DOCUMENT RESUME

ED 119 038 95 CG 008 429

AUTHOR Speedie, Stuart M.; And Others

TITLE Teaching Problem-Solving Skills: Development of an Instructional Model Based on Human Abilities Related to Efficient Problem Solving. Final Report.

SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.

BUREAU NO BR-2-E-051

PUB DATE Aug 73

GRANT OEG-5-72-0042(509)

NOTE 169p.; Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.

DESCRIPTORS *Ability Identification; Elementary Education; *Problem Solving; *Process Education; *Productive Thinking; Research Projects; *Skill Development; Student Ability; Teaching Procedures

ABSTRACT

This study investigates one way humans cope with change--problem-solving. It concentrates on the human abilities important to efficient problem-solving and the processes involved in problem-solving. The objectives of this study were: (1) to develop three group administration tasks that measure problem-solving processes; (2) to evaluate the reliability of these tasks; (3) to determine the predictability of performance on these tasks from a set of human ability measures specified by the Structure of Intellect model; and (4) to determine the underlying structure of the ability measures and problem-solving criteria. Seventeen tests of ability, along with four of the problem-solving criterion tasks, were administered to a sample of 490 fifth-grade students. Test-retest reliabilities were found to be quite low, but regression analysis revealed that convergent production of semantic transformations and two memory abilities were related to performance on simulated problems; memory and evaluation abilities were related to Verbal Maze performance; and logical reasoning and noticing details were important to concept identification. The lack of important results indicates the need for replication of the study with several important changes in design. (Author)

5P

FINAL REPORT
Project No. 2–E–051
Grant No. OEG–5–72–0042(509)

TEACHING PROBLEM-SOLVING SKILLS: DEVELOPMENT OF AN INSTRUCTIONAL MODEL BASED ON HUMAN ABILITIES RELATED TO EFFICIENT PROBLEM SOLVING

Stuart M. Speedie
Donald J. Treffinger
John F. Feldhusen

Purdue University
West Lafayette, Indiana 47907

August, 1973

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
Office of Education
Bureau of Research

Final Report

Project No. 2-E-051

Grant No. OEG-5-72-0042(509)

Teaching Problem-Solving Skills: Development of an Instructional Model Based on Human Abilities Related to Efficient Problem Solving

Stuart M. Speedie
Donald J. Treffinger
John F. Feldhusen

Purdue University

West Lafayette, Indiana 47907

August, 1973

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Speedie, Stuart Mitchell. Ph.D., Purdue University, August, 1973. A factor analytic study of specific cognitive abilities related to human problem solving. Major Professors: Donald J. Treffinger and John F. Feldhusen.

The problem of change is one of the most urgent and formidable challenges faced by all individuals in modern society. Yet we know little of the means by which individuals adjust to change. The purpose of this study was to investigate one way by which humans cope with change - problem solving. More specifically, this study concentrated on what human abilities are important to efficient problem solving. First, however, since present problem-solving tasks are insufficient samples of problem-solving behavior, new tasks were developed which focused on the processes involved in problem solving. Thus the objectives of this study were as follows: (a) to develop three group administration tasks - simulated problem situations, Verbal Mazes, and concept identification tasks - that measured problem-solving processes; (b) to evaluate the reliability of these tasks; (c) to determine the predictability of performance on these problem-solving tasks from a set of human ability measures specified by

the Structure of Intellect model; and (d) to determine the underlying structure of the ability measures and problem-solving criteria.

Four forms of each problem-solving task were developed, consisting of single and multiple solutions, and an alternate form of each. Two forms of each task were evaluated for reliability by means of a two week test-retest situation. The tasks were also evaluated for alternate forms reliability by simultaneous administration of the alternate forms of each task.

The representative sample of human ability measures consisted of 17 tests from the Structure of Intellect model measuring each of the processes operating on semantic content to produce several different types of products. These 17 tests, along with four of the problem-solving criterion tasks, were administered to a sample of 490 fifth grade students over a period of five hours. The resulting data was subjected to regression analysis, conventional factor analysis with Promax oblique rotation, canonical correlation analysis, and extension loadings analysis.

Test-retest reliabilities were found to be quite low, ranging from .30 to .74. Alternate forms reliabilities were essentially zero except for the Verbal Maze problems. Regression analysis revealed that convergent production of semantic transformations and two memory abilities were related to performance on the simulated problems; memory and evaluation abilities were related to Verbal Maze performance; and logical reasoning and noticing details were important to concept identification. The factor analysis revealed five factors: (a) a general test performance factor, (b) a simulated problem factor, (c) a concept identification factor, (d) a divergent thinking measure, and one insignificant crossover factor. Canonical correlation analysis and extension loading analysis added nothing more.

It was concluded that, although a number of actors may have affected the results, the most probable cause of the lack of important results was the low reliabilities of the problem-solving criteria. The regression analyses yielded some information, but hardly of great significance, since multiple Rs were generally low. Thus the study requires replication with several important changes in order to reveal significant insights into the problem-solving process. Also it was concluded that no empirically based model of instruction could be based on the results of this study.

INTRODUCTION

The problem of change is one of the most urgent and formidable challenges faced by all individuals in modern society. Yet we know little of the means by which individuals adjust to change, especially rapid change. In a world where technology has made change a constant fact of life, it is necessary to find ways to educate individuals in coping effectively, or face the alternative of being drowned by the rising tides of our own inventiveness and technology. In his book Future Shock, Toffler (1970) contends that we shall need individuals

> ... who can make critical judgements, who can weave their way through novel environments, who are quick to spot new relationships in the rapidly changing reality. It (modern society) requires men who in C. P. Snow's compelling term, have the future in their bones. (p. 403).

Since solving the problems generated by change appears to be of prime importance for surviving in this presentday world, a central role of education should be to assist each individual in developing the abilities and skills necessary for coping with this change. Thus, schools should have the development of problem-solving skills in each person as one of their primary goals. Or as Toffler (1970) argues:

> The new education must teach the individual how to classify and reclassify information, how to evaluate its veracity, how to change categories when necessary, how to move from the concrete to the abstract and back, how to look at problems from a new direction--how to teach himself. (p. 414).

Such goals are often stated in the general philosophy and objectives of school systems but not generally implemented to the fullest possible extent. Students are often taught known solutions to familiar or age-old problems, and little time is spent in explicitly teaching strategies for encountering and dealing with new problems. (cf. Kozol, 1967; Holt, 1964).

It is not reasonable, however, to blame the schools entirely for this deficit since social and behavioral scientists know so little themselves about how people solve the type of problems one encounters in the fastpaced world of today. What seems to be called for is the development of innovative approaches to instruction aimed at training individuals to deal with such problems. Such approaches would necessarily concentrate on teaching complex cognitive skills and strategies necessary for confronting the problems of fast-paced change.

However, the effectiveness of many innovative instructional programs is open to question. This has been documented by Charles Silberman in Crisis in the Classroom (1970):

> ... the 1950's and '60s saw one of the largest and most sustained educational reform movements in American history, an effort that many observers, this writer included, thought would transform the schools. Nothing of the sort has happened; the reform movement has produced innumerable changes, and yet the schools themselves are largely unchanged. (p. 158).

Perhaps one of the reasons that few educational innovations have been effective is that these programs are seldom derived from empirically tested principles of instruction, nor based on proven ideas concerning their subject matter. Most are quite effective in training students to achieve knowledge level objectives but few have concentrated on the processes or strategies important to their particular concerns.

To summarize, our rapidly changing society requires persons who are able to solve a great variety of challenging problems. This implies that schools must emphasize courses of instruction which dwell on the skills and abilities necessary for problem solving. Yet, few such programs of instruction exist. Thus, a new and innovative approach would seem to be indicated although educational innovations have a history of failure. How, then, can we proceed to design innovative courses of instruction which will help our children to meet future challenges?

This project assumes that a number of goals must be accomplished in order to design such a program. First, a clear and precise definition of problem-solving must be developed. Second, the strategies which people solve the multitude of problems they face must be elucidated. Third, the intellectual skills and abilities necessary for solving problems must be established. Fourth, instructional strategies must be designed which are derived from the findings of the first three goals and the empirically established principles of learning and instruction.

Obviously, the accomplishmentof these four goals is a highly ambitious undertaking and one which is beyond the scope of this dissertation. Instead, two goals will occupy the primary focus. That is, which intellectual skills and abilities are important to efficient problem-solving and can an instructional program based on these abilities be designed?

REVIEW OF LITERATURE

The following sections will review the literature with respect to definitions of problem-solving. Several models of the problem-solving process will be reviewed. The operational definitions of problem-solving will be discussed and evaluated. From the components of the problem-solving models and the literature concerned with intellectual skills involved in problem-solving, a set of important skills will be hypothesized and a study proposed.

Definitions of problem-solving.

By its semantic nature, problem solving is an action performed upon an object. Thus in defining problem solving, one must first deal with the object, i.e., the nature of "problem" must be explained. Johnson (1972, p. 25) stated that a problem exists when an individual is motivated toward a goal and his first attempt to reach it is unrewarding. Newell and Simon (1972) contend that a person is confronted with a problem when he wants something and does not know immediately what specific actions he can perform to get it. It appears to be a generally accepted definition (cf. Duncker, 1945; Wertheimer, 1959; Duncan, 1959; Ausubel and Robinson, 1969; Bourne, Ekstrand and Dominowski, (1971) that a problem is a situation in which a person is motivated toward a goal from some present state, but does not know the specific means of achieving that goal, given the present state.

Problem solving is, then, the means by which an individual moves from his present state to some goal. However, what exactly constitutes this process is a matter for considerable debate. Davis (1966) claimed that problem solving is any semi-complex learning task which has not already been identified by another name. The more concrete definitions, however, have been expressed as identifiable subgroups of intellectual processes which are linked in some order. These links have been of two kinds: linear and feedback.

Earlier definitions of problem solving generally listed several intellectual processes linked in a linear manner. Dewey (1933) proposed that problem solving consisted of five phases:

> (1) suggestions in which the mind leaps forward to possible solution; (2) an intellectualization of the difficulty or perplexity that has been <u>felt</u> into a problem to be solved, a question for which the answer must be sought; (3) the use of one suggestion after another as a leading idea, or hypothesis, to initiate and guide observation and other operations in the collection of factual materials; (4) the mental elaboration of the idea or supposition ...; and (5) testing the hypothesis by overt or imaginative action. (p. 107).

Wallas (1926) suggested four phases of problem solving which were preparation, incubation, illumination, and verification. Later definitions of problem solving suggested interactions among the various phases. Merrifield, Guilford, Christensen and Frick (1962) proposed that the phases of preparation, analysis, production, verification, and reapplication interacted in the problem solving process. Johnson (1972) suggested three broad classes of processes: preparation, production, and judgement. Others, such as Gagné (1966) maintain that problem solving is the complex interaction of a number of subordinate learnings which load to the learning of a new rule.

Later approaches to defining problem solving have usually considered it to be the operation of an information processing system, a system of processing and feedback. A paradigm of such an information processing system is given in Figure 1.

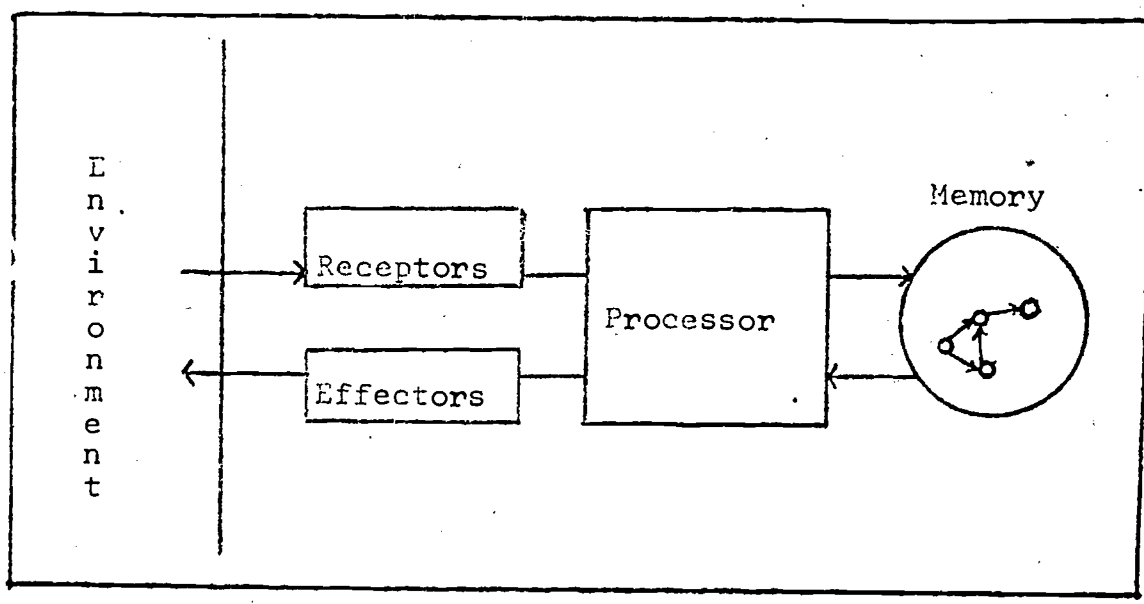


Figure 1. General structure of an information processing system (After Newell and Simon, 1972).

Newell and Simon (1972) make four propositions concerning problem solving in this information processing system:

1. A few, and only a few, gross characteristics of the human IPS are invariant over task and problem solver.
2. These characteristics are sufficient to determine that a task environment is represented (in the IPS) as a problem space, and that problem-solving takes place in a problem space.
3. The structure of the task environment determines the possible structures of the problem space.
4. The structure of the problem space determines the possible programs that can be used for problem-solving.

These four propositions mean, in more conventional terms, that:(a) there are only a few characteristics which are true of all problems and problem solvers; (b) but these few characteristics are enough to determine that the problem is perceived by the problem solver; (c) and these objective characteristics of the problem task determine how the problem solver views the problem; (d) and finally, how the problem solver perceives the problem determines how he will solve the problem. Newell and Simon have derived this model from a great deal of simulation work with three problem tasks: theorem proving, cryptarithmetic, and chess. For these tasks, possible problem spaces have been specified, and programs which appear to imitate human behavior have been developed. However, one must be cautioned that these problem spaces and programs (solution strategies) are highly task specific.

Guilford (1967) proposed a model of problem solving involving processes from the Structure of Intellect model. This model is represented in Figure 2. In this model the problem exists in terms of environmental and somatic input. This input is filtered by arousal and attentional mechanisms and the problem is either ignored (Exit I) or the problem is sensed and structured through cognition processes. These cognition processes may call on evaluation processes or new input. After passing through this stage, the problem may again be

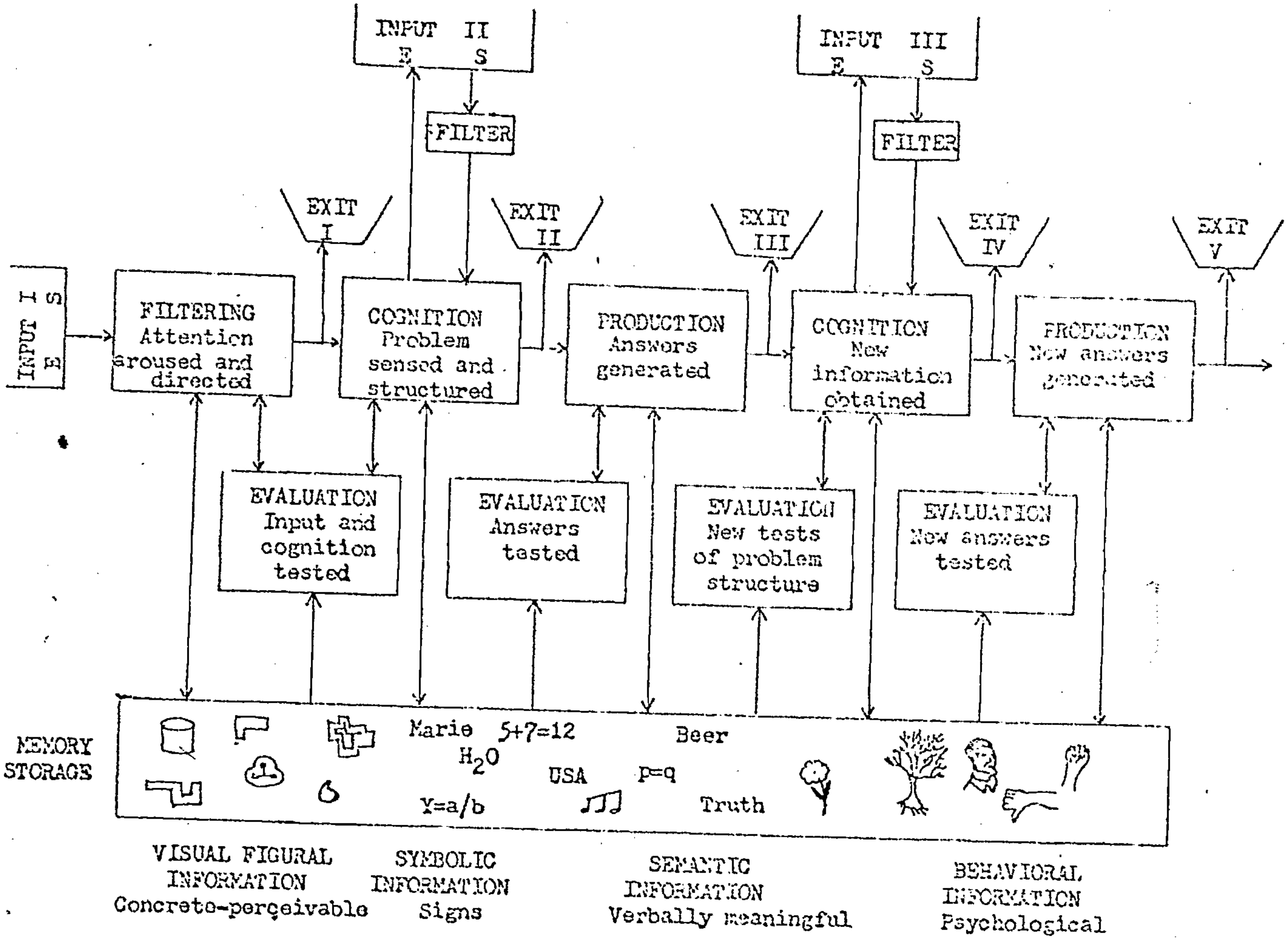


Figure 2. J. P. Guilford's model of problem solving.

exited (Exit II) or answers generated either by convergent or divergent production. These answers are evaluated and either the problem is exited (Exit III) or a new cycle of cognition, production, evaluation is started. Underlying all of these processes is memory accession, upon which the other processes are dependent. The Guilford model of problem solving is an information processing model in which a variety of intellectual processes act upon environmental and somatic input, with constant reference to memory, to generate problem solutions.

