28	28	28	29
	r	-.03	.74	1.00	.60	-.08	-.06	-.11	-.08	-.16	-.11	-.10	.14	.17
3	*N*	33	33	33	33	32	32	32	30	28	28	28	28	29
	r	-.03	.74	1.00	.60	-.08	-.06	-.11	-.08	-.16	-.11	-.10	.14	.17
4	*N*	54	33	33	54	53	53	53	49	45	43	48	47	49
	r	-.05	-.10	-.07	.57	.16	.06	.22	.26	-.07	.16	.29	-.21	-.07
5	*N*	54	32	32	53	54	54	53	49	45	43	48	47	49
	r	-.04	.16	-.04	.16	.32	-.03	.20	.28	-.12	.03	.07	.14	.04
6	*N*	54	32	32	53	54	54	53	49	45	43	48	47	49
	r	-.04	-.08	-.06	-.02	-.08	.79	.13	.00	.30	.20	-.09	.06	.16
7	*N*	54	33	32	53	53	53	54	50	46	44	49	48	49
	r	-.03	a	a	-.05	-.09	-.06	-.03	-.10	-.18	-.05	-.08	-.23	.16
8	*N*	50	30	30	49	49	49	50	50	46	44	48	47	48
	r	-.04	-.07	-.05	-.04	-.08	-.08	.12	.44	.28	-.17	-.00	.12	.10
9	*N*	46	28	28	45	45	45	46	46	46	44	45	44	45
	r	-.07	-.11	-.08	-.01	-.11	.03	.00	.10	.49	-.15	-.16	.20	.17
10	*N*	44	28	28	43	43	43	44	44	44	44	43	42	44
	r	.08	-.00	.12	.23	-.04	.02	-.12	-.06	-.23	.37	-.06	-.09	-.16
11	*N*	49	28	28	48	48	48	49	48	45	43	49	47	48
	r	.18	-.11	-.08	-.01	-.15	-.09	.07	-.01	-.16	.07	.46	.06	.04
12	*N*	48	28	28	47	47	47	48	47	44	42	47	48	47
	r	-.11	-.17	-.12	.08	-.21	.43	-.02	-.08	.33	.24	-.07	.13	.18
13	*N*	49	29	29	49	49	49	49	48	45	44	48	47	49
	r	-.03	-.03	.07	.26	-.22	.02	.14	-.01	.12	-.09	-.09	.03	.29

[a]Correlation not computed due to near zero standard deviations.

Table F-4
Cross-Correlations of Number of Illegal Moves (Vertical) and Number of Repeated Moves (Horizontal) for 13 Problems