A third model of problem solving is suggested by Ausubel and Robinson (1969). The model, as depicted in Figure 3, postulates four levels in the problem-solving process: (1) problem setting, (2) definition of the problem, (3) gap filling, and (4) verification of the solution. Central to problem solving is the gap filling process. This process starts with the given problem information plus an individual's background. These are manipulated by rules of inference guided by a strategy in order to reduce the gap between the initial proposition and the final proposition or goal. This gap filling process is similar in many respects to the means-ends analysis of Newell, Shaw and Simon (1958) which they employed in their General Problem Solver.

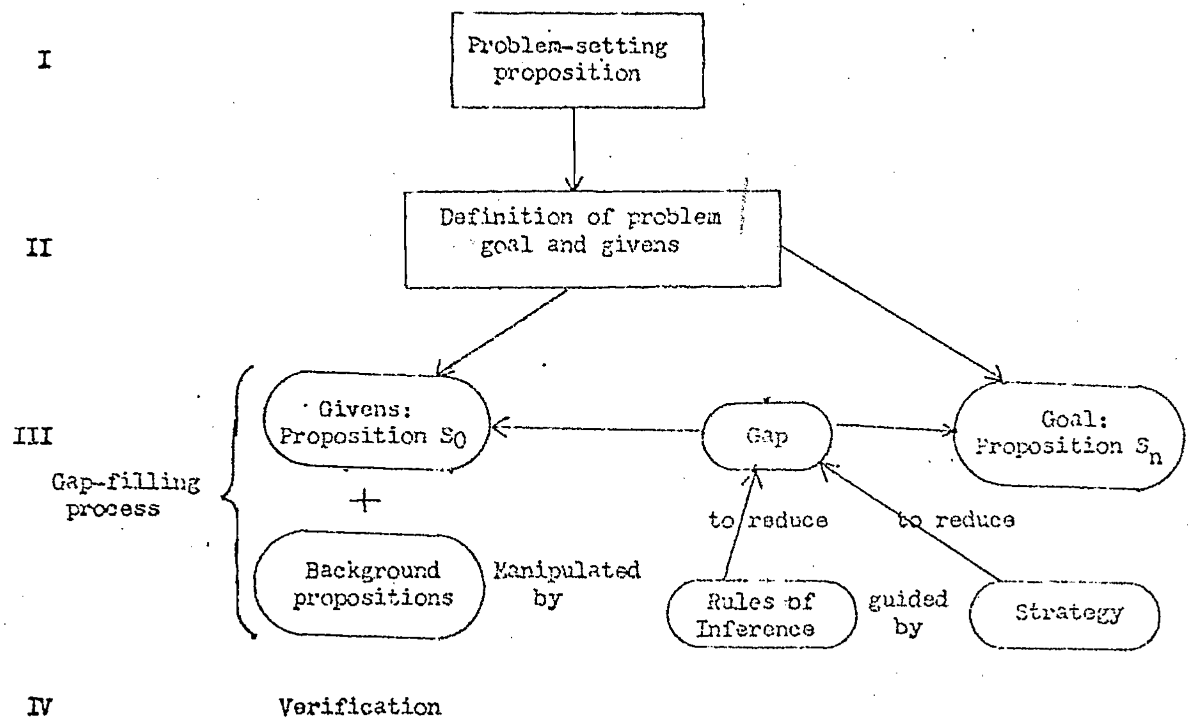


Figure 3. Problem solving paradigm - Ausubel and Robinson (1969, p. 505

A fourth model of problem solving was developed by Feldhusen, Houtz, and Ringenbach (1972) through a review and synthesis of the problem-solving literature. As a result of their review, they postulated 12 different types of problem-solving processes:

1. Sensing that a problem exists;
2. Defining the problem;
3. Clarifying the problem;
4. Asking questions;
5. Guessing causes;
6. Judging if more information is needed;
7. Noticing relevant details;
8. Using familiar objects in unfamiliar ways;
9. Seeing implications;
10. Solving multiple solution problems;
11. Solving single solution problems; and
12. Verifying solutions.

A fifth, perhaps more elemental model of problem solving was put forth by Miller, Galanter, and Pribram in their book Plans and the Structure of Behavior (1960). They postulated that all variety of problems confronting human beings are solved through the use of "plans". The basic element of the "plan" is the TOTE. The TOTE is a basic feedback system illustrated in Figure 4. The basic process is to test a given condition, exit if the condition is satisfied, otherwise perform some one of a small class of operators and test the condition again. It is evident that this model does not specify any particular processes involved in problem

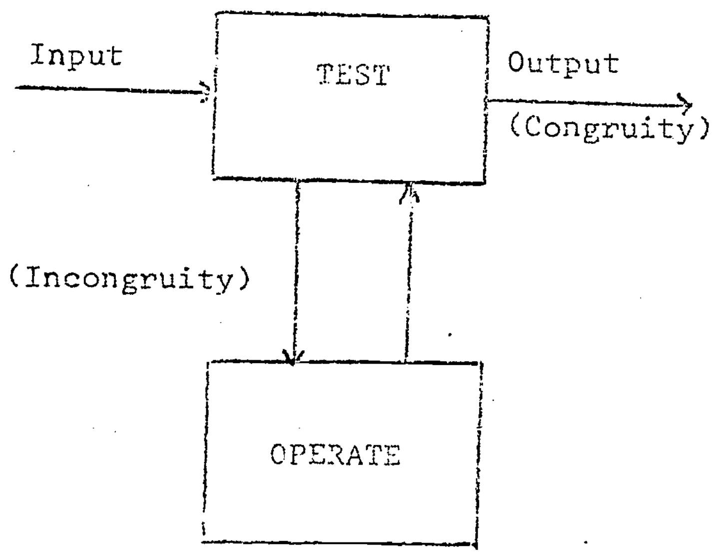


Figure 4. The basic TOTE feedback system.

solving, but it does specify the form that any such process will take; namely a TOTE hierarchy as given in Figure 5.

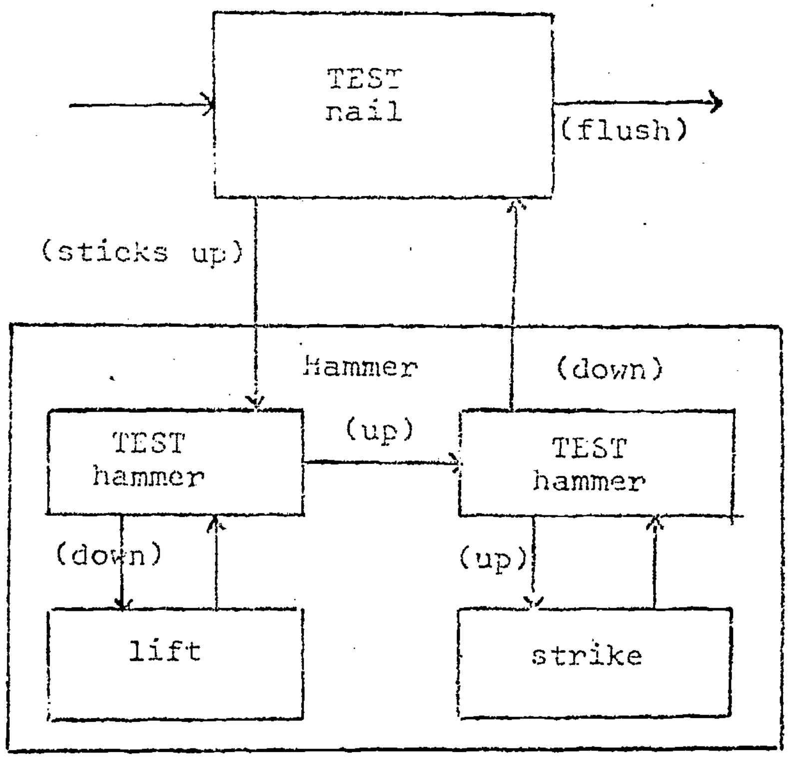


Figure 5. A TOTE hierarchy for hammering a nail.

To summarize, there are a number of definitions of problem solving, embodied in several different models. These definitions all specify that problem-solving is the process of reaching a goal state from an initial state, where the means is not at first known. However, the intellectual skills and abilities that choose, initiate, and execute these processes is a matter of debate. It remains a matter for empirical study to identify these intellectual skills and abilities.

Empirical studies of abilities important to problem solving.

Basic empirical work in this area was done by Merrifield et al. (1962) in their factor analytic study of problem solving. Further analysis of the data collected were presented in Guilford and Hoepfner (1971). In this study 37 measures representing 16 different Structure of Intellect (SI) factors were administered to a group of 232 Naval personnel. The 16 SI factors represented are listed below:

1. CMU cognition of semantic units (verbal comprehension)
2. CMC cognition of semantic classes (conceptual classification)
3. CMR cognition of semantic relations (education of conceptual relations)
4. CMS cognition of semantic systems (general reasoning)
5. CMT cognition of semantic transformations (penetration)
6. CMI cognition of semantic implications (conceptual foresight)
7. DMU divergent production of semantic units (ideational fluency)
8. DMC divergent production of semantic classes (spontaneous flexibility)

9. DMR divergent production of semantic relations (associational fluency)
10. DMT divergent production of semantic transformations (originality)
11. NMC convergent production of semantic classes
12. NMR convergent production of semantic relations (education of conceptual correlates)
13. NMT convergent production of semantic transformations (conceptual redefinition)
14. NMI convergent production of semantic implications (deduction)
15. EMR evaluation of semantic relations (logical evaluation)
16. EMI evaluation of semantic implications (sensitivity to problems)

central reason for choosing these particular 16 the authors reasoned hypothesis that these lities were central to the production phase of problem ving. They had chosen to concentrate on this phase to the nature of their problem-solving tasks.

The problem-solving criteria of the study were sing Links, in which the task was to produce three ds to complete a chain of associations between an initial and final word. The second criteria was Predicaments in which the task was to produce two ways in which given objects could be used to solve a specific problem. A third criteria was Transitions in which the S was required to write a coherent account logically connecting the given initial and final situations of a short story; this test yielded measures of coherence and logical connection. A summary of their revised results is given in Table 1.

The most recent analysis (Guilford and Hoepfner, 1971) indicated that there was no discrete problem-solving factor. Rather, the problem-solving tests loaded on a number of SI factors. Transitions was associated with an ability they labelled verbal comprehension (or in SI terms, cognition of semantic units). Performance on the Predicaments test was related to performance on a factor they labelled conceptual foresight or cognition of semantic implications. The ability labelled sensitivity to problems or cognition of semantic implications was related to performance on both the Predicaments and Transitions tests. Missing Links was associated with convergent production of semantic transformations and finally, originality or divergent production of semantic transformations and performance on the Transitions test were shown to be associated.

However, it should not be concluded from the report of Merrifield et. al. (1962) that these are the only abilities related to human problem solving. One serious limitation of the study involves the representativeness of the criterion tests with respect to problem-solving behavior. Three of the four measures were based on the determination of verbal associations, utilizing simple word to word associations in the Missing Links test or the more complex "implied meaning" in the Transitions test. In the Predicaments test, the dominant ability would appear to be the recall and evaluation of uses for the objects specified. In none of these tests was there a provision for measuring the strategies of problem solution, or provision for multiple solutions, asking relevant questions about the problem, clarifying the goal, or defining the problem. Thus, the tests employed by Merrifield et. al. (1962) did not include measures for many possible facets of efficient problem solving.

Werdelin (1966) reported two factor analyses in which the criteria were problem-solving tests in mathematics. In one study 36 human ability tests were administered to a group of boys, and in the second study 29 tests were given to similar samples. There were 18 tests in common in the two studies. Factor analysis resulted in 5 common factors for the human abilities.

They were general reasoning, deductive reasoning, verbal comprehension, space and numerical factors. The problem-solving criteria loaded most heavily on the general reasoning factor. Again, however, the generalizability of the factors was limited by the specific nature of the criterion tests.

A third factor analytic study, carried out by Bunderson (1967), used problem-solving criteria of concept learning tasks. The primary purpose of the study was to determine the relative importance of a number of human abilities at several stages of concept learning. A total of 30 tests representing 10 factors and 26 concept problems were administered to a sample of 145 Princeton undergraduates. For complex positive and negative problems, the concept problems were divided into 6 sequential blocks of 3 problems each and three measures were derived. The remaining eight problems (double negative) were divided into 4 sequential blocks of 2 problems each. Extension loadings of the concept measures on the ten factors were calculated using a differential performance function, for each of the blocks.

Bunderson reported that factors of Chunking Memory, Perceptual Speed, and Spatial Scanning were important in the initial attempts at solving the problems. The factors of Verbal and General Reasoning, Induction, and Flexibility came to play a more important function as Ss improved in solving concept problems. And finally the factors of Memory Span, Associative Memory, and Spatial Scanning became most important in the latter stages of the concept problems, presumably when strategies for solution had been learned, recall became the principle means of solution. Again, however, generalizability of the results to other problem-solving tasks is limited by the specificity of the criteria.

These three studies comprise the significant literature in the factor analytic studies of problem solving. However, two other studies have used a correlational approach to the investigation of what human abilities are important to problem solving. Harootunian and Tate (1960) report a study in which they attempt to relate a composite problem-solving criterion to a number of human ability factors through regression analysis. The factors included Problem Recognition, Word Fluency, Ideational Fluency, Closure, and Judgement. These were represented by a total of 14 tests. The problem-solving criterion consisted of a linear composite of verbal and abstract reasoning scores from the Differential Aptitude Test, the Davis-Eells Games, and a specially constructed test entitled Thought Problems. The battery of tests was administered to a group 632 seventh and eighth grade students. Factor scores were additive composites of their representative tests. Correlations of the factors with the problem solving criterion were respectively .624, .417, .290, .396, and .707. In the light of the nature of the criterion, these correlations are not surprising. There is a high degree of similarity between the tests representing Problem Recognition and Judgement and the three tests making up the criterion. Thus, this study does not provide much information concerning the important factors in problem solving, though it does make an attempt to use more a comprehensive measure of problem solving.

The second correlational study was by Stevenson, Hale, Klein, and Miller (1968). In this study the authors investigated the relationships between a number of learning measures and several problem-solving tasks. The learning measures consisted of 12 tasks, two paired associate tasks, three discrimination tasks, a probability learning task, an incidental learning task, and a verbal memory task. The problem-solving criteria were two concept of probability learning tasks, a conservation of volume task, and an anagram task. These tasks were administered to a group of bright, average, and dull seventh graders.

It was evident from their data that there were a number of sex differences. For both sexes, however, the correlations show that associative memory as reflected in the paired associate tasks is important to performance on the problem-solving tasks. This conclusion is reinforced by the importance of verbal memory in the problem-solving tasks for girls. The authors conclude that

> "The high proportion of significant correlations across learning and problem-solving tasks reveals the common dependence of the two categories of tasks upon Ss' ability to acquire new information and to apply this information on subsequent trials and in the solution of new problems." (p. 47)

There is considerable difficulty in drawing any conclusions from these studies with respect to human abilities important to problem-solving. The primary reason for this is the lack of similarity among the problem-solving criteria used in the different studies. The most that can be said is that some verbal reasoning and memory abilities appear consistently across studies. In order to establish firmly what abilities are important to problem solving, it is necessary to have comprehensive and representative measures of human problem-solving which are operational definitions of the definition established earlier in this paper.

Characteristics of Ideal Problem Solving Tests

Cronbach (1955) was concerned with the testing situation which would elicit the maximum response. Thus, he proposed that problems must be meaningful to the subject, must not degenerate into exercises which make no demands on the higher mental processes, and that the problem setting should be considered carefully to remove distracting factors. Ray (1955) suggested two additional requirements for problem-solving tests; that they should have a scoring continuum rather than a simple pass-fail score; and that they should provide the test administrator with the maximum amount of knowledge about what the subject is doing. John (1957) put forth a number of criteria for problem-solving tests in accordance with his dictum that the problem-solving process should be accessible for direct observation. He proposed that such tests should (1) start the subject with a standard minimum of information about a problem and then require him to structure his own presolution behavior with a minimum of externally imposed constraints; (2) be maximally free of special skills, special knowledge, or experiences peculiar to a given culture; (3) be constructed so that the effect of familiarity with the generic tasks is minimal in order to facilitate the construction of equivalent forms;

and (4) have a format which presents the subject with the necessity to interact with real events rather than abstract.

These criteria have been principally concerned with problem solving in a laboratory setting. Keisler (1969) took a different point of view, in that he was interested in tests usable in the classroom. His criteria are summarized in the following 11 points:

1. The population of problems sampled by the test should be appropriately defined.
2. The population of problems should be an important one.
3. The test should not be inadvertently confounded with assessment of prerequistie learnings.
4. The conditions under which the test is given must be clearly specified.
5. The test should indicate how well the child was following the appropriate procedure leading to the solution.
6. Increasing number of cues should be used to find out how much assistance the learner requires.
7. On some occasions the student should be required to provide an oral or written account of his own covert procedures, assuming such self-report

does not interfere with the problem-solving process.

8. Quality of problem solving can be assessed through the records of procedures and solutions.
9. Problem setting, as distinct from problem solving, should also be assessed under standardized conditions.
10. The effects of affective attitude change should also be considered.
11. At a practical level, a test of problem solving should be amenable to group presentation if at all possible.

Covington (in press) reflected these concerns by concluding that techniques are needed which reflect, as fully as possible, the rich complexities of creative thought, which allow for the distinctive dispositions and cognitive styles of the creative person, but which, at the same time, permit a reasonable degree of standardization and the use of objective scoring procedures.

In terms of concerete criteria for problem-solving tests all of the points made by the above writers appear to be summarized in the following:

1. Tasks selected for problem-solving tests should be complex; i.e., they should not be merely simple exercises, but rather problems in which

there are a large number of steps from an initial state to a final state, or a reasonably large number of attributes.

2. Performance on the test should be minimally related to previous learning which could differentiate individuals at the time of the test.
3. The problems should command the attention and interest of the subject so as to insure an adequate level of motivation for optimum performance.
4. The test should yield a variety of continuous measures concerning the outcomes of problem solving, the processes, and the intellectual skills involved.
5. The test should contain a minimum number of constraints on the types of problem-solving behavior the individual may engage in.
6. The test should demonstrate both reliability and validity.
7. The test should be practical for group administration.

Problem Solving Tests

A number of tests have been developed to measure problem solving in general or in specific areas, but it is debatable whether or not any meet the above criteria for an ideal test. Feldhusen et. al. (1971) report four classes of tasks present in the literature which have been used to investigate problem solving. These are (1) puzzle-insight problems, (2) process problems, (3) component tasks, and (4) real-life, relevant tasks.

The first set of tasks, the puzzle-insight problems, have been the tools for most of the classical investigations of problem solving. Perhaps the most famous of these are Maier's Two-String and Hatrack (1945) problems, Duncker's (1945) box problem, Luchin's (1942) Water Jar problems, Katona's (1940) Matchsticks problems and anagrams (Johnson, 1966). All these tasks are basically artificial intellectual games in which the initial conditions and final goal are precisely stated, and there are a severely limited number of ways to go between the two.

The second type of problem task, the process problems, have been used in research which focuses primarily on the processes of problem solving. This type includes switchlight problems (John, 1957; Tyler, 1958; Davis, 1966; Davis, Manske, and Train, 1968), the verbal maze problems of Hayes (1965), simulated problem situations, (Glaser, Damrin, and Gardner, 1954; Rimoldi, 1960; McGuire and Babbott, 1967; Streufert, Kliger, Castore, and Driver, 1967; Nattress, 1971), and concept identification (Bruner, Goodnow and Austin, 1956;

Clark, 1971; Bourne, Ekstrand, & Dominowski, 1971). The common element to each of these tasks is a structure where there are a number of discrete decision points where an individual's choices can be recorded. The result is a record of all a subject's decisions throughout the course of the problem which can be studied and processes and strategies inferred.

The component skills and abilities involved in problem solving are the primary concern of the type labelled "component problems." Included in this category are the battery of tests used by Guilford (1967) to establish the Structure of Intellect model of intelligence, the Torrance Tests of Creative Thinking (TTCT; 1966a), Unfinished Stories (Lundsteen and Michael, 1966), and the Purdue Elementary Problem Solving Inventory (Feldhusen et. al., 1971). This type of test is based on a theoretical approach which maintains that problem solving is carried out by the efficient employment of a wide variety of skills. Thus in order to measure problem solving behavior, it is necessary to observe behavior related to each one of these skills. Consequently, each of the tests is designed to tap one specific ability by presenting highly specific problem situations which precisely defined initial conditions and criteria for problem solutions.

The fourth type of problem solving test consists of those tests which place great emphasis on problem situations which represent "real-life" situations and are relevant to the concerns of the student. Tests of this type are Crutchfield and Covington's (1966) The Old Black House, The Man in the Pit, and The Missing Jewel, Miles (1968) Creative Design Test, and Treffinger's (1970) Solving Problems #2. These are tests similar to the component type in that the measures are the same but they place greater emphasis on motivating Ss to perform as well as possible by solving relevant problems.

Evaluation of Problem Solving Tests

The instruments from these four categories of problem-solving measures, however, do not necessarily meet all of the criteria for an ideal problem-solving test. With respect to the criterion of problem complexity the puzzle-insight problems do not qualify, since the tasks are fairly simple in nature, having only a few components to work with. The other three types do meet this criterion in that for any of these problems the structure may be as complex as desired. As an example consider the medical diagnosis simulation. When the S is first given the problem he has a number of choices as to action and contingent upon this choice is additional information and other choices which lead through a large network of decisions to a final diagnosis and treatment.

The second criterion, that of effect of previous learning, is particularly important to those types of tests which are highly task specific in nature. Most of the puzzle-insight problems have a relatively simple if unusual solution or strategy for solution which is easily remembered once it has been encountered. Thus for those people who have encountered the task before it becomes a test of memory rather than a test of problem solving. This would appear to be a difficulty with all those which strive for relevancy, since if the problem situation resembles too closely past experience, memory becomes the dominant influence rather than more complex, and perhaps more important, skills. Those tasks which depend little on past experceince or on a level of experience common to all Ss, such as concept identification, verbal mazes, some of Guilford's tests, or the TTCT, should be more able to tap skills other than memory.

Unfortunately, those same tests which are the least affected by past experience are those most likely fail the third criterion, an adequate level of motivation. The prime purpose of this criterion is to insure a generally uniform level of motivation to undertake the problem task. When the task cannot be related to past experience, other types of motivating forces, such as intellectual curiosity, social pressure, or experimenter pressure must be relied upon to produce the motivating force. Yet these vary greatly from person to person in terms of susceptibility so that a much larger range of motivation is produced and there is consequently a spuriously increased range of performance on the task. Thus those tests which are least likely to hold the interest of the student such as the matchstick problems, concept identification problems, or the TTCT are most likely to fail to meet this criterion, while the medical simulation, the war game simulation, the PEPSI, and The Old Black House are more likely to capture the interest of the student and thus provide more uniform levels of motivation.

The fourth criterion, which is concerned with the measures provided by the test, is particular to the puzzle-insight problems. In order for a problem-solving test to be of maximum utility, it must yield a maximum amount of information about the problem-solving behavior of the student. The puzzle-insight problems have been primarily concerned with time to solution, number of mistakes or inappropriate solutions, and number of hints necessary for solution. These are strictly end-product measures and provide little information about the processes or skills involved in problem solving. The process problems supply a great deal more information, due to the fact that the S is followed through the problem by the decisions he makes at each point. Thus in addition to the end-product measures, information is also available on the strategies used, the efficiency of the strategies, and the sequence of typesof decisions. The component and the "real-life" problems provide a different kind of information. Again these are end product measures, but these have to do with the proficiency of the individual in the individual skills which theoretically facilitate problem solving behavior. Thus it would appear that the process tasks yield the greatest amount of information about actual problem-solving behavior and consequently best satisfy the fourth criterion.

Constraints, the subject of the fifth criterion, are evident in all of the tasks cited above. The only issue is which one demonstrates the least number of constraints, since it is obvious that some constraints are necessary in order that problem solving behavior can occur at all. However, some types of problems such as the puzzle-insight type present a high number of constraints upon the strategies usable for solution or the solution itself. The component problems generally suffer from the same difficulty due to the highly specific nature of the tasks. Perhaps the process problems again best satisfy this criterion since they provide a number of possible paths to solution, though the multiple solution component problems such as the TTCT also have a low number of constraints on the solution.

The reliability and validity of problem-solving tests, the sixth criterion, has not been a subject for intense investigation. In most research in problem solving, the task has been the object of investigation rather than the task as a manifestation of a larger and more generalized set of behaviors. In such cases reliability was unimportant since the individual only performed the task once, and validity was irrelevant since generalization to other types of problems was not attempted. However, some validity evidence does exist. Mendelson, Griswold and Anderson (1966) reported that performance on anagrams was related to a measure of intelligence. A consistent positive relationship between intelligence and concept identification performance has been demonstrated (Hoffman, 1955; Osler and Fivel, 1961; Denny, 1966; Whitman, 1966; Bunderson, 1967). The validity of the TTCT has been investigated and found acceptable in a number of studies (Torrance, 1966b). But, for the majority of existing problem solving tasks, the criterion is not satisfied.

The final criterion, practicality for classroom administration, is satisfied by few of the existing measures. Most of the puzzle-insight problems are individually administered as are the verbal mazes, switchlight problems, and concept identification tasks. The Matchstick problems (Katona, 1940), anagrams, simulated problem situations and all of the component and "real-life" problems exist in a form for group administration. However scoring is often a difficulty with these tests, since responses to such tests as the TTCT, and the Creative Design Test, must be rated by judges in order to derive numerical scores. Thus it would appear that most of the tests are beset with problems in relation to practical concerns.

In choosing a representative set of tests of problem-solving behavior according to the stated criteria, it would appear that those measures in the process category are best qualified as ideal problem-solving tests. That is, they should yield the maximum amount of information about the problem-solving process and reflect the utilization of a sizeable number of human abilities and skills. The component problems, while they appear to qualify by most of the criteria, are not useable due to the fact that they are based on pre-existing theories concerning the skills involved in problem-solving behavior.

Yet as stated above, the tests in the process category are far from perfect. A number of improvements need to be made before they would come close to meeting the criteria for an ideal test. The verbal maze problems need a higher level of motivation, evidence for reliability and validity of the measures yielded, and a form which is amenable to a large group testing situation. The simulated problem situations satisfy most of the criteria except that they need to be constructed so as to rely less on past experience, and be supported by more evidence for their reliability and validity. The concept identification tasks need a higher level of motivation, and a form which is useable in classroom testing. Finally all of the tests are restricted to single solution problems while for purposes of complexity and verisimilitude multiple solution problems would be appropriate.

Thus it would appear that a set of tests which would provide the most representative sample of problem solving behavior are the process problems, specifically modifications of the verbal mazes, the simulated problem situations, and the concept identification tasks.

Summary.

Problem-solving has been defined as the process by which the problem solver starts from an initial state and proceeds to a goal, where the specific means by which this is accomplished are not initially known to the problem solver. The models of problem solving which were discussed, hypothesized that a number of different processes and human abilities were involved in problem solving, i.e. it is a complex cognitive skill. The empirical studies of human problem-solving abilities have revealed that this hypothetical assertion has some basis in reality. Finally a number of criteria for a sufficient operational definition of problem solving were discussed, and it was concluded that tests which are concerned with processes in problem situations met most of these criteria.

Problem Statement

The purpose of this study is to establish a set of human abilities which are related to efficient human problem solving. Specifically, this study will build upon the results of Merrifield, et al. (1962) by attempting to relate a number of human abilities as specified by the SI model of Guilford (1967) to a set of problem-solving measures which meet the criteria for an ideal problem-solving test. These SI tests will be concerned with all the operations applied to semantic content that result in products identified as relations, transformations, and implications, through which in certain cases may include units, classes and systems. The only content dealt with is semantic since the tests must necessarily present problem situations through verbal statements and responses must be made in the same manner. All operations are included since it is possible to reasonably identify actions taken in problem solution which might involve cognition, memory, divergent and convergent production, and evaluation. Also the products emphasized are relations, systems, transformations, and implications since these products would appear to be more important to complex problem situations. Though, these tests are specified by the SI model, they include tests representative of almost all factors important to problem-solving as cited in the review of empirical studies.

Finally, the population of interest will be fifth grade public school students. Children at the age level associated with this grade have generally mature cognitive skills. They are able to deal with verbal material with reasonable ease (McCandless, 1967) and probably lack the background of experience in solving problems which might obscure the workings of problem-solving strategies. Thus problem-solving tests for this age level should yield a valid sample of problem solving behavior.

In order to fulfill the purposes of this study the following specific objectives have been formulated:

1. To adapt the verbal maze, the simulated problem situation, and concept identification tasks with both single and multiple solutions

to group administration with elementary school students.

2. From the test results gathered on a set of SI tests and the problem-solving criteria, to determine the simple relationship between each of the abilities and the measures from the criteria.

3. To determine a structure underlying performance on the ability measures and the criteria.

4. To determine the predictability of performance on the criteria from knowledge of the abilities.

METHODOLOGY

In this chapter the development of the criterion instruments will be described in terms of the samples used, the developmental procedures followed, the content of the instruments, the scoring procedures, and the collection of reliability data. Then the primary study will be described in terms of the sample employed, the instruments used, and the analyses carried out on the collected data.

Criterion development

Samples.

Two different samples were required. The first sample consisted of approximately 30 fifth grade students from a rural Indiana elementary school. This sample was employed in the preliminary testing of the criterion instruments. The second sample consisted of 140 fifth grade students from two upper middle-class elementary schools. These students comprised six classes, three in each school. This sample was used for determining the test-retest and alternate forms reliability of the criterion instruments. Two classes, one from each school, were used in collecting data for each of the three types of instruments.

Instrument development.

Simulated problem situations. The simulated problem situations were based on the work done by McGuire and others (McGuire and Babbott, 1967) at the Center for Educational Development, University of Illinois College of Medicine in Chicago. The prototype of this instrument was originally used to measure physician's competencies in patient management situations. The rationale for use of this type of instrument was that it provided problem situations which could be relevant to the concerns of the population of interest and at the same time yield a number of measures of the problem-solving process.

Four forms of the simulated problem situation were developed: a multiple solution problem, a single solution problem, and an alternate form for each. The problem situations chosen were as follows:

1. The New Bike - multiple solution problem. In this problem the S is presented with a situation where he desires to purchase a bicycle he has seen in a store window. A number of ways of obtaining the money are made available to the S, and he must choose among these in such a way as to earn the necessary amount of money in a minimum amount of time.

2. Free Ice Cream - multiple solution problem, alternate form. In this form, the S is presented with a situation where the local ice-cream store is offering a prize of a year's worth of free ice-cream in an ecology drive to clean up the town. In order to win the prize, the S must earn a given number of points, points are earned by planting trees, recycling a variety of materials, or picking up trash in public places. Again a time limit is set.

3. The State Fair - single solution problem. In this form the problem situation is one where the S wants to go to the State Fair with friends, but must find the price of admission. The money is somewhere in the S's house, but the location is not known. The S must search through the house in order to find the money which is hidden, taped to the bottom of a dresser drawer in the S's bedroom.

4. The Missing Friend - single solution problem, alternate form. In this form the problem situation is one where the S is supposed to take a friend to a club meeting after school so that the friend may join the club. However, the friend is not at school in the afternoon. The S must

find the friend and take him to the meeting before the club adjourns for the afternoon.

After choosing the problem situations, a problem map was created for each problem. Figure 6 details the problem map for The State Fair. The purpose of the problem map is to specify all the steps the S may take in solving the problem, and the order that they may be taken. The problem map involves three different symbols. Squares denote information provided to the S. The hexagons denote a decision that the S must make. The triangles represent points in the problem where the S may quit or where a solution has been found. Arrows indicate the direction of movement through the problem map.

To indicate how this works, refer once again to Figure 6. The box entitled Start specifies that the S is given the statement of the problem; the S is told that he must find the money in order to go to the State Fair. Then the S proceeds to a decision point where he must decide to do one of three things: (a) look in the hall closet, (b) think of possible places the money might be, or (c) ask mother for the money. The first and third choice give the S additional information and direct him back to the decision point, with the instruction to make another choice. The S

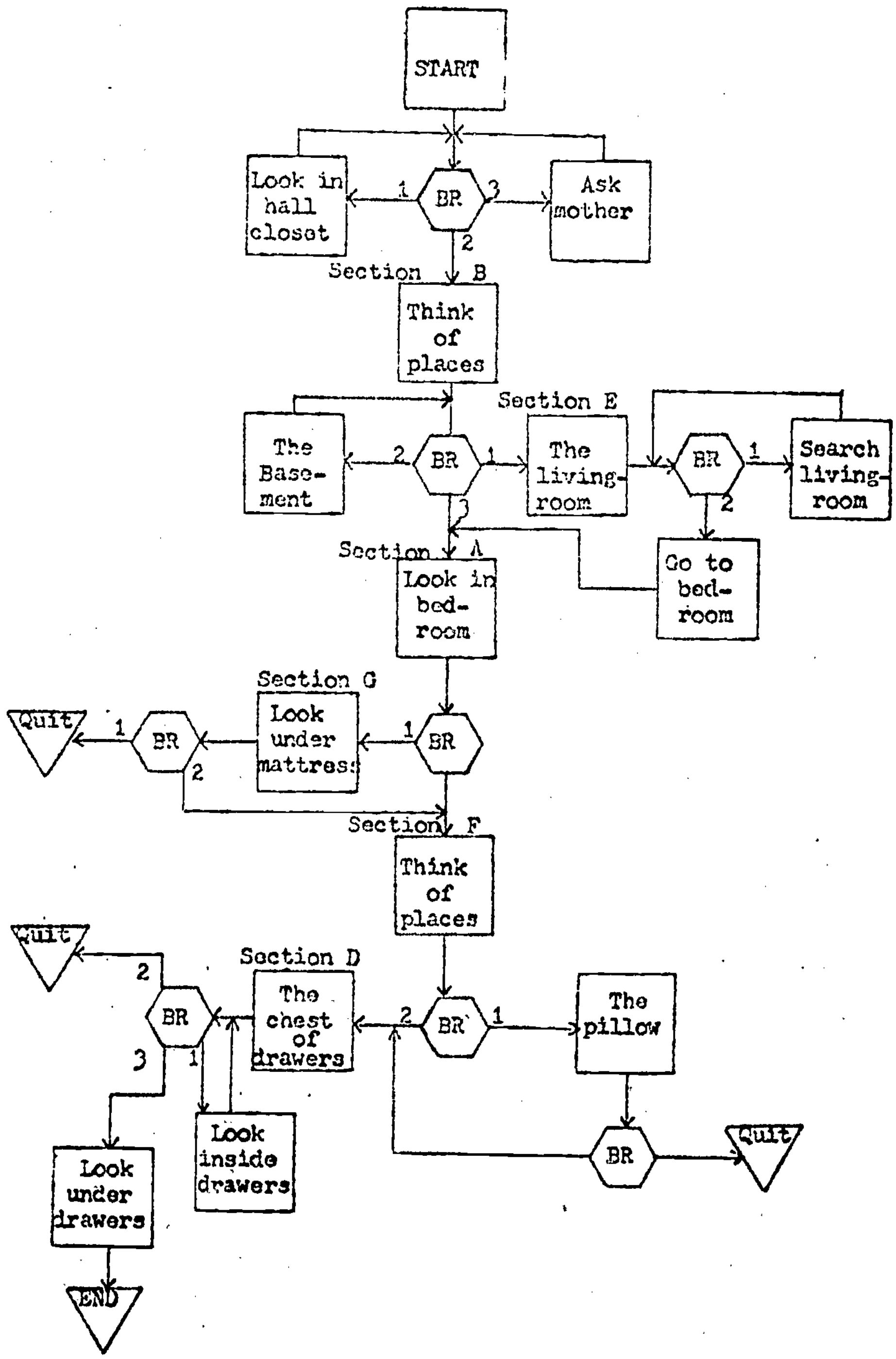


Figure 6. Problem map for The State Fair.

proceeds in similar way through other information boxes and decision points until he reaches an exit box. It should be noted that in the cases of The New Bike, and Free Ice Cream that information boxes also sometimes direct the S to the fact that he has earned a certain amount of money or points, and that it has taken him a given number of days. It should be noted that the path taken by different S's will depend on the choices he makes, and may differ from S to S.