Problem		Problem 1	2	3	4	5	6	7	8	9	10	11	12	13
1	*N*	*55*	*33*	*33*	*54*	*54*	*54*	*54*	*50*	*46*	*44*	*49*	*48*	*49*
	r	.07	-.09	-.09	-.13	-.13	-.12	-.09	.10	.03	.01	.05	-.22	.01
2	*N*	*33*	*33*	*33*	*33*	*32*	*32*	*32*	*30*	*28*	*28*	*28*	*28*	*29*
	r	-.14	-.14	-.14	.27	.20	-.11	a	.17	-.24	-.17	.10	-.16	.09
3	*N*	*33*	*33*	*33*	*33*	*32*	*32*	*32*	*30*	*28*	*28*	*28*	*28*	*29*
	r	-.08	.29	.29	-.12	.25	.06	a	-.11	.10	.07	.03	-.18	.23
4	*N*	*54*	*33*	*33*	*54*	*53*	*53*	*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	.01	.16	.16	.49	.10	-.11	.01	-.17	-.18	.38	-.07	.03	.21
5	*N*	*54*	*32*	*32*	*53*	*54*	*54*	*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	-.08	.11	.11	.45	.37	-.15	-.11	-.03	.01	.28	-.04	.06	.23
6	*N*	*54*	*32*	*32*	*53*	*54*	*54*	*53*	*49*	*45*	*43*	*48*	*47*	*49*
	r	-.09	-.10	-.10	.12	.21	.03	-.13	-.10	.08	.14	-.19	-.08	.16
7	*N*	*54*	*32*	*32*	*53*	*53*	*53*	*54*	*50*	*46*	*44*	*49*	*48*	*49*
	r	-.08	-.09	-.09	.41	.43	.03	-.02	-.11	-.13	.17	-.01	.25	.25
8	*N*	*50*	*30*	*30*	*49*	*49*	*49*	*50*	*50*	*46*	*44*	*48*	*47*	*48*
	r	-.10	-.12	-.12	.28	.20	.11	-.14	.03	-.10	.08	.02	.08	-.11
9	*N*	*46*	*28*	*28*	*45*	*45*	*45*	*46*	*46*	*46*	*44*	*45*	*44*	*45*
	r	-.10	.00	0	.18	-.13	.08	-.05	-.08	.18	-.21	.03	.07	.14
10	*N*	*44*	*28*	*28*	*43*	*43*	*43*	*44*	*44*	*44*	*44*	*43*	*42*	*44*
	r	-.14	.15	.15	.03	.28	-.06	-.01	-.18	-.07	.08	-.08	-.03	-.33
11	*N*	*49*	*28*	*28*	*48*	*48*	*48*	*49*	*48*	*45*	*43*	*49*	*47*	*48*
	r	-.09	.06	.06	.36	.02	-.12	-.13	-.05	.07	-.05	.19	-.09	.09
12	*N*	*48*	*28*	*28*	*47*	*47*	*47*	*48*	*47*	*44*	*42*	*47*	*48*	*47*
	r	-.14	-.05	-.05	.47	.14	.01	-.20	.06	.31	.09	-.09	.22	.43
13	*N*	*49*	*29*	*29*	*49*	*49*	*49*	*49*	*48*	*45*	*44*	*48*	*47*	*49*
	r	.01	-.11	-.11	.49	.34	-.12	.01	-.09	-.10	.22	-.04	.09	.26

[a] Correlations not computed due to near zero standard deviations.

Table F-5
Product-Moment Correlations Between Initial Move Latency for Each Problem and Performance Measures, by Problem

Performance Measure	Problem 1	2	3	4	5	6	7	8	9	10	11	12	13
Individual Problem Score	.03	.31	-.08	-.01	-.01	-.11	-.02	-.16	-.23	-.19	-.07	-.03	-.26
Total 2	-.17	-.01	-.22	-.15	-.10	-.23	-.03	-.14	-.07	-.18	-.09	-.15	-.16
MRATE	.16	-.04	.21	.09	-.03	.15	.01	.01	.06	.13	.07	.07	.18

Table F-6
Product-Moment Correlations Between Average Move Latency for Each Problem and Performance Measures, by Problem

Performance Measure	Problem 1	2	3	4	5	6	7	8	9	10	11	12	13
Individual Problem Score	-.06	.00	-.03	.07	.07	.00	-.01	-.07	-.16	.00	.10	.00	.07
Total 2	-.11	-.14	-.14	-.05	-.04	-.15	-.10	-.01	-.20	-.15	.10	-.22	-.29
MRATE	.09	.04	.08	.02	-.09	.07	-.01	-.08	.06	.04	-.12	.06	.13

DISTRIBUTION LIST

Navy

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 COMNAVMILPERSCOM (N-6C)
Dept. of Navy
Washington, DC 20370

1 Dr. Richard Elster
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Patrick R. Harrison
Psychology Course Director
LEADERSHIP & LAW DEPT. (7b)
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 Dr. Kneale Marshall
Scientific Advisor to DCNO(MPT)
OP01T
Washington DC 20370

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock Co
Newport News, VA 23607

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. George Moeller
Head, Human Factors Dept.
Naval Submarine Medical Research Lab
Groton, CN 06340

1 Dr William Montague
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. William Moonan
Code 303
Navy Personnel R&D Center
San Diego, CA 92152

1 LCDR W. Moroney
Code 55MP
Naval Postgraduate School
Monterey, CA 93940

1 Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Library
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92138

1 Ted M. I. Yellen
Technical Information Office, Code 201
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

5 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

1 Director, Navy Personnel R&D Center
Washington Liason Office
Building 200, 2N
Washington Navy Yard, DC 20374

5 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Special Asst. for Education and
Training (OP-01E)
Rm. 2705 Arlington Annex
Washington, DC 20370