The next step in the instrument development process was to put the problem map into a format suitable for classroom administration. For this purpose, the problem map was divided into sections containing a decision point and one or several information boxes. These sections became pages in a test booklet. Figure 7 is a reproduction of the first section from The State Fair. The initial paragraph gives the pertinent information for that section. After reading this, the S must make a decision based on this information plus any other information he may have collected. Decisions are made by choosing one of the alternatives listed at the bottom of the page. The box beside each alternative then gives the S feedback directions on which section to go to next, how much money or points he may have earned and days he may have spent by choosing the alternative, or it tells him that he is finished with the problem.

Figure 7. Sample section of The State Fair simulated problem.

THE STATE FAIR

Let's pretend that somewhere in your house you have hidden a $5 bill. You have been saving it for something special. Now something special has come along - the State Fair starts today. There will be lots of rides and other fun things to do. Your friends are all planning to go to the fair this afternoon. You need the money so that you can get in with your friends and enjoy the rides. There is only one problem - you cannot remember where you hid the money. You are sure it is somewhere in the house but you don't know exactly where it is. Which of these things would you do first in order to get the money to go to the fair?

1. Look in the hall closet because closets are good places to hide things.	1. There is nothing in the hall closet but coats. Make another choice.
2. Sit down and think of some of the places you might have hidden the money.	2. Go to Section B.
3. Ask your mother for the $5 since she just got paid two days ago.	3. Your mother will not give you the money since it must be used to buy food. Make another choice

Figure 7

A special process was employed to provide the informative feedback given in the boxes next to each alternative. In order to prevent the S from reading all the informative feedback, and then choosing the most reasonable alternative, all information in the feedback boxes in each section was printed invisibly on the page. This was accomplished through the use of the Latent Image Process system manufactured by the A. B. Dick Company. In this method normal DITTO Masters are prepared with the visible printing, and then the printing intended to be invisible is placed on the master using regular master and a special imprinting sheet, the Latent Image Master. When the resultant master is run on a duplicating machine, the printing which is intended to be invisible is imprinted into the page, without becoming visible. A special pen, called the Latent Image Developer Pen, is used to bring out the invisible printing. The pen is rubbed over the area where the printing is located and the printing becomes visible against a light yellow background. Since the S develops only those feedback boxes corresponding to the actual decisions he makes, a trace of the decisions which the S makes while solving the problem is preserved.

The next step in the developmental process involved trying out the preliminary version of the instruments on a small group of fifth graders. Feedback was collected on a variety of topics such as readability, clarity of directions, necessary time limits, workability of the Latent Image materials, the amount of assistance needed by the Ss. The feedback information gathered from the first testing session was used to revise the instrument. The most significant revision was the creation of a sample of the instrument to acquaint the Ss with the workings of the Latent Image process and how to follow directions given in the instrument. This tryout and revision process was carried out twice more after this first trial, before the instrument was judged ready for reliability testing.

In order to score the simulated problem instruments, the order of developed responses as they appeared in the test booklet was transferred to a mark sense scoring sheet. These score sheets were read, and punched cards were produced which contained an ID and a number of punched columns corresponding to the possible number of choices for that particular form. Thus for The State Fair there were 30 columns since the problem has 30 possible choices. Each column contained a 1 if the corresponding feedback box was developed and a 0 otherwise. These cards were submitted to a computerized scoring program which calculated a number of different measures depending on the particular form of the test.

For the two multiple solution instruments, The New Bike and Free Ice Cream, the scoring program calculated nine measures. The first two variables dealt with the type of finish; the first indicated whether the S quit, stopped or successfully solved the problem; the second indicated whether or not the S had met the conditions of a successful solution. The third variable was the number of dollars or points earned. The fourth variable indicated the number of days the S used in earning the points or money. The fifth variable was a strategy measure. This was calculated by comparing the decisions made by the S with those necessary for a sure logical solution to the problem. The number of decisions in common divided by the total number of decisions in the sure logical solution times 100 became the strategy measure. The sixth measure served as an indication of the amount of information used by the S. The measure is also a ratio of the S's decisions to seek information to the total number of possible information seeking decisions, multiplied by 100. The seventh calculated measure indicated the number of decisions the S made which were inconsistent with the information the S possessed or direction, he was given. The score was the total number of inconsistent decisions the S made while working on the problems. The final measure derived from the instrument was an indicator of the total number of decisions the S made in solving the problem. Due to the fact that the S was allowed to pass through several decision points more than once, the total number of steps had to be inferred rather than actually counted. These inferences assumed that at these particular decision points the S always performed logically and according to the given directions. Total time to solve the problem was also collected.

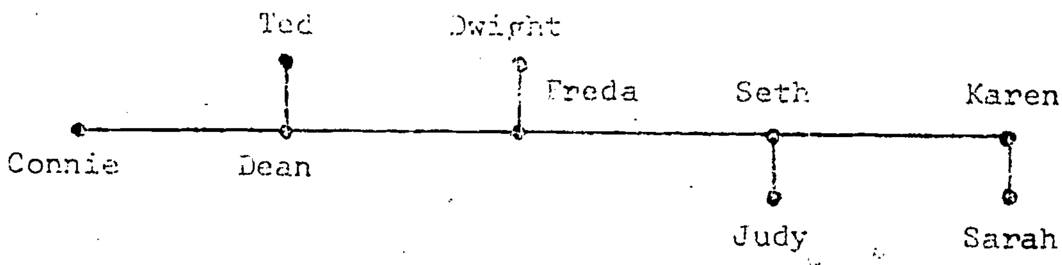
For the two single solution problems, The State Fair and The Missing Friend, the scoring program yielded five measures. The measures consisted of the finish type, the strategy, inconsistency, number of steps, and time scores. The calculations of these scores were exactly the same as described in the previous paragraph.

Verbal maze problems. This set of problem-solving criteria was adapted from the tasks proposed by Hayes (1965). In the original tasks, Ss learned lists of paired names to a criterion of three perfect trials. They were then told that the name pairs represented links between pairs of spies. They were then presented with a series of problem tasks in which the names of two spies were given and the S was asked to find a route to get the message from the first to the second. Ss were required to carry out these tasks using only what they could "work out in their heads."

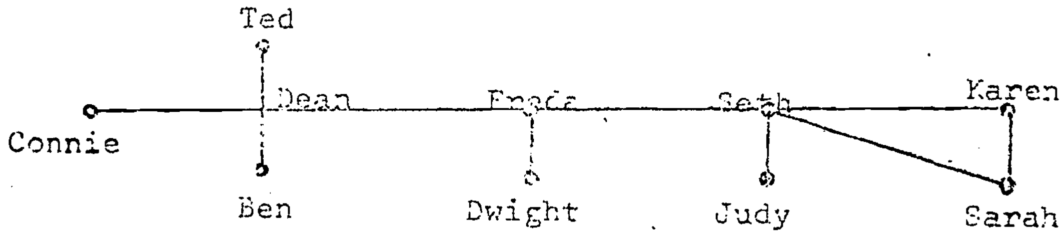
In order to adapt this type of task for use in elementary school classrooms, it was necessary to develop a group administered mode. The objective was to develop four forms in this mode: a single solution problem, a multiple solution problem and an alternate form of each. In order to design these forms, it was necessary to develop the linkages between names which could again be portrayed as problem maps. The maps for the four problem are given in Figure 8. The nodes in these maps represent the spies who are linked by the interconnecting lines.

After the maps were specified, common English first names were assigned to each of the nodes. These names were then formed into pairs based on the linkages given in the maps. Test booklets were then constructed with an example, a single solution problem and a multiple solution problem included. The format of each problem consisted of a brief introductory paragraph giving the name of the spy sending the message and the name of the spy who was to receive it. Then followed a list of spy name pairs which completely described the problem map. The rest of the page consisted of lines of connected boxes. Figure 9 is an example of such a page. The S was told to use each line to attempt to find a solution path by writing connected spy names in successive boxes. If the S came upon a dead-end, he was instructed to restart on a new line. Finally Ss were instructed to write down the clock time when they had found a solution.

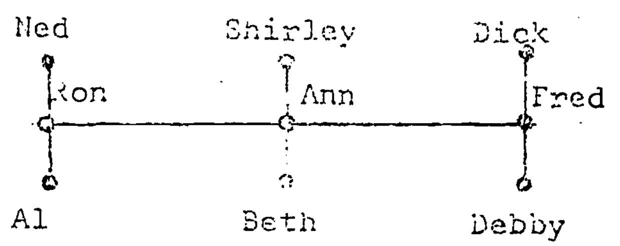
1. Single solution Verbal Maze problem map.



2. Multiple solution Verbal Maze problem map.



3. Single solution Verbal Maze problem map - alternate form.



4. Multiple solution Verbal Maze problem map - alternate form.

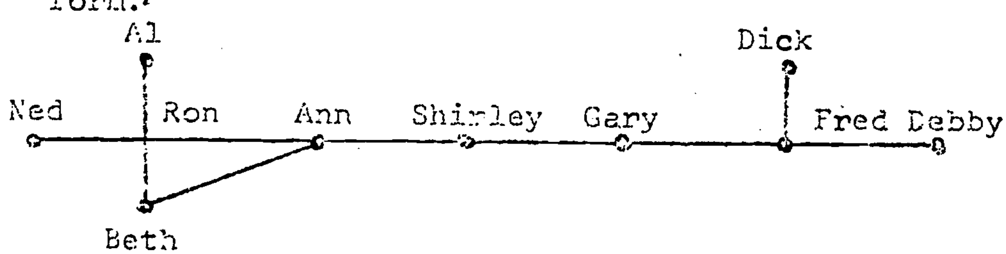


Figure 8. Problem maps for the Verbal Maze problems.

With some groups of spies, there may be more than one way to pass the message through the ring. Below is a list of spy-name pairs that contains more than one path. Using the same directions as before, try to find as many paths as you can for the message to travel from Judy to Ted.

Seth -- Freda
Dean -- Ben
Judy -- Sarah
Freda -- Dean
Dwight -- Ben
Dean -- Connie
Sarah -- Seth
Freda -- Dwight
Seth -- Karen
Dean -- Ted
Judy -- Seth

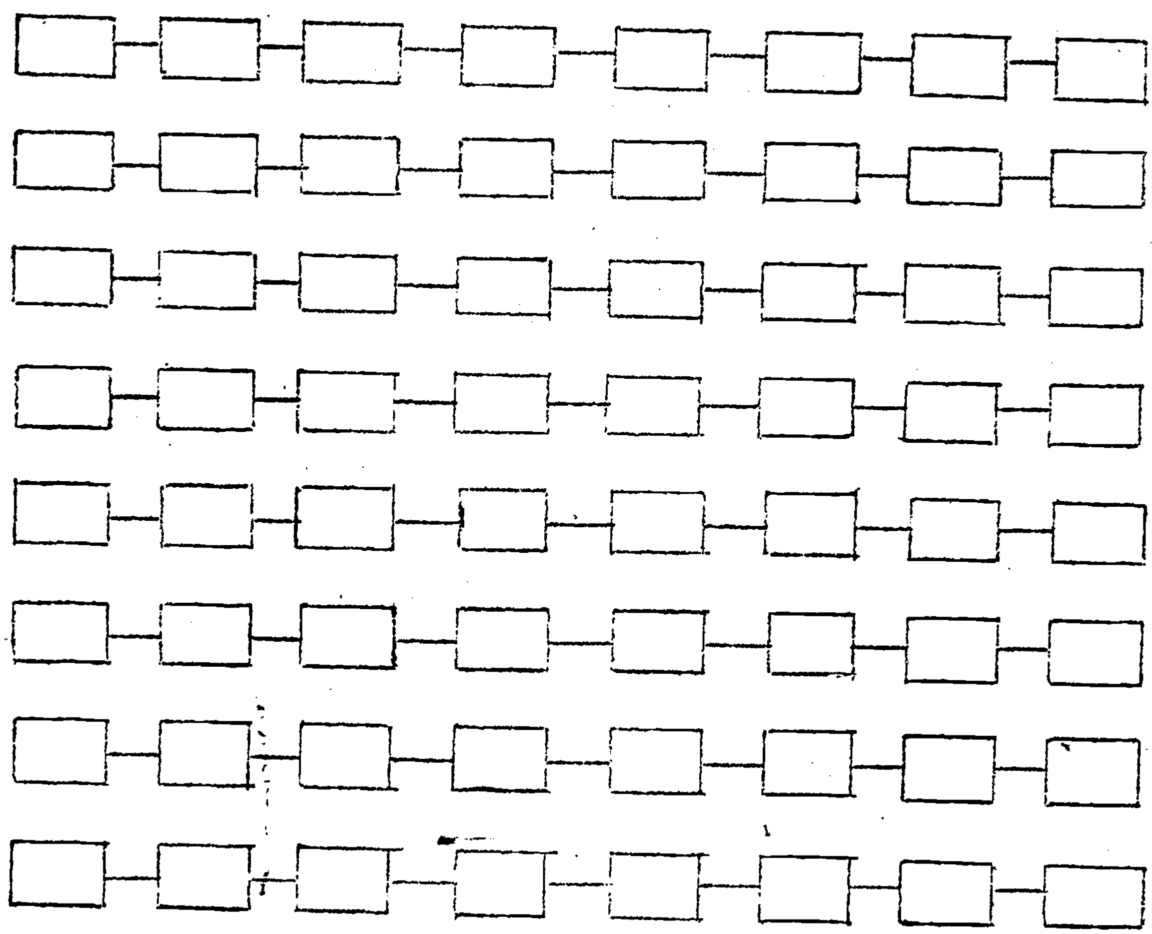


Figure 9. Example of a Verbal Maze problem.

The first developed forms of these instruments were administered to the tryout sample. Information was collected with respect to clarity of directions, ease of utility of the problem format, and time requirements for completion of the task. The forms were revised twice with intervening tryouts, before the instrument was judged ready for reliability testing.

The verbal maze instrument yielded three measures for the combined single and multiple solution problems. The first of these measures was the total number of trials used in finding all solutions. The second measure was the total time taken to find all solutions. The third and final measure was the total number of solutions found in both the single and multiple solution problems.

Concept Identification Problems. The primary objective in the development of concept identification tasks was to construct a suitable format for group administration of concept identification tasks. First, however, the concepts had to be defined. In parallel with the other developmental efforts, four target concepts were identified. Exemplars consisted of single abstract geometric forms with four bi-valued dimensions. The dimensions were: (1) Large - Small, (2) Square - Circle, (3) Two holes - Four holes, (4) White - Crosshatched. Two dimensions were relevant and two were irrelevant. The four target concepts consisted of two conjunctive concepts and two exclusive disjunctive concepts. The conjunctive concepts were LURBs (Two hole, Squares) and FLIXes (Four Hole Circles). The disjunctive concepts were DROOGs (Crosshatched or Square but not both) and RILGs (White or Square but not both).

The presentation mode was modeled after the selection paradigm (Bourne, et al., 1971). Thus all permutations of the exemplars were presented simultaneously on one page. Figure 10 details the form of each exemplar. The letter identifies the particular exemplar.

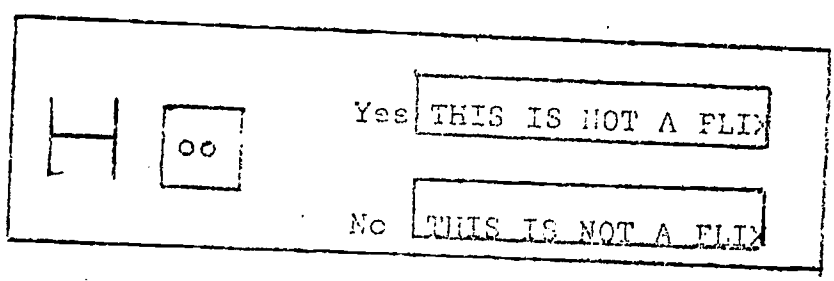


Figure 10. Typical exemplar for the concept identification task.

Feedback was provided to the S by means of the boxes following the YES and NO. For each exemplar the S had to decide whether it was or was not an example of the target cocnept. If it was, the S chose the YES box, otherwise the NO box. The feedback was provided in Latent Image form and the S indicated his decision by coloring in one box or the other with the Latent Image Developer Pen. The revealed printing told the S whether or not the exemplar was or was not and example of the target concept. For each problem, one positive exemplar was chosen as a focus. In order to record the sequence of exemplars the S considered, he was required to write the identifying letters across the top of the page in the order he looked at the exemplars. Ss were required to respond to all 16 exemplars. All four forms are presented in the Appendix.

The four forms were developed and administered to the tryout group. Feedback was collected on the clarity of the directions, the difficulty of the problems, and usability of the problem format. It was found that the directions were quite difficult to understand, and so sample problems were constructed to explain each problem and familiarize the Ss with the format. This revised form was tried out twice more and revised before it was judged ready for the reliability testing.

The concept identification tasks yielded four measures. In order to obtain these measures, the sequences of exemplars as they were considered by S and the decisions made were transferred to punched cards. These cards were submitted to a computerized scoring program which calculated three of these measures. The first measure was the number of errors in exemplar identification made by the S. The next two measures were based on the scoring procedures detailed in Laughlin and Jordan (1967) and were measures of the focusing and scanning behavior of the S. In order to calculate these measures it was necessary to assume that the number of errors was equivalent to the number of exemplars used by S to identify the target concept. The final measure was the total elapsed time the S took to work his way through the 16 exemplars.

Reliability Study.

Procedures. Two types of reliability measures were obtained for each of the three types of problem-solving criteria. These were immediate alternate forms reliability and two week test-retest reliability. Table 1 summarizes the types of reliability information obtained for each form.

Table 1

Summary of the types of reliability information obtained for each problem solving criterion.

	Test-retest	Alternate Forms
Simulated problem situations		
1. The New Bike	Yes	Yes
2. Free Ice Cream		Yes
3. The State Fair		Yes
4. The Missing Friend		Yes

Table 1 (continued)

	Test-retest	Alternate Forms
Verbal Maze Problems		
5. Single solution #1	Yes	Yes
6. Multiple solutions #1	Yes	Yes
7. Single solution #2		Yes
8. Multiple Solutions #2		Yes
Concept Identification Tasks		
9. Conjunctive LURBs	Yes	Yes
10. Disjunctive DROOGs	Yes	Yes
11. Conjunctive FLIXes		Yes
12. Disjunctive RILGs		Yes

In the first testing session, two classes (one from each of the sample schools) received The New Bike and Free Ice Cream problems, two classes received the single and multiple solutions verbal Maze problems #1 and #2 and two classes received all four concept identification tasks. Due to procedural difficulties, The State Fair and The Missing Friend were not administered. For each type of problem-solving criterion, the test administrator first introduced the problem tasks. Then a short example of each task was worked out with the class as a whole, to make sure that Ss understood the nature of the task and the mechanics of the testing situation. Ss were instructed to write down the clock time after they had solved each problem. Testing sessions were limited to a period of 60 minutes.

In the second testing session the test-retest data was collected. The first two classes again received The New Bike plus two other problem solving tasks. The second two classes again received the concept identification tasks with targets conjunctive concept LURB and disjunctive concept DROOG, plus two other problem solving tasks. In the third set of two classes, one of the classes again received the Verbal Maze problem single and multiple solutions #1 plus another problem solving task, while the other received the same verbal maze task plus The State Fair and The Missing Friend. Again the testing sessions were limited to 60 minutes. However, due to procedural difficulties, not all classes were instructed to write down the clock time upon completion of their problems.

Analyses. The data was subdivided into test-retest and alternate forms reliability groups for each of the problem-solving criteria. Pearson Product Moment correlation coefficients, means, and standard deviations were calculated for each of the test-retest and alternate forms groups for each of the measures listed below:

A. The New Bike and Free Ice Cream

1. Money or points earned
2. Days used
3. Information score
4. Strategy score
5. Inconsistency score
6. Number of steps

B. The State Fair and The Missing Friend

1. Strategy score
2. Inconsistency score
3. Number of Steps

C. Conjunctive Concepts LURBs and FLIXes

1. Number of errors
2. Focusing score
3. Scanning score

D. Disjunctive concepts - DROOGs and RILGs

1. Number of errors

E. Verbal Maze Problems

1. Total time to solution
2. Total number of trials
3. Total number of solutions

Summary. The reliability study of the problem-solving instruments completed the development portion of the project. The criterion instruments had been developed and judged ready for use in satisfying the primary objectives of the overall study.

The Problem Solving Abilities Study

In this section the sample, instrument preparation, and administration procedures will be described. In addition, test scoring procedures and data analyses will be delineated.

Sample.

The sample for this study consisted of 490 fifth grade students from 18 classrooms. These classrooms were located in six schools of the Tippecanoe School Corporation. Ss came from a wide variety of SES levels, but generally shared a common cultural heritage. Ss were generally of average intelligence, but the sample covered a wide range of intelligence levels. The age of Ss averaged 11 years.

Instruments.

The instruments used in this study consisted of two sets. The first set was comprised of 17 tests drawn from the set of tests used by Guilford (1967) to establish the Structure of Intellect model. These are summarized in Table 2. The second set of tests were the problem-solving tasks developed in the first part of this study.

Table 2. A Classification of the 17 SI tests.

Measure	Processes					Products					
	C	M	D	N	E	U	C	R	S	T	I
Word Completion	X					X					
Word Linkage	X							X			
Plane Flight Test	X								X		
Apparatus Test	X										X
Remembered Relations		X						X			
Learned Information		X							X		
Related Alternatives		X									X
Utility Test-Fluency			X			X					
-Flexibility			X				X				
Controlled Associates			X					X			
Expressional Fluency			X						X		
Symbol Production			X							X	
Sentence Order				X					X		
New Uses				X						X	
Sequential Associations				X							X
Unlikely Things					X				X		
Logical Reasoning					X						X
Judging Object Adaptations					X					X	

Word Completion. The test represents the SI ability of cognition of semantic units, or verbal comprehension. Merrifield et al. (1962) found tests of this ability were related to their problem-solving measures. The instrument consisted of twenty words, and the Ss task was to write a short definition for each. To accomodate the developmental level of the population of interest, the instrument had to be modified by substituting a less difficult set of words. However, the structure was maintained by using the same parts of speech and choosing words of the same relative difficulty at the fifth grade level, as the original words were at the adult level. The reliability of the instrument was .82 (Sheridan Psychological Services, 1972). The score was the number of correct definitions.

Word Linkage. This test represents the SI ability of cognition of semantic relations, or the comprehension of verbal relations. This test was included as a parallel to Word Completion. Each item of the test consisted of two words, between which there was a semantic connection. The task of the S was to choose from a list of three words, that word which represented this semantic connection. To accomodate the developmental level of the Ss, several items were eliminated due to difficult words. Thus the test was shortened from 30 to 25 items. The reliability, using the Spearman-Brown

Prophecy Formula was .70 (Sheridan Psychological Services, 1972). The score was the number of correct responses.

Plane Flight Test. This test should represent the SI ability of cognition of semantic systems, a factor sometimes called general reasoning. It was modelled after the Ship Destination Test (Guilford, and Hoepfner, 1971). The ability measured by this test was found to be related to problem solving by both Werdelin (1966) and Bunderson (1967). S's task was to calculate the time it will take an airplane to travel from one point to another on a coordinate map. Travel between any two points takes two hours. Several factors are introduced which can affect this travel time at successive stages in the problem. No reliability information was available on this test so that the communality of .33 was taken as the lower bound of the reliability. The score was the number of correct responses.

Apparatus Test. This test represents the SI ability of cognition of semantic implications. The ability measured by this test was found to be related to their measures of problem solving by Merrifield, et al. (1962). Each item names a familiar object, and S's task is to think of two possible improvements for the object. Four items were eliminated due to the unfamiliarity of the objects. The reliability of the test, using the Spearman-Brown Prophecy Formula was .72 (Sheridan

Psychological Services, 1972). The score was the number of acceptable improvements according to preestablished criteria.

Remembered Relations. This test represents the SI ability of memory for semantic relations, thatis, the ability to store and retrieve information abc .t verbal relations. The test was included because both Stevenson, et al. (1968) and Bunderson (1967) found that memory for verbal associations was related to problem-solving behavior. The S is presented with a list of relations between a variety of objects and is told to study it for 2.5 minutes. The S then turns the page and answers a series of multiple-choice questions concerning the relations on the previous page. Due to problems with time limits, the test was reduced from 40 to 20 items. The reliability of the Test, using the Spearman-Brown Prophecy Formula, was .46 (Sheridan Psychological Services, 1972). The score was the number of correct responses.

Learned Information. This test represents the SI ability of memory for semantic systems. It was again included because of the findings of Stevenson et al. (1968) and Bunderson (1967). The S was presented with a set of paragraphs and given 5 minutes to read and memorize it. Then the S was told to turn the page and reproduce the paragraphs in the order given, to the best of his ability. At the top of the second page was a list of keywords to stimulate recall. Scoring was based on the correct order of the reproduction of the main ideas of each paragraph, with a maximum score of 15 points. Since the original paragraphs concerned the SI model and the vocabulary was of high difficulty, it was decided to substitute a parallel set of paragraphs with a vocabulary suitable for fifth grades. The reliability of the original instrument was .64. Since no reliability was available on the revised instrument, the communality of .23 was used as a lower bound of the reliability.

Related Alternatives. This test represented the SI ability of memory for semantic implications. It was included for the same reasons as the other two memory tests. In this measure the S was given a study page which contained a list of surnames associated with occupations, and given 2.5 minutes to study it. Then the S was told to turn the page and respond to a series of multiple-choice items. Each item gave a surname and the S was required to choose from a list of four objects, the object which was associated with that person's occupation. The score was the number of correct responses. Due to several items which contained occupations quite unfamiliar to fifth grade students, 10 items were eliminated. Thus the test was reduced from 30 to 20 items. The reliability, using the Spearman-

Brown Prophecy Formula, was .49 (Sheridan Psychological Services, 1972).

Utility Test. This test represented two different SI abilities, divergent production of semantic units and classes. These are more commonly called ideational fluency and spontaneous flexibility. Merrifield et. al. (1962) reported that divergent production of semantic units was related to their problem-solving tasks. Harootunian and Tate (1960) found that spontaneous flexibility was correlated with their particular measure of problem solving. Thus, this test was included in the study. The items are of a very simple nature. The S is asked to think of as many uses as he can for a brick and a wooden pencil. The responses are scored for number of relevant responses, a measure of ideational fluency. They are also scored for the number of different ideas the S produces; a measure of spontaneous flexibility. The reliabilities of these measures are respectively .74 and .64.

Controlled Associations Test. This test represents the SI ability of divergent production of semantic relations. Merrifield et al. (1962) reported that this ability was related to their measure of problem-solving, and thus it was included in the study. In this test, the S was given eight words and told to write down as many words as he could that meant the same or about the same as the given words. The score was the total number of words that did, in fact, have a similar meaning as the given words. The reliability of the instrument was given by Sheridan Psychological Services (1972) as .70.

Expressional Fluency. This test represents the SI ability of divergent production of semantic systems. It was included in the study in order to discover if the higher order divergent production abilities were related to efficient problem solving. The form of the task was that the S was given a sequence of four letters. He was then asked to produce as many sentences of four words as he could, in which each word started with one of the four letters and the words were in the same order as the letters. The S's score was the number of sentences he could produce which met the stated criteria. Sheridan Psychological Services (1972), give the reliability of this test as .67.

Symbol Production. This test represents the SI ability of divergent production of semantic transformations. This was included for the same reason as Expressional Fluency. In this test the S was given a series of brief statements such as "Airplane takes off". He was then asked to produce an abstract symbol for "airplane" and another abstract symbol for "takes off". Due to time limit considerations, the instrument was reduced from

61 to 33 items. The S's score was the number of symbols which were abstract, and in the judgement of the scorer, represented the object or phrase. Reliability, using the Spearman-Brown Prophecy Formula, was .77 (Sheridan Psychological Services, 1972).

Sentence Order. This test represented the SI ability of convergent production of semantic systems. It was included because Merrifield et al. (1962) reported that higher order convergent production abilities were related to problem solving, but had not included this ability in their study. In this test, the S was presented with sets of three sentences about a particular subject or event. The S was required to determine the proper order of the sentences by numbering them in the order they should appear. The S's score on this test was the number of sentence triples he ordered correctly. Sheridan Psychological Services (1972) gives the reliability of this test as .56.

New Uses. This test represents the SI ability of convergent production of semantic transformations. Merrifield et al. (1962) reported that this ability was related to their problem-solving measures. In this test the S was presented with a picture containing a number of objects in a given setting. Below the picture were listed a number of functions. The S was asked to find objects in the picture which could fulfill the listed functions, with the restriction that the function would be a new or unusual use of the object. The S's score on the instrument was the number of items to which he responded successfully in accordance with the given criteria. The reliability of the test, as given by Sheridan Psychological Services (1972) was .47.

Sequential Associations. This test represents the SI ability of convergent production of semantic implications. Merrifield et al. (1962) also reported that this ability was related to their problem-solving measures. In this test, the S was given sequences of four words. The task was to number these words in such a way that there existed a semantic connection between suceeding pairs of words. The S's score consisted of the number of sequences which were correctly ordered. The reliability of the test was .75 (Sheridan Psychological Services, 1972).

Unlikely Things. This test represents the SI ability of evaluation of semantic systems. It was included in the battery because both Harootunian and Tate (1960) and Bunderson (1967) reported that judgement or evaluation was significantly related to their problem-solving measures. In this test the S was presented with a series of pictures, in which two or more things were out of place, incongruous, or unlikely. Beside each picture four alternatives were given. The S's task was to choose the two alternatives which were <u>most</u> unlikely, incongruous, or out of place. The score was the number of correctly chosen alternatives. Sheridan Psychological Services (1972) gives the reliability as .54.

Logical Reasoning. This test represents the SI ability of evaluation of semantic implications, or verbal reasoning. Three studies found that this ability was related to their problem-solving measures; Harootunian and Tate (1960), Werdelin (1966) and Bunderson (1967). The original form of this test was too long and the vocabulary too difficult for fifth graders. In order to adapt the instrument, every other item was deleted from the original form. Then parallel items were constructed, so that the form of the logical proposition was kept intact, but the vocabulary was greatly simplifed. Thus each item remained a logical premise with four alternative conclusions from which the S must choose the one which logically follows. The instrument was consequently reduced from 40 to 20 items. However, S's score remained the number of correctly chosen alternatives. The Spearman-Brown correction yielded a reliability of .72 for the shortened test. However, since the content of the items was changes the communality of .35 was taken as the lower bound of the reliability.

Judging Object Adaptations. This test represents the SI ability of evaluation of semantic transformations.

It was included in the test battery because Harootunian and Tate (1960) found that a judgement factor was significantly related to their measure of problem-solving. In this test, each item consists of an object name and three alternative uses. The S was instructed to choose the alternative which was relevant and the most unusual, ingenious, or clever. Due to the unfamiliarity of some of the objects names, five items were deleted from the test form. Similarly, the directions required extensive rewriting to bring the vocabulary down to the level of fifth graders. The S's score was the number of correct choices he made. The reliability of the instrument, employing the Spearman-Brown formula, was .42 (Sheridan Psychological Services, 1972).

Problem Solving Criteria. Five problem-solving criterion instruments were employed in this study. These were The New Bike, The State Fair, LURBs, DROOGs, and Verbal Mazes single and multiple solutions #1. These instruments were thoroughly described in the previous section on criterion development. The measures yielded were as follows:

1. The New Bike - number of dollars earned, number of days used, information score, strategy score, inconsistency score and number of steps.

2. The State Fair - strategy score, inconsistency score, and number of steps.
3. LURBs - number of errors, focus score, scanning score, and total elapsed time.
4. DROOGs - number of errors and total elapsed time.
5. Verbal Mazes - total time, total trials and total number of solutions.

Reliabilities of these measures are given in the RESULTS section.

The above 22 tests yielded 31 reliable measures. These measures became the basic data for analysis in this study.

Procedures.

After the SI tests had been revised, the 22 instruments were assigned to one of five one hour testing sessions. Arrangements were made with each of the participating schools for five one hour testing sessions in each of their fifth grade classrooms. These sessions were spaced at one week intervals, over a period of five weeks.

Eight test administrators were recruited, and each was assigned to one or more classes. These persons received pretraining in general test administration procedures and were acquainted with the goals of the project. On each Friday preceeding a week of testing, the administrators met and were introduced to the next week's tests. They were acquainted with the mechanics of each test, and possible testing problems were discussed. In the following week those tests were administered to Ss in the sample. During each testing session, the administrator kept a written log containing anecdotal information about the progress of the testing and appropriate timing information. Table 4 gives the sequence of tests as the Ss received them.

After all testing was completed, the instruments were scored. In the case of the problem-solving criteria, the test responses were transcribed directly to data sheets by clerical personnel, and then prepared for computer scoring. For scoring the SI tests, three scoring judges were recruited. The 17 tests were divided among these three people and each one became exclusively responsible for scoring the assigned tests. The procedure for scoring each test involved three steps. The scorer first became familiar with the test key and scored a sample of about ten tests. The scorer then went over these tests with the person in overall charge of the project, to insure that the tests were being scored properly. Then the scorer finished the entire set of tests before moving on to the next set.

In cases where test scoring required judgement upon the part of the scorers, the reliability of the judging was assessed. After an entire set had been scored, a random sample of 20 tests were drawn and rescored by the same person. These sets of rescores were then correlated with the original scores to determine the reliability of the judges.

Table 4. Test administration sequence and timing information.

Session	Instruments	Administration time (min.)
1	Utility Test	10
	Controlled Associates	12
	Verbal Maze #1	35
2	Expressional Fluency	8
	Remembered Relations	9
	Apparatus Test	14
	Unlikely Things	10
3	Sentence Order	8
	Sequential Associations	6
	Judging Object Adaptations	8
	New Uses	9
	Word Linkage	6
	The State Fair	15
4	Word Completion	7
	Plane Flight Test	8
	Learned Information	7 1/2
	Logical Reasoning	10
	Concept LURB	15
	Concept DROOG	15
5	Symbol Production	8
	Related Alternatives	9
	The New Bike	45

As each S's SI test was scored, it was recorded on a data card for that S. Each S was assigned a unique identification number which was recorded on each test booklet. All this information was then combined with the computer scoring output to make up individual cases of data for each S consisting of all his test scores. Finally, this information was placed on computer cards and only those cases with complete information were selected out for analysis.

Analyses.

The analyses for this study were aimed at providing answers for three basic questions:

1. What are the simple relationships among the 31 measures derived from the SI and problem-solving tests?
2. How predictable are the performances on the problem-solving measures, given the SI tests?
3. What is the nature of the underlying structure of the SI tests and problem-solving criteria?

Information regarding the first question was derived from the simple Pearson Product Moment Correlation coefficients among the 31 measures. First, descriptive statistics were generated for each of the measures. These included means, variances, standard deviations, ranges, and skewness indices. The skewness index was examined for each variable and suitable transformations were carried out for those variables which were highly skewed. Using the set of variables, where the transformed variables replaced their counterparts, the correlation coefficients were calculated.

In order to obtain information regarding the second question, stepwise regression analysis was performed for each of the problem-solving measures. The computer program used for these analyses was the REGRESSION sub-program of the Statistical Package for the Social Sciences (SPSS; Nie, Bent, and Hull, 1970). For each of the problem-solving measures, this program yielded a multiple R and R^2 for each step, the change in R^2 from step to step, and the order of entrance of predictors into the regression equation. In order to determine the significant set of predictors for each problem-solving measure, an F test,

$$F = \frac{(R_1^2 - R_2^2)(N - m_1 - 1)}{(1 - R_1^2)(m_1 - m_2)} \qquad d.f. = (m_1 = m_2),\ (N - m_1 - 1)$$

for the change in R^2 from step to step was applied. R_1 is the multiple correlation for equation 1 which contains m_1 predictor variables. R_2 is the multiple correlation for the second equation with m_2 predictors, and N is the total number of cases in the sample (Draper and Smith, 1966). The significant set of predictors was the set from the last step in the stepwise regression analysis which caused a significant increase in R^2 at the 5% level. The final R^2 was then interpreted as the amount of criterion variance accounted for by the SI tests in the significant predictor set.

The approach to obtaining information concerning the third question, that of underlying structure, was threefold. First, a conventional factor analysis was carried out. An iterated principal factor solution (Harman, 1967) was calculated from the basic correlation matrix (R_{psa}). This yielded 27 factors with positive roots. These eigenvalues were plotted against the factor number as specified in the Scree Test (Cattell, 1966a, 1966b). The scree slope was judged to begin at the sixth factor, and thus the first five factors were considered to be relatively free of error variance. These five factors were then orthogonally rotated by the varimax procedure (Harman, 1967). The resulting solution was judged not completely adequate due to a low hyperplane count in some factors and factor overlap on several tests. An oblique Promax rotation (Hendrickson and White, 1964) was then carried out. In this procedure, a varimax solution with elements a_{ij} is used to generate a target matrix W with elements $w_{ij} = a_{ij}^{M+1}/a_{ij}$. Thus each element of W is the corresponding element of the varimax solution raised to the Mth power with the original sign retained. A least squares transformation is then calculated between the varimax solution and this target matrix W by the "Procrustes" technique due to Mosier (1939) and Hurley and Cattell (1962). The resulting transformation is then applied to the original varimax solution, to yield an oblique reference structure. This structure was then used to compute the primary factor loadings, the intercorrelations among the primary factors, and the test - factors correlation matrix using the formulae found in Harman (1967).