1 Office of the Chief of Naval Operations
Research, Development, and Studies Branc
(OP-102)
Washington, DC 20350

1 Long-Range Manpower, Personnel, and
Training Planning Branch (OP-110)
Room G828
Arlington Annex
Washington, DC 20350

1 Head, Manpower Training and Reserves
Section (Op-964D)
Room 4A478, The Pentagon
Washington, DC 20350

1 Captain Donald F. Parker, USN
Commanding Officer
Navy Personnel R&D Center
San Diego, CA 92152

1 LT Frank C. Petho, MSC, USN (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Dr. Gary Poock
Operations Research Department
Code 55PK
Naval Postgraduate School
Monterey, CA 93940

1 The Principal Deputy Assistant
Secretary of the Navy (MRA&L)
4E780, The Pentagon
Washington, DC 20350

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

1 DR. H.M. WEST III (OP-01)
Head, Program Development Branch(OP-120)
ARLINGTON ANNEX
WASHINGTON, DC 20350

1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Director, Research & Analysis Division
Plans and Policy Department
Navy Recruiting Command
4015 Wilson Boulevard
Arlington, VA 22203

1 Dr. Bernard Rimland (03B)
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group (TAEG)
Dept. of the Navy
Orlando, FL 32813

Army

1 Technical Director
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. J. Barber
HQS, Department of the Army
DAPE-ZBR
Washington, DC 20310

1 Col Gary W. Bloedorn
US Army TRADOC Systems Analysis Activity
Attn: ATAA-TH
WSMR, NM 88002

1 DR. RALPH DUSEK
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Ed Johnson
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 LTC Michael Plummer
Chief, Leadership & Organizational Effectiveness Division
Office of the Deputy Chief of Staff for Personnel
Dept. of the Army
Pentagon, Washington DC 20301

1 DR. JAMES L. RANEY
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Commandant
US Army Institute of Administration
Attn: Dr. Sherrill
FT Benjamin Harrison, IN 46256

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

1 U.S. Air Force Office of Scientific Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Research and Measurment Division
Research Branch, AFMPC/MPCYPR
Randolph AFB, TX 78148

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Dr. Frank Schufletowski
U.S. Air Force
ATC/XPTD
Randolph AFB, TX 78148

1 Jack A. Thorpe, Maj., USAF
Naval War College
Providence, RI 02846

1 Brian K. Waters, Lt Col, USAF
Air War College (EDV)
Maxwell AFB, AL 36112

Marines

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Director, Office of Manpower Utilization
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 Headquarters, U. S. Marine Corps
Code MPI-20
Washington, DC 20380

1 Special Assistant for Marine Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 Major Michael L. Patrow, USMC
Headquarters, Marine Corps
(Code MPI-20)
Washington, DC 20380

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Other DoD

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

1 Dr. William Graham
Testing Directorate
MEPCOM/MEPCT-P
Ft. Sheridan, IL 60037

1 Director, Research and Data
OASD(MRA&L)
3B919, The Pentagon
Washington, DC 20301

1 Military Assistant for Training and Personnel Technology
Office of the Under Secretary of Defense for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Lorraine D. Eyde
Personnel R&D Center
Office of Personnel Management of USA
1900 EStreet NW
Washington, D.C. 20415

1 Jerry Lehnus
REGIONAL PSYCHOLOGIST
U.S. Office of Personnel Management
230 S. DEARBORN STREET
CHICAGO, IL 60604

1 William J. McLaurin
Rm. 301, Internal Revenue Service
2221 Jefferson Davis Highway
Arlington, VA 22202

1 Dr. Andrew R. Molnar
Science Education Dev. and Research
National Science Foundation
Washington, DC 20550

1 Personnel R&D Center
Office of Personnel Managment
1900 E Street NW
Washington, DC 20415

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Ms. Carole A. Bagley
Minnesota Educational Computing Consortium
2354 Hidden Valley Lane
Stillwater, MN 55082

1 Dr. Jackson Beatty
Department of Psychology
University of California
Los Angeles, CA 90024

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 DezWPs im Streitkraefteamt
Postfach 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. Lyle Bourne
Department of Psychology
University of Colorado
Boulder, CO 30309

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52240

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. William E. Coffman
Director, Iowa Testing Programs
334 Lindquist Center
University of Iowa
Iowa City, IA 52242