The second approach taken to the question of underlying structure was to perform a canonical correlation analysis (Hotelling, 1936). The first set was the SI tests and the second was the problem-solving criteria. First, canonical roots were extracted and correlated linear functions of the SI tests and the problem-solving criteria were calculated. In order to evaluate the relative importance of the resulting canonical variates, the redundancy for each pair of linear functions as defined by Stewart and Love (1968) was calculated. The redundancy is a non-symmetric index which indicates the proportion of variance extracted by a canonical factor of the first set which may be predicted from the corresponding canonical factor of the second set. To calculate the redundancy coefficients for each set of linear functions, the test-canonical factor correlations (i.e. the canonical factor structure) had to be calculated. This was done by multiplying the matrix of intercorrelations of the SI tests (R_{si}) by the coefficients of the linear functions for the first set. A similar process was carried out for the linear functions of the problem-solving criteria. This yielded a set of factor loading vectors s_i for the SI tests and a similar set t_i for theproblem-solving criteria. The proportion of variance extracted by each canonical factor of the first set was given by the expression $s_i s_i/p$ where p is the number of tests in the first set. A similar expression involving the t_i's and the number of variables in the second set gives the proportion of variance extracted by each factor in the second set. The redundancy for the i^{th} function of the first set is then calculated by the formula

$$Red_i = \frac{\underline{s}_i' \; \underline{s}_i}{p} \times R_{c_i}$$

where R_{c_i} is the canonical correlation for the first set of linear functions. Again, the redundancy for the i^{th} function of the second set is calculated by a similar formula. These redundancy measures were then used to evaluate the importance of the pairs of linear functions.

The third approach in investigating the underlying structure was to determine the relationships of the problem-solving criteria to the SI structure represented by the SI tests. In order to accomplish this, it was necessary to attempt to force the SI tests into independent factors, as postulated by the SI model. This was accomplished by extracting all factors with positive roots from the SI tests intercorrelation matrix by the iterated principal factor method. This yielded 13 factors which were then rotated to a varimax criterion to achieve maximum uncorrelated simple structure. In order to study the relationships between the problem-solving criteria and the SI tests, this factor structure was then extended to include the problem-solving criteria by the method of Dwyer (1937) and Mosier (1938). The procedure is specified in the following formula:

$$V = R'_{ps-si} \, F \, (F'F)^{-1}$$

where V is the matrix of extension loadings of the problem-solving criteria on the SI factor matrix, R'_{ps-si} is the intercorrelation matrix of the problem-solving criteria with the SI tests. F is the varimax rotated factor loading matrix of the SI tests. These extension loadings were interpreted as estimated factor loadings for the purposes of elucidating the inter-battery relationships.

Summary

This concludes the chapter describing the methodology of the project. The development of the problem-solving criteria has been described and the study of their reliability has been delineated. The instruments employed in the primary study have been described, and the administration procedures have been outlined. Finally, the analyses of the data from the main study of problem solving were described in detail. In the following chapter the results of the analyses will be presented.

RESULTS

In this chapter the results of the two studies will be reported. First, the results of the reliability study will be presented. Then, the results of the primary study of problem solving will be reported.

Reliability Study

The reliability study attempted to determine the reliability of the problem-solving criteria developed in the first part of the study, by two different methods. The first method sought to establish the reliability of the instruments by administering them to the same group of subjects twice, with an interval of two weeks between administrations. The means and standard deviations for each of the measures derived from the instruments for the first and second testing sessions are given in Table 5. In addition, the distribution of finish types and legal finishes for The New Bike are reported in Table 6.

For The New Bike only 17 of 42 Ss indicated that they solved the problem of the first occasion and 5 of these were not justified, while in the second session 26 Ss successfully solved the problem. Since the inconsistency score was extremely positively skewed, a $\log_{10}$ transformation was applied, and these are the means reported in the table.

Table 5. Two week test-retest means and standard deviations for the reliability study of the problem-solving criteria.

Measure	Session 1 Mean	Session 1 S.D.	Session 2 Mean	Session 2 S.D.
The New Bike N=42				
Money earned	24.57	9.78	26.74	9.20
Days used	9.43	5.12	11.60	4.81
Information score	55.93	15.93	56.19	17.01
Strategy score	35.67	14.83	35.76	14.15
Inconsistency score	.43	.90	.47	.93
Number of steps	28.05	9.28	28.76	9.48
Concept LURB N=48				
Number of errors	4.58	2.91	4.92	3.28
Focus score	2.98	2.20	3.85	2.32
Scanning score	1.09	1.21	1.46	1.23
Concept DROOG N=48				
Number of errors	4.90	2.15	5.46	2.88
Verbal Maze #1 N=46				
Total time	9.19	4.31	5.67	3.23
Total trials	6.20	2.25	5.11	1.92
Total solutions	3.39	2.06	3.39	1.96

Table 6. Distribution of finish types and legal finishes for simulated problem situations.

Form	Finish Types			Legal Finish	
	Solved Problem	Quit	Did not finish	Yes	No
The New Bike (1)	17	8	17	37	5
The New Bike (2)	33	6	3	35	7
Free Ice Cream	37		7	17	27
The State Fair	13	7	3		
The Missing Friend	13	3	7		

Two particular pieces of information should be taken into consideration regarding these means in Table 5. First, test administrators consistently reported that few Ss understood the nature of the disjunctive concept DROOG which they were supposed to identify. Second, it should be noted that the Verbal Maze #1 was originally designed to have a maximum total of three solutions. Yet Ss discovered new solutions that met the criteria stated in the instructions, thus a mean total solutions of 3.39.

The test-retest correlations for these instruments are presented in Table 7. For The New Bike, the highest test-retest reliabilities were demonstrated by the information and strategy scores (r_{tt} = .50) of The New Bike. The other measures had reliabilities ranging from .19 (inconsistency score) to .48 for the number of steps measure. The measures for concept LURB demonstrated a wide range of test-retest reliabilities. They ranged from a high of .52 for number of errors to -.07, which was a not significantly different from zero, for the scanning score. The test-retest correlation of number of errors for concept DROOG was also non-significant. However, the reliabilities for Verbal Maze #1 were all significnat and ranged from .36 for total time to a respectable .74 for total solutions.

Table 7. Two week test-retest reliabilities - Reliability study.

Measure	N	r_{tt}
<u>The New Bike</u>	42	
Money earned		.38
Days used		.32
Information score		.50
Strategy score		.50
Inconsistency score		.19
Number of steps		.48
<u>Concept LURB</u>	48	
Number of errors		.52
Focus score		.34
Scanning score		.07 (n.s.)[1]
<u>Concept DROOG</u>	48	
Number of errors		.07 (n.s.)
<u>Verbal Maze #1</u>	46	
Total time		.35
Total trials		.52
Total solutions		.74

[1]Correlation not significantly different from 0.0 at the 0.05 level.

The second approach taken to assessing the reliability of the problem-solving criteria was to compare the performance of the same Ss on two supposedly alternate forms of the instrument. The alternate forms for the simulated problem situations were The New Bike and Free Ice Cream multiple solution problems and The State Fair and The Missing Friend single solution problems. The alternate forms for the concept problems were the LURBs and FLIXes conjunctive concepts and the DROOGs and ALGs disjunctive concepts. Finally, the alternate forms of the verbal mazes were Verbal Maze #1 and Verbal Maze #2. The means and standard deviations for these instruments are presented in Table 8. In addition, the finish types and legal finishes where appropriate for the four forms of the simulated problems, are given again in Table 8. From this table, it is evident that both The New Bike and Free Ice Cream were quite difficult since only 12 Ss successfully solved the first and only 10 solved the second. The State Fair and The Missing Friend appeared to be somewhat easier since 18 out of 23 solved each of these.

The alternate form means of the simulated problem situations cannot be directly compared since the forms have different numbers of sections and are only structurally similar. The concept problems, on the other hand, can be compared since the concept pairs are of equal

Table 8. Means and standard deviations for the alternate forms of the problem-solving criteria-reliability study.

Measure	Form 1 Mean	Form 1 S.D.	N	Form 2 Mean	Form 2 S.D.
Simulated problems	The New Bike			Free Ice Cream	
Money/points earned	24.65	9.60	44	32.89	15.91
Days used	9.39	5.01		5.23	2.73
Information score	55.66	15.94		37.82	18.93
Strategy score	35.25	14.90		30.68	13.89
Inconsistency score	.41	.88		.92	1.07
Number of Steps	27.86	9.23		17.93	8.91
	The State Fair			The Missing Friend	
Strategy score	75.96	25.04	23	34.48	12.46
Inconsistency score	.82	1.05		2.35	.18
Number of steps	7.35	2.44		8.91	1.65
Concept Tasks	LURBs			FLIXes	
Number of errors	4.67	2.86	48	4.85	3.14
Focus score	3.06	2.16		4.02	2.79
Scanning score	1.14	1.21		1.84	1.63
	DROOGs			RILGs	
Number of errors	4.94	2.17	47	5.00	2.49
Verbal Maze problems	#1			#2	
Total time	9.15	4.36	45	5.72	3.30
Total trials	6.11	2.20		4.91	1.81
Total solutions	3.47	2.01		2.80	1.85

complexity. The mean number of errors for conjunctive LURs (4.67) and FLIXes (4.35) were highly similar. However, the focus and scanning score means for the FLIXes were slightly higher than those for the LURBs. For the Verbal Mazes #1 and #2, the means for all three measures were less for the #2 form than for the #1 form. It should be noted that for all alternate forms, the first mentioned form was administered immediately before the second mentioned form.

The alternate form reliability coefficients for each of the instruments is presented in Table 9. These results cast considerable doubt on the assertion that the different forms of the instruments were, in fact, alternate forms. Twelve out of the 18 reliability coefficients were not significantly different from zero. The remaining coefficients ranged from a high of .84 for Verbal Maze total solutions to a low of .34 for Verbal Maze total time.

The results of this reliability study were used to choose the criteria for the primary study of problem solving. It should be obvious from Tables 7 and 9 that the obtained reliabilities of both types were far from ideal. Normally, acceptable reliability coefficients lie in the range of .70 to .99, which excludes every measure except Verbal Maze total solutions.

Table 9. Alternate form reliabilities - Reliability study.

Measure	N	r_{tt}
The New Bike - Free Ice Cream	44	
Money/points earned		-.02 (n.s.)[1]
Days used		-.01 (n.s.)
Information score		-.19 (n.s.)
Strategy score		-.03 (n.s.)
Inconsistency score		-.08 (n.s.)
Number of steps		-.04 (n.s.)
Concepts LURB-FLIX	48	
Number of errors		.56
Focus score		.20 (n.s.)
Scanning score		-.23 (n.s.)
Concepts DROOG-RILG	47	
Number of Errors		.19 (n.s.)
Verbal Mazes #1 - #2	45	
Total time		.34
Total trials		.46
Total solutions		.82
The State Fair - The Missing Friend	23	
Strategy score		-.03 (n.s.)
Inconsistency score		.00 (n.s.)
Number of steps		.21 (n.s.)

[1]Correlation is not significantly different from 0.0 at the .05 level.

However, it was decided that for research purposes that all measures which exhibited at least one type of reliability coefficient different from zero would be used in the primary study. The rationale for this decision was based on the assertion made by McGuire and Babbott (1967) and Treffinger and Poggio (1972) that for many kinds of problem-solving tests, conventional measures of reliability were inappropriate. This can be due to number of causes, including differential learning across a number of occasions of solving the same problem, or differential transfer from one task to the next. The result of this decision was that Concept DROOG and The State Fair were almost entirely eliminated and the scanning score from concept LURB was also deleted.

The Problem Solving Abilities Study

In primary study, the data for analyses were derived from 31 measures gathered on a sample of 490 subjects. Only those Ss with complete data on all variables were selected for analyses. Table 10 reports the descriptive statistics for each of the variablesincluding the 17 SI tests and the four problem-solving instruments. This table gives the mean, standard deviation, range, skewness, and reliability for each of the variables. Due to extreme positive skewness (5.67) of the inconsistency score from The New Bike, the variable labelled inconsistency score in the table is actually the log base 10 of the inconsistency score.

Intra-judge reliability.

The first analysis that was carried out was to determine the reliability of the scorers on those tests in which scoring required substantial judgement. Each test was scored by only one individual. Table 11 presents the score-rescore correlation coefficients for each of these types of tests. All coefficients were within a range of excellent scorer reliability (i.e. greater than .90) except for Symbol Production (.79). However, this last was judged to be sufficient for research purposes.

Table 10. Descriptive statistics of the human abilities measures and problem-solving criteria N=320.

Measure	Abbreviation	Mean	S.D.	range	Skewness
SI Ability Test					
Utility Test - Fluency	(UTFLU)	3.68	4.12	0.00 - 20.00	1.39
Utility Test - Flexibility	(UTFLX)	15.32	7.52	0.00 - 48.00	.71
Controlled Associations	(CA)	9.19	5.01	0.00 - 29.00	.60
Expressional Fluency	(EF)	3.08	2.56	0.00 - 15.00	1.09
Remembered Relations	(RR)	15.03	3.24	2.00 - 20.00	-.74
Apparatus Test	(AT)	11.53	8.59	0.00 - 31.00	.20
Unlikely Things	(UT)	20.17	3.74	7.00 - 28.00	-.88
Sentence Order	(SO)	5.62	2.44	0.00 - 14.00	.26
Sequential Associations	(SA)	2.94	1.67	0.00 - 8.00	.53
Judging Object Adaptations	(JOA)	11.15	3.28	1.00 - 19.00	-.97
New Uses	(NU)	10.45	4.64	0.00 - 28.00	.19
Word Linkage	(WL)	13.11	3.59	0.00 - 21.00	-.38
Word Completion	(WC)	9.14	4.38	0.00 - 19.00	-.15
Plane Flight Test	(PFT)	9.99	3.42	0.00 - 20.00	-1.09
Learned Information	(LI)	8.42	4.23	0.00 - 15.00	-.18

Table 10. (Continued)

Measure	Abbreviation	Mean	S.D.	range	Skewness
Logical Reasoning	(LR)	7.32	3.35	0.00 - 17.00	.47
Symbol Production	(SP)	14.46	5.90	0.00 - 31.00	.06
Related Alternatives	(RA)	11.56	4.40	0.00 - 20.00	-.13
Problem-solving Criteria					
The New Bike					
Dollars earned	(UENE)	21.53	11.57	0.00 - 45.00	-.40
Days used	(DUNB)	8.22	5.39	0.00 - 21.00	.30
Information score	(INFNB)	46.19	20.83	3.00 - 94.00	-.17
Strategy score	(STRNB)	30.43	17.48	0.00 - 97.00	.59
Inconsistency score	(INCONNB)	.66	1.03	0.00 - 3.00	.95
Number of steps	(NSTPNB)	25.95	13.05	3.00 - 84.00	.80
The State Fair					
Number of steps	(NSTPSF)	7.71	2.92	3.00 - 22.00	1.77
Verbal Maze #1					
Total time	(TMSPY)	4.74	3.66	0.00 - 15.00	.48
Total trials	(TTRSPY)	4.13	1.87	1.00 - 14.00	1.61
Total solutions	(TSSPY)	1.72	1.56	0.00 - 5.00	.30

Table 10. (Continued)

Measure	Abbreviation	Mean	S.D.	range	Skewness
<u>Concept LURB</u>					
Number of errors	(ERRCJ)	6.82	3.95	0.00 - 16.00	.33
Focus score	(FOCCJ)	3.39	2.74	0.00 - 14.00	.45
Total elapsed time	(TMCJ)	4.15	2.75	0.10 - 9.00	-.09

Table 11. Intra-judge reliability determinations for SI tests requiring substantial scorer judgement.

Measure	Score $\bar{X}_1$	Score S.D.$_1$	Rescore $\bar{X}_2$	Rescore S.D.$_2$	t_{12}	r_{12}
UTFLU	14.35	6.51	14.85	7.62	n.s.	.94
UTFLX	4.10	4.28	4.2	4.6	n.s.	.97
CA	11.00	6.23	11.00	6.32	n.s.	.99
EF	2.40	1.98	2.35	1.68	n.s.	.93
AT	11.00	7.71	10.83	8.20	n.s.	.98
SO	5.30	2.69	5.30	2.69	n.s.	1.00
SA	2.95	1.40	2.95	1.40	n.s.	1.00
NU	8.90	4.99	8.95	4.80	n.s.	.99
WC	8.90	4.89	8.75	4.82	n.s.	.99
LI	9.20	4.00	9.70	3.96	n.s.	.95
SP	14.05	5.34	13.45	4.47	n.s.	.79

Variable Intercorrelations.

The Pearson Product Moment intercorrelations for all the possible pairs of variables are reported in Table 12. For this table, any correlation greater than .11 is significantly different from zero at the 5% level. For purposes of describing the results the table will be subdivided into three sections; the SI test intercorrelations, the problem-solving criteria intercorrelations, and the SI test-problem-solving criteria intercorrelations.

The mean intercorrelation of the SI tests was .33 with a range from .10 to .61. Thus all the SI tests were positively correlated to a moderate degree. This positive manifold increased the possibility of underlying common factors.

For the problem-solving criteria, the mean magnitude of the intercorrelations was .19 with a range of correlations from -.28 to .94. 44 of the 78 correlations did not exceed the value (.10) necessary to be significantly different from zero. The largest intercorrelations appeared among measures from the same instrument as would be expected. The negative correlations are also to be expected, since some measures such as ERRCJ are measures of errors on a test while others are measures of positive performance. However, a

Table 12. Intercorrelations of SI Ability tests and problem-solving criteria.