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N.W.
Washington, DC 20036

1 Dr. Kenneth B. Cross
Anacapa Sciences, Inc.
P.O. Drawer Q
Santa Barbara, CA 93102

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Professor Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ross Green
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Frederick Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Dr. Douglas H. Jones
Rm T-255
Educational Testing Service
Princeton, NJ 08450

3 Journal Supplement Abstract Service
American Psychological Association
1200 17th Street N.W.
Washington, DC 20036

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
Honeywell Plaza MN12-3166
Minneapolis, MN 55408

1 Professor John A. Keats
University of Newcastle
AUSTRALIA 2308

1 Dr. Mazie Knerr
Litton-Mellonics
Box 1286
Springfield, VA 22151

1 Dr. Stephen Kosslyn
Harvard University
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Michael Levine
210 Education Building
University of Illinois
Champaign, IL 61820

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat
Groningen
NETHERLANDS

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77004

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Dr. Mark Miller
Computer Science Laboratory
Texas Instruments, Inc.
Mail Station 371, P.O. Box 225936
Dallas, TX 75265

1 Professor Jason Millman
Department of Education
Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army Navy Drive
Arlington, VA 22202

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MINRAT M. L. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 1328
D-53 BONN 1, GERMANY

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
4 Hill Hall
Columbia, MO 65211

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Dr. Kazuo Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 980
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Kathryn T. Spoehr
Department of Psychology
Brown University
Providence, RI 02912

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Dr. David Stone
ED 236
SUNY, Albany
Albany, NY 12222

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Brad Sympson
Psychometric Research Group
Educational Testing Service
Princeton, NJ 08541

1 Dr. Kikumi Tatsuoka
Computer Based Education Research Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 Dr. John Wannous
Department of Management
Michigan University
East Lansing, MI 48824

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2
WEST GERMANY

PREVIOUS PUBLICATIONS

Proceedings of the 1977 Computerized Adaptive Testing Conference. Jul

Research Reports

Final Report: Computerized Adaptive Performance Evaluation. February
80-1. Effects of Immediate Knowledge of Results on Achievement Test Performa and Test Dimensionality. January 1980.
79-7. The Person Response Curve: Fit of Individuals to Item Characteristic Models. December 1979.
79-6. Efficiency of an Adaptive Inter-Subtest Branching Strategy in the Meas of Classroom Achievement. November 1979.
79-5. An Adaptive Testing Strategy for Mastery Decisions. September 1979.
79-4. Effect of Point-in-Time in Instruction on the Measurement of Achieveme August 1979.
79-3. Relationships among Achievement Level Estimates from Three Item Charac Curve Scoring Methods. April 1979.
Final Report: Bias-Free Computerized Testing. March 1979.
79-2. Effects of Computerized Adaptive Testing on Black and White Students. 1979.
79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve February 1979.
78-5. An Item Bias Investigation of a Standardized Aptitude Test. December
78-4. A Construct Validation of Adaptive Achievement Testing. November 1978
78-3. A Comparison of Levels and Dimensions of Performance in Black and Whit on Tests of Vocabulary, Mathematics, and Spatial Ability. October 197
78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Tes Performance and Psychological Reactions to Testing. September 1978.
78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Stra August 1978.
77-7. An Information Comparison of Conventional and Adaptive Tests in the Me ment of Classroom Achievement. October 1977.
77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October
77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievemen September 1977.
77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977
77-3. Accuracy of Perceived Test-Item Difficulties. May 1977.
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Resp Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977.
Final Report: Computerized Ability Testing, 1972-1975. April 1976.
76-5. Effects of Item Characteristics on Test Fairness. December 1976.
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive A Testing. June 1976.
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Abil Test Performance. June 1976.
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976.
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. Marc
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975.
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. Nov 1975.
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlation March 1975.
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Te February 1975.
74-5. Strategies of Adaptive Ability Measurement. December 1974.
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974.
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability July 1974.
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 197
74-1. A Computer Software System for Adaptive Ability Measurement. January
73-3. The Stratified Adaptive Computerized Ability Test. September 1973.
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-1. Ability Measurement: Conventional or Adaptive? February 1973.

Copies of these reports are available, while supplies last, from:
Computerized Adaptive Testing Laboratory
Psychometric Methods Program, Department of Psychology
N660 Elliott Hall, University of Minnesota
75 East River Road, Minneapolis, Minnesota 55455

EN