	UTFLX	CA	EF	RR	AT	UT	SO	SA	JOA	NU	WL	WC	PFT
UTFLU	.52	.48	.34	.26	.34	.25	.30	.17	.20	.36	.29	.38	.17
UTFLX		.38	.27	.19	.33	.17	.21	.10	.11	.34	.20	.34	.12
CA			.40	.43	.41	.32	.36	.24	.34	.41	.33	.58	.29
EF				.27	.36	.20	.28	.21	.20	.32	.34	.42	.19
RR					.38	.37	.35	.27	.28	.34	.44	.47	.39
AT						.32	.24	.21	.26	.50	.37	.53	.30
UT							.30	.21	.33	.38	.33	.41	.35
SO								.26	.20	.41	.40	.40	.31
SA									.29	.22	.27	.36	.24
JOA										.28	.26	.38	.28
NU											.41	.61	.29
WL												.54	.36
WC													.40
PFT													
LI													
LR													
SP													
RA													
VENB													
DUNB													
INFNB													
STRNB													
INCONB													
NSTPNB													
NSTPSF													
TMSPY													
TTRSPY													
TSSPY													
ERRCJ													
FOCCJ													

Table 12. (Continued)

	LI	LR	SP	RA	VENB	DUNB	INFNB	STRNB	INCONB	NSTPNB
UTFLU	.22	.32	.26	.28	.05	.04	.15	.08	.04	.09
UTFLX	.26	.25	.28	.31	.12	.07	.14	.08	.05	.09
CA	.31	.41	.39	.40	.15	.15	.19	.10	-.03	.07
EF	.23	.23	.29	.31	.08	.06	.11	.06	-.01	.05
RR	.34	.40	.37	.46	.20	.18	.19	.12	-.10	.11
AT	.28	.33	.42	.39	.19	.18	.19	.10	-.03	.08
UT	.17	.32	.37	.32	.19	.20	.21	.12	-.11	.09
SO	.20	.38	.26	.35	.24	.19	.28	.18	-.03	.15
SA	.14	.26	.22	.28	.16	.17	.18	.15	-.12	.12
JOA	.20	.22	.26	.22	.11	.10	.09	.03	-.06	.02
NU	.35	.38	.44	.41	.32	.29	.38	.25	.01	.24
WL	.29	.40	.31	.47	.24	.20	.26	.13	-.09	.12
WC	.37	.49	.45	.54	.25	.21	.26	.14	-.05	.12
PFT	.15	.30	.26	.33	.09	.07	.11	.01	-.12	.01
LI		.21	.26	.30	.31	.25	.28	.22	-.02	.23
LR			.42	.35	.25	.22	.28	.16	.00	.16
SP				.32	.21	.19	.22	.16	.07	.15
RA					.28	.24	.31	.22	-.08	.20
VENB						.86	.87	.75	-.01	.76
DUNB							.81	.71	.05	.75
INFNB								.88	.09	.89
STRNB									.07	.94
INCONB										.09
NSTPNB										
NSTPSF										
TMSPY										
TTRSPY										
TSSPY										
ERRCJ										
FOCCJ										

Table 12. (Continued)

	NSTPSF	TMSPY	TTRSPY	TSSPY	ERRCJ	FOCCJ	TMCJ
UTFLU	-.07	-.08	.04	.25	-.23	-.11	-.08
UTFLX	-.06	-.04	-.01	.17	-.13	-.08	-.07
CA	-.15	-.06	.10	.31	-2.3	-.08	.00
EF	-.10	.00	.01	.16	-.09	.03	-.03
RR	-.07	-.05	-.05	.35	-.12	-.05	.02
AT	-.04	-.12	-.04	.28	-.12	-.06	-.02
UT	.06	.00	.02	.33	-.24	-.13	-.01
SO	-.09	-.04	-.08	.28	-.19	-.09	-.01
SA	-.13	-.01	.00	.18	-.18	-.03	.01
JOA	-.17	-.03	.01	.26	-.15	-.05	.11
NU	-.10	-.08	.01	.29	-.23	-.07	-.03
WL	-.06	-.02	.00	.34	-.12	-.03	.11
WC	-.08	-.01	.03	.33	-.23	-.13	-.02
PFT	-.14	.02	-.01	.32	-.20	-.05	.11
LI	-.15	-.09	-.01	.24	-.16	-.07	.01
LR	-.12	-.10	-.06	.33	-.25	-.13	-.06
SP	-.04	-.10	-.02	.28	-.13	-.03	.05
RA	-.09	-.11	-.02	.29	-.09	-.02	-.07
VENB	-.02	.01	-.01	.26	-.17	-.13	.03
DUNB	-.03	.00	-.03	.23	-.19	-.14	.00
INFNB	-.01	-.07	.01	.27	-.17	-.14	-.01
STRNB	.03	-.13	.02	.18	-.10	-.17	.01
INCONB	.22	-.07	.00	-.13	.03	-.05	-.05
NSTPNB	.04	-.08	.04	.18	-.11	-.14	.02
NSTPSF		-.04	-.03	-.10	.22	-.01	.00
TMSPY			.17	.03	-.13	.01	.05
TTRSPY				-.05	.03	.13	.01
TSSPY					-.28	-.09	.04
ERRCJ						.61	.06
FOCCJ							.13

result difficult to explain, is the sizable correlation of the two concept measures ERRCJ and FOCCJ. One of these is the number of errors an S makes in the process of identify the concept, while the other is an indicator of the S's use of a focusing strategy. Yet, this result appears more reasonable in the light of the assumption made using the number of errors. It was that the number of errors was assumed to be an index of the number of exemplars used by S to identify the concept. Since the focus score is calculated only on those exemplars the S uses before he solves the concept, a positive correlation between the measure of error and focus strategy is possible.

The mean magnitude of the intercorrelations of the problem-solving criteria with the SI tests was .12 with a range from -.25 to .30. Of the 234 correlation coefficients between these two batteries, 111 did not exceed the magnitude (.11) necessary for a significant difference from zero. Four measures - money earned, days used, and the information score from The New Bike, and the Verbal Maze total solutions - demonstrated a consistent pattern of low to moderate correlations with the SI tests. The New Bike inconsistency score, the total time and total trials from the Verbal Maze problem, and the total elapsed time for the concept problem presented consistently zero correlations with the SI tests. Thus it would appear that the possibility of common underlying factors is not high. In order to delineate the relationships among the variables, a number of analyses were carried out on the correlation matrix. These analyses will be reported in the following sections.

Regression analyses of the SI tests on the problem-solving criteria

In order to explore the predictability of the problem-solving criteria given the SI tests, step-wise regression analyses were carried out for each one of the problem-solving criteria, using the SI tests as predictors. Multiple correlations and percents of variance accounted for were calculated for each criterion. These figures were based on an optimum predictor set for each criterion which was determined by F tests on the change in multiple correlation. Table 13 summarizes the results of the regression analyses. Two particular points should be noted. First, the F test should be construed as a decision function rather than an exact test. Second, the multiple regression procedure takes advantage of all sources of variance including error variance. Thus multiple R's may be spuriously inflated.

For the problem-solving instrument, The New Bike, the multiple correlations (R) ranged from a high of .43 for the amount of money earned to .25 for the incon-

Table 13. Prediction of problem-solving criteria from SI ability measures.

Criterion	R	% predictable variance	SI predictors contributing significant variance
The New Bike			
Money earned	.43	19	NU, LI, RA, UTFLU, LR
Days used	.37	13	NU, LI, LR, UTFLU
Information score	.45	20	NU, RA, LI, LR
Strategy score	.30	9	NU, LI, RA
Inconsistency score	.25	6	UT
Number of steps	.32	10	NU, LI, RA, WC
The State Fair			
Number of steps	.21	4	JOA, LI
Verbal Maze #1			
Total time	.18	3	AT, RA, WC
Total trials	.16	2	CA, SO
Total solutions	.47	23	RR, UT, LR, PFT, LI
Concept LURB			
Number of errors	.33	11	LR, UT, UTFLU
Focus score	.13	2	WC
Total elapsed time	.23	5	PFT, RA, WL, LR

sistency score. The percent of variance accounted for in the criterion variable by the predictor set ranged from 19% for money earned to 6% for the inconsistency score. For every variable except inconsistency score, the SI test New Uses was the best predictor of performance. The SI tests Learned Information and Related Alternatives appeared in secondary or tertiary positions of importance in four out of the six optimum predictor sets. The fluency measure from the Utility Test contributed significant variance to the prediction of money earned and days used. Logical Reasoning was a significant predictor of money earned, days used, and the information score. Finally, Word Completion was a significant but minor predictor of number of steps and Unusual Things was the only significant SI predictor of the inconsistency score.

The number of steps measure for The State Fair had an R of .21 with its optimum predictor set, which accounted for only 4% of the variance. The optimum set consisted of Judging Object Adaptations and Learned Information.

The total time and total trials measures from Verbal Maze #1 had multiple R's of only .18 and .16 respectively accounting for only 3% and 2% of the variances. However the multiple R for the total solutions measure was a respectable .47 which accounted for 23% of the criterion variance. The significant predictor set for this variable consisted of five SI tests including Remembered Relations, Unlikely Things, Logical Reasoning, Plane Flight Test, and Learned Information.

For the concept identification task LURB, multiple R's ranged from a high of .33 and 11% of the variance to a low of .13 and 2% of the variance. Number of errors had the highest R, and was predicted by three tests with Logical Reasoning being the most important. Since so little variance was predictable for the other two measures, the predictor sets were judged to be of minor importance.

To summarize the regression analyses, no multiple R exceeded .50 and thus less than 25% of the criterion variance was predictable from the SI tests. However, it is important to recognize that this may have been due to more than one cause, since the multiple R between a criterion and a predictor set is also dependent on the reliabilities of all measures, and the problem-solving criterion reliabilities were generally low.

Structural analyses

As was stated in the Methodology chapter, a three-fold approach was taken to the question of underlying structure. The variable intercorrelations were analyzed by conventional factor analysis, canonical correlation analysis, and extension of the problem-solving criteria into a predetermined SI factor structure.

Conventional factor analysis. The first structural analysis carried out on the data was an iterated principal factor analysis extracting 31 principal factors. The eigenvalues and cumulative percent variance accounted for by each of the factors are given in Table 14. The scree test (Cattell, 1966) was carried out by plotting the eigenvalue against factor number for each of the factors extracted. This plot appears in Figure 11. This test proposes that eigenvalues of factors with a high proportion of error variance will form a straight sloping line or "scree slope." From the plot it was judged that the "scree slope" began at the sixth factor so the first five factors were judged to be relatively error free. The matrix of correlations was then subjected once again to an iterated principal factor analysis which extracted five factors. These factors were then rotated to a varimax criterion. The factor structure and communalities are reported in Table 15. This solution was judged not completely adequate in terms of simple structure since a number of tests and measures such as the Utility Test, Controlled Associates,

Table 14. Eigenvalues and cumulative percent of variance for the 31 factor principal factor solution.

Factor	Eigenvalue	Cumulative % variance	Factor	Eigenvalue	Cumulative % of variance
1	7.52	39.0	17	.19	96.4
2	3.51	57.1	18	.18	97.3
3	1.42	64.5	19	.17	98.2
4	1.13	70.3	20	.13	98.9
5	.74	74.2	21	.11	99.4
6	.57	77.2	22	.08	99.9
7	.51	79.8	23	.06	100.2
8	.48	82.3	24	.05	100.4
9	.43	84.5	25	.03	100.6
10	.31	86.5	26	.02	100.7
11	.36	88.4	27	.01	100.7
12	.36	90.2	28	-.01	100.7
13	.30	91.8	29	-.03	100.5
14	.26	93.1	30	-.05	100.3
15	.23	94.3	31	-.06	100.0
16	.22	95.4			

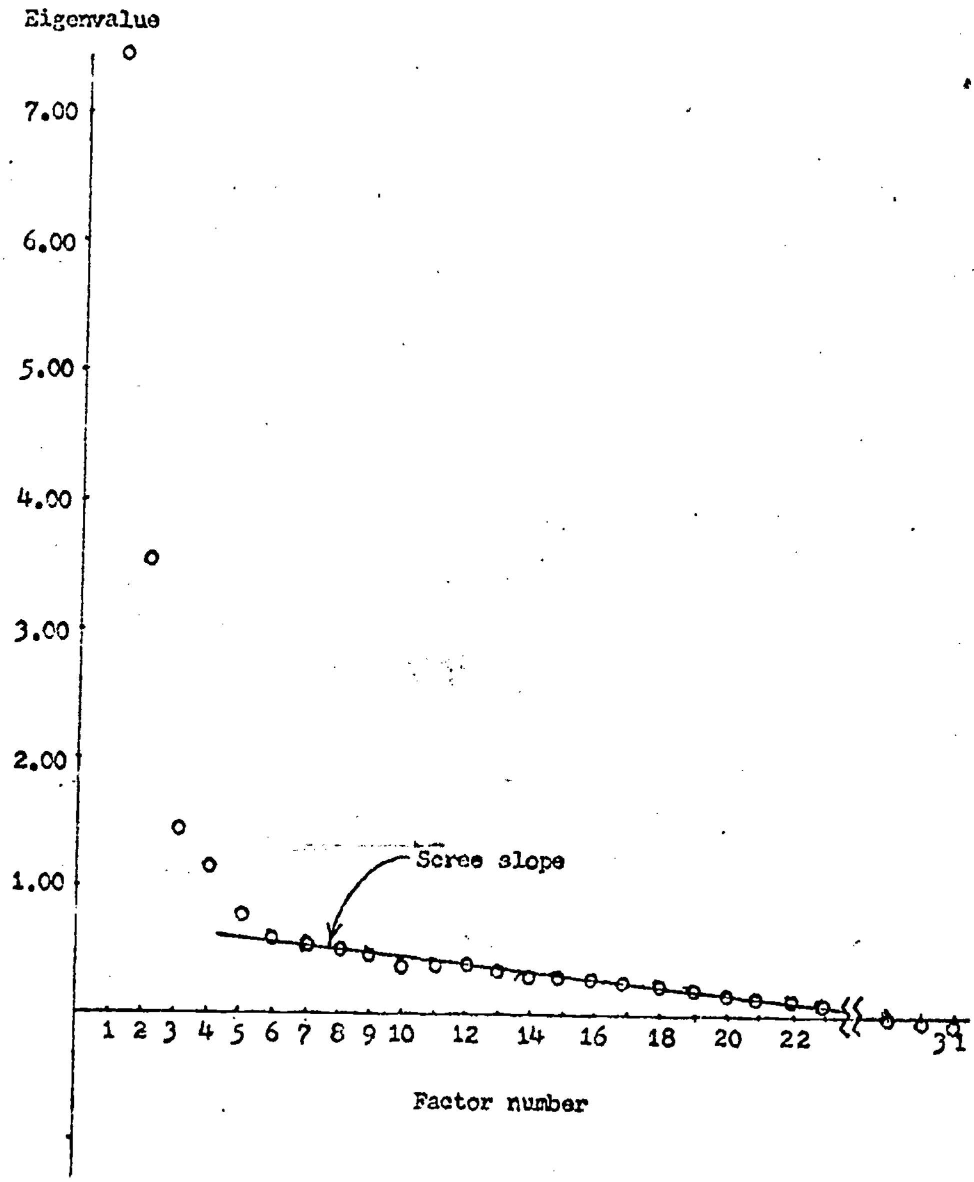


Figure 11. Plot of eigenvalues vs. factor number for 31 principal factors.

Expressional Fluency, and several measures from The New Bike had important loadings on more than one factor.

This varimax solution was then subjected to an oblique rotation by the Promax procedure (Hendrickson and White, 1966). Table 16 presents the primary factor loading matrix. Table 17 gives the factor intercorrelations and Table 18 reports the test-factor correlation matrix.

Comparing Tables 15 and 16 it is evident that the Promax solution yielded a slightly simpler structure in terms of factor loadings. The loadings for the two measures from the Utility Test increased slightly in factor IV and decreased in factor I. The major improvement was in the hyperplane count (loadings of magnitude less than .10) which improved for each factor. In addition, most tests had an important loading on only one factor. However, the factors yielded by Promax procedure were no longer orthogonal. Table 17 indicates that factor I is most correlated with factors II, IV and V, yet the magnitude is quite moderate (.34). Factor II is slightly negatively correlated with factor III and essentially uncorrelated with factors IV and V. Factors III, IV and V are essentially uncorrelated. It would appear then that the oblique rotation left most of the factors orthogonal while moving factor I to a

Table 15. Five factor varimax solution for the SI tests and problem-solving criteria.

	Factor I	Factor II	Factor III	Factor IV	Factor V	
Eigenvalue	7.52	3.51	1.42	1.13	.74	
Cum. % of variance	39.0	57.1	64.5	70.3	74.2	
Measure						Communalities
UTFLU	.38	.00	-.10	.60	.01	.52
UTFLX	.30	.04	-.14	.57	-.06	.42
CA	.59	.03	-.03	.40	.12	.53
EF	.43	.00	.08	.31	.05	.29
RR	.66	.06	.01	-.01	-.01	.44
AT	.59	.05	.01	.23	-.14	.42
UT	.56	.08	-.12	-.02	.06	.34
SO	.52	.13	-.08	.10	.00	.30
SA	.41	.10	-.03	.00	.15	.20
JOA	.46	.00	-.03	.01	.17	.25
NU	.60	.21	-.05	.26	-.06	.48
WL	.66	.10	.05	.02	.04	.45
WC	.77	.07	-.05	.23	.01	.66
PFT	.56	-.03	-.06	-.11	.15	.35
LI	.38	.21	-.01	.19	.03	.28
LR	.57	.13	-.14	.12	-.07	.38
SP	.55	.10	.00	.15	-.13	.35
RA	.61	.17	-.06	.13	-.06	.42
UENB	.24	.86	-.04	-.04	.03	.80
DUNB	.20	.82	-.08	-.05	.01	.73
INFNB	.23	.94	-.04	.07	-.05	.94
STRNB	.08	.90	-.04	.05	-.10	.84
INCONB	-.13	.08	-.03	.16	-.30	.14
NSTPNB	.05	.94	-.02	.08	-.06	.89
NSTPSP	-.15	.02	.07	-.04	-.35	.15
TMSPY	-.08	-.03	-.05	-.05	.34	.13
TTRSPY	-.06	.03	.13	.13	.25	.10
TSSPY	.48	.16	-.14	.00	.14	.30

Table 15. (continued)

	I	II	III	IV	V	Communalities
ERRCT	-.22	-.10	.80	-.11	.28	.79
FOCCJ	-.05	-.10	.76	-.04	.12	.61
TMCJ	.05	.02	.13	-.16	.13	.06

Table 16. Primary factor matrix for the five factor Promax solution.

Measure	I	II	III	IV	V
			Factor		
UTFLU	.23	-.05	-.09	.59	.06
UTFLX	.18	.01	-.02	.56	-.02
CA	.49	-.04	.00	.37	.14
EF	.39	-.05	.10	.28	.05
RR	.74	-.05	.05	-.11	-.06
AT	.64	-.05	.06	.14	-.16
UT	.57	-.02	-.10	-.09	.04
SO	.51	.05	-.05	.03	-.01
SA	.37	.05	-.02	-.04	.14
JOA	.44	-.06	-.02	-.03	.16
NU	.57	.12	.00	.19	-.06
WL	.70	.01	.09	-.07	.00
WC	.77	-.05	.00	.14	-.01
PFT	.60	-.13	-.04	-.18	.17
LI	.32	.17	.01	.15	.04
LR	.58	.02	-.11	.03	-.08
SP	.59	.01	.04	.07	-.16
RA	.64	.08	.11	.04	-.09
VENB	.09	.87	.00	-.09	.06
DUNB	.06	.83	-.04	-.10	.04
INFNB	.07	.94	.01	.01	-.02
STRNB	-.07	.93	.00	.01	-.06
INCONB	-.09	.08	-.02	.14	.30
NSTPNB	-.14	.98	.02	.05	.02
NSTPSF	-.01	.02	.08	-.08	-.39
TMSPY	-.21	.00	-.07	.01	.38
TTRSPY	-.19	.09	.13	.19	.28
TSSPY	.43	.09	-.12	-.05	.14
ERRCJ	.04	.00	.83	-.11	-.38
FOOCJ	.03	.00	.78	.01	.06
TMCJ	.06	.03	.14	-.15	.11

slightly oblique position with respect to all of the other factors.

In order to understand the psychological meaning of the five factors, it is necessary to examine the test-factor correlation matrix. Table 18, which presents this matrix, reveals that while the primary factor structure became simpler through the oblique rotation, a greater degree of test overlap on factors is present. Factor I appears to be a general test performance factor. All measures with reliabilities greater than .30 have important correlations with this factor. Factor II appears to be primarily a test specific factor having high correlations with The New Bike measures, though the New Uses is also moderately correlated. This is probably due to correlated errors. Since these same tests also have important correlations with factor I, the correlation of the two factors is explained.

Factor III is another test specific factor which had high correlations with error and focus scores from the concept identification problem. The low negative correlations of this factor with factors I and II are most probably a result of the low negative correlations of a number of SI tests and The New Bike with this factor. This in turn is most likely due to the fact that the two measures assess degree of poor performance on the task.

Factor IV appears to represent a divergent production factor. Four of the seven tests which had important correlations with this factor have been identified by Guilford (1967) as divergent production tests. Two of the remaining tests, Apparatus Test and New Uses, call for "original" thinking by S in order to successfully perform on the tests. "Original" thinking has often been identified as divergent in nature. Finally, Word Completion has a moderate correlation with this factor, probably due to the semantic nature of all the other tests. Factor IV has its highest correlation with factor I, most probably because a general test performance factor affects performance on these divergent production tests.

Factor V has moderate correlations with several tests in both the SI tests and problem-solving criteria. This factor appears to indicate that the SI abilities of evaluation of semantic transformations (Judging Object Adaptations) and cognition of semantic systems (Plane Flight Test) are related to problem-solving measures of the errors made in problem solving (The New Bike inconsistency score, number of steps from The State Fair, and number of errors in the concept problem). The measure of total time in the Verbal Maze problem also had an important correlation with this factor. Factor V was moderately correlated with only factor I,

Table 17. Factor intercorrelations. Five factor Promax solution for the SI tests and problem-solving criteria.

	Factor			
Factor	II	III	IV	V
I	.34	-.21	.34	.31
II		-.21	.10	-.03
III			-.11	.08
IV				-.11

Table 18. Factor-test correlation matrix. Five factor [illegible] solution for the SI test and problem-solving criteria.

Measure	Factor I	II	III	IV	V
UTFLU	[illegible]	[illegible]	-.[illegible]	.67	.06
UTFLX	[illegible]	.13	-.1[illegible]	.53	-.02
CA	.5[illegible]	[illegible]	-.[illegible]	.52	.25
[illegible]	[illegible]	.0[illegible]	.0[illegible]	[illegible]	.15
[illegible]	[illegible]	.13	-.0[illegible]	[illegible]	.19
[illegible]	.2[illegible]	.1[illegible]	-.10	.37	.03
[illegible]	[illegible]	.13	-.23	.11	.22
[illegible]	[illegible]	.2[illegible]	-.17	.22	.14
[illegible]	[illegible]	.17	-.09	.08	.26
[illegible]	[illegible]	.06	-.08	.10	.30
[illegible]	[illegible]	.33	-.1[illegible]	[illegible]	.09
VL	[illegible]	.22	-.05	.17	.23
[illegible]	[illegible]	.23	-.17	.40	.22
[illegible]	[illegible]	.06	-.11	.00	.32
[illegible]	[illegible]	.29	-.10	.27	.12
[illegible]	.5[illegible]	.2[illegible]	-.2[illegible]	.25	.09
SP	.5[illegible]	.22	-.11	.28	.02
RA	[illegible]	.2[illegible]	-.05	.27	.12
[illegible]	[illegible]	.5[illegible]	-.19	.02	.07
[illegible]	[illegible]	[illegible]	-.21	.01	.04
[illegible]	[illegible]	[illegible]	-.2[illegible]	.13	-.02
[illegible]	.2[illegible]	[illegible]	-.18	.08	-.11
[illegible]	-.[illegible]	.[illegible]	-.06	.15	-.35
[illegible]	.20	.93	-.1[illegible]	.1	-.09
[illegible]	-.17	.[illegible]	.05	-.04	-.37
[illegible]	-.0[illegible]	-.06	.0[illegible]	-.09	.[illegible]
[illegible]	-.0[illegible]	.01	.15	.09	.20
TSSPY	.51	.2[illegible]	-.21	.10	.27
ERROJ	-.[illegible]	-.1[illegible]	.8[illegible]	-.15	-.30
FOCCJ	-.[illegible]	-.13	.7[illegible]	-.07	.13
[illegible]	.03	.01	.[illegible]	-.15	.16

which can be accounted for by the important correlations of Judging Object Adaptations, Plane Flight Test, and number of errors in the concept problem with both factors.

To summarize the results of this conventional factor analysis, it revealed little about underlying relationships between human abilities as specified by the SI tests and the problem-solving criteria. Only factors I and V showed any crossover between the two sets of tests. However Factor I explained little since it appeared to be a general test performance factor. Only factor V presented some evidence that the human abilities, evaluation of semantic transformations and cognition of semantic systems, are related to errors in problem-solving tasks.

Canonical correlation analysis. Since the conventional factor analysis revealed little of the relationships between the set of human abilities tests and the problem-solving criteria, it was decided that a different, more suitable approach would be employed. This approach chosen was canonical correlation analysis which was designed by Hotelling (1936) to analyze relationships between sets of variables.

Five canonical variates were extracted from the set of SI tests and the problem-solving criteria. The pertinent information concerning these canonical variates is presented in Table 19. The first two

Table 19. Description of canonical variates: SI tests versus the problem-solving criteria.

Root	R	R^2	Significance Level	Variance extracted	Redundancy	Proportion of total redundancy
				SI Tests		
1	.66	.44	.000	.17	.074	.64
2	.40	.16	.112	.14	.022	.19
3	.37	.14	.486	.05	.008	.07
4	.34	.11	.853	.07	.008	.07
5	.29	.03	.981	.05	.004	.03
Sum				.60	.116	
			Problem-solving criteria			
1	.66	.44	.000	.31	.136	.87
2	.40	.16	.112	.04	.006	.04
3	.37	.14	.486	.05	.007	.04
4	.34	.11	.353	.04	.004	.03
5	.29	.08	.981	.04	.003	02
Sum				.48	.156	

canonical correlations were significant according to the Wilks lambda test. However, five canonical variates were extracted in order to account for a maximum amount of variance in the two sets. The five canonical variates extracted 60% of the variance of the SI tests and 48% of the variance of the problem-solving criteria. The first two variates were the only ones to extract appreciable variance from the first set, while only the first canonical variate extracted appreciable variance from the second set.

The redundancy figures reported in Table 19 were the amounts of variance extracted by the canonical variates from one set that were predictable from the corresponding canonical variates of the other set. Thus for the first pair of canonical variates which were correlated .66, the proportion of variance predictable in the second set from the first set is .074, while the proportion of variance in the first set predictable from the second set was .136. The total proportion of extracted variance predictable in the set of problem-solving criteria from the SI tests was .116. The reverse accounted for a total proportion of variance of .156. Finally, the first pair of canonical variates accounted for 64% of the redundancy in the SI tests and 87% of the redundancy in the second set. This indicates that the only canonical variates of importance were the first extracted.

pair of variates in the analysis.

In order to analyze the contributions of the various tests due to the first pair of canonical variates, the correlations between the tests and the canonical factors were calculated. This canonical factor structure is given in Table 20. Since the first canonical factor accounted for most of the redundancy, this was the only one which was closely examined. All the SI tests had moderate correlations with this factor. The test exhibiting the highest correlation was New Uses with a coefficient of -.75. For the problem-solving criteria, several of the measures also had reasonable size correlations with the factor. These measures included four of the measures from The New Bike, the Verbal Maze total solutions, and concept problem total errors. The pattern of correlations across the two batteries appear quite similar to the general test performance factor found in the conventional factor analysis.

Extension analysis. Since the canonical correlation analysis also revealed little about the relationships between the SI tests and the problem-solving criteria, a third approach was attempted. This analysis attempted to answer the question: assuming that the SI tests represent a structure of human abilities, how do the

Table 20. Test correlations with the 5 independent canonical factors.

Measure	Canonical factors I	II	III	IV	V
	SI Tests				
UTFLU	-.48	-.13	-.32	.45	-.41
UTFLX	-.32	.10	-.08	.13	-.31
CA	-.62	-.24	.01	.21	.19
EF	-.32	-.03	-.09	.13	.12
RR	-.53	-.01	.35	-.15	-.10
AT	-.53	.03	.23	-.17	-.11
UT	-.64	-.35	.02	-.28	.00
SO	-.63	.00	-.06	-.15	-.13
SA	-.42	-.08	-.13	-.09	.50
JOA	-.44	-.20	.40	-.15	.22
NJ	-.75	.12	-.21	-.05	.01
WL	-.62	-.11	.34	-.05	-.10
WC	-.69	-.22	.05	-.23	-.13
PFT	-.53	-.39	.27	.20	-.05
LI	-.48	.46	.25	.18	.08
LR	-.69	.13	.01	.08	-.24
SP	-.50	.13	.24	.08	.00
RA	-.62	.22	.00	-.27	-.04
	Problem-solving Criteria				
VENB	-.57	.53	.11	-.36	.13
DUNB	-.32	.42	.05	-.36	.31
INFNB	-.65	.48	-.19	-.21	.05
STRNB	-.40	.53	-.22	-.21	.23
INCONB	.19	.32	.00	.41	-.15
NSTPNB	-.38	.58	-.19	-.09	.10
NSTPSF	.27	-.05	-.05	-.37	-.42
TNSPY	.15	-.42	-.01	-.14	.05
TTRSPY	.02	-.28	-.17	.17	.25
TSSPY	-.71	-.14	.37	.08	-.17
ERRCJ	.49	.23	.24	-.30	.00
FOCCJ	.19	.12	.08	.05	.37
TMCJ	.02	-.22	.57	.24	.28

problem-solving criteria fit within this structure? In order to answer this question, it was first necessary to reproduce the structure specified by the Structure of Intellect model (Guilford, 1967) from the intercorrelations of the SI tests. The first attempt at reproducing this structure is reported in Table 21. This table gives the results of an iterated principal factor solution, a scree test on the eigenvalues, and a varimax rotation of the two factors judged to be relatively error free. It is evident that this solution did not reproduce the factor structure supposedly represented by the 16 SI tests.

A second approach was then carried out. In this analysis all factors were used which extracted positive variance from the intercorrelation matrix. This 13 factor varimax rotated solution is reported in Table 22. These factors accounted for 100% of the variance in the SI tests. It is evident that the 18 SI factors supposedly represented by the test battery were not fully represented. However, these factors do appear to represent 13 factors similar to those factors specified by the SI model. Nine of the 13 factors had a single test which loaded higher than any other test, though two of these loadings were so low in magnitude (factor XII and factor XIII) that the factors could only be tentatively identified as representing the particular

Table 21. Two factor varimax solution for SI tests.

	Factor I	Factor II
Eigenvalue	6.[illegible]	.98
Cum. % of variance	63.1	72.1
Measure		
UTPLU	.17	.67
UTPLX	.30	.70
CA	.45	.54
EP	.32	.40
ER	.52	.22
AT	.45	.44
UT	.54	.17
SO	.47	.27
SA	.42	.11
JCA	.44	.14
NI	.55	.46
WR	.61	.25
WC	.68	.46
PFT	.55	.08
LI	.31	.32
LR	.50	.31
SP	.46	.34
FA	.54	.33

Table 22. Thirteen factor varimax solutions for 18 SI tests.

	Factor I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
Eigenvalue	6.18	.88	.47	.41	.37	.32	.27	.23	.20	.18	.13	.10	.07
Cum. % of var.	63.1	72.1	76.8	81.0	84.8	88.1	90.8	93.2	95.2	97.0	98.3	99.4	100.1
Measure													
UTFLU	.69	.09	.03	.06	.08	.11	.27	.11	.14	.04	.10	.09	.09
UTFLX	.67	.15	.15	.15	.12	.01	.04	.01	.05	.02	-.01	-.05	-.07
CA	.40	.12	.11	.21	.23	.16	.17	.21	.26	.12	.41	.00	.04
EF	.25	.16	.13	.12	.09	.04	.14	.17	.46	.04	.07	.02	.00
RR	.09	.05	.31	.35	.26	.21	.20	.18	.11	.30	.11	-.01	(.22)
AT	.25	.49	.17	.15	.23	.13	-.02	.12	.26	.20	.03	.03	.09
UT	.11	.17	.16	.06	.20	.52	.17	.18	.04	.13	.03	.03	.02
SO	.17	.12	.14	.09	.13	.12	.57	.17	.11	.11	.05	.01	.01
SA	.06	.06	.25	.05	.12	.05	.15	.49	.09	.06	.00	.02	.03
JOA	.03	.12	-.02	.16	.06	.33	.02	.46	.08	.12	.13	.02	-.03
NU	.24	.52	.10	.24	.18	.21	.31	.12	.07	.02	.04	.01	-.04
WL	.11	.50	.37	.20	.14	.22	.31	.17	.21	.15	.03	(.26)	-.01
WC	.22	.46	.32	.21	.22	.14	.18	.35	.17	.12	.30	.16	-.06
PFT	.04	.13	.20	.04	.11	.25	.20	.23	.05	.42	.05	.04	-.01
LI	.18	.17	.13	.51	.09	.05	.07	.10	.09	.02	.04	.02	-.01
LR	.20	.13	.17	.67	.48	.09	.29	.19	.01	.12	.14	.15	.04
SP	.16	.27	.03	.16	.52	.24	.07	.15	.16	.05	.01	-.07	-.02
RA	.20	.20	.52	.22	.13	.12	.18	.18	.12	.14	.05	-.01	.00
SI Factor	DMU DMC	CMI NMT CMU	NMI	MMS	EMI DMP	EMS	NMS	NMI ENT	UMS	CMS	DMR	(CMR)	(EMR)

tests with the highest loadings. Four factors had essentially equivalent high loading for several tests and thus were identified as representing more than one SI factor. Undoubtably, the structure could have been improved by oblique rotation, but in conformance with Guilford's criterion of orthogonality, this structure was accepted as the most representative structure of the SI model that could be derived from available data.

Using the factor structure given in Table 22, extension loadings of the problem-solving criteria on the 13 factors were calculated according to the procedures of Mosier (1938). This had the effect of adding a row to the factor loading matrix for each of the problem-criteria. These added rows are presented in Table 23. Almost all of the extension loadings are less than .30, indicating that most of the problem-solving criteria were not related to the factor structure. There are, however, a few exceptions. The number of steps and money earned from The New Bike are moderately related to factor IV, which was associated with the SI factor of memory for semantic systems. The information score from the same test appeared to be related to factor VII, which was associated with convergent production of semantic systems. The New Bike strategy score and the concept problem total elapsed time were related to factor X,

Table 23. Extension loadings of the problem-solving criteria on the 13 factor varimax solution for the SI tests.

Measures	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
VAMB	-.02	.19	.19	[illegible]	.20	.06	.24	.10	-.14	-.19	.13	[illegible]	[illegible]
CUNR	-.05	.19	.20	.29	[illegible]	[illegible]	.18	.13	-.13	[illegible]	-.05	[illegible]	[illegible]
INTRL	.0[illegible]	.22	.21	[illegible]	.18	[illegible]	.34	.10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
STAMB	.02	.13	.19	[illegible]	.14	[illegible]	.24	.13	-.11	[illegible]	[illegible]	[illegible]	[illegible]
INFORB	.08	.34	-.17	.32	[illegible]	-.11	.00	[illegible]	[illegible]	[illegible]	.30	[illegible]	[illegible]
ASTPRB	.04	.11	[illegible]	.30	[illegible]	-.0[illegible]	.22	[illegible]	[illegible]	[illegible]	-.17	[illegible]	[illegible]
ASTPRJ	-.06	.06	[illegible]	-.21	.01	[illegible]	-.09	-.24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
INTPY	-.07	-.06	[illegible]	-.09	-.14	[illegible]	.01	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
VERBAY	.04	-.01	.02	[illegible]	-.10	.14	-.10	-.02	.04	[illegible]	[illegible]	[illegible]	[illegible]
ASSPY	.15	.06	.06	[illegible]	[illegible]	[illegible]	.13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
EMRCJ	-.10	-.09	.16	-.03	-.09	-.13	-.22	-.21	.15	[illegible]	[illegible]	[illegible]	.33
FOCCJ	-.13	-.07	.04	[illegible]	[illegible]	[illegible]	.06	-.04	.25	[illegible]	[illegible]	-.12	[illegible]
[illegible]CMJ	-.14	-.12	-.21	.17	.01	[illegible]	.04	.04	.13	[illegible]	[illegible]	[illegible]	[illegible]

which was associated with cognition of semantic systems. Finally, the total elapsed time for the concept problem was also related to factor XIII which was tentatively identified as memory for semantic relations.

To summarize the extension loading analysis, little of the relationships between the SI tests and the problem-solving criteria were revealed. Only seven loading exceeded .30, and these were widely scattered across the factors. The outcome once again appears to be that there was essentially no relationships between the SI tests and the problem-solving criteria.

Summary

This concludes the results chapter. The findings of the reliability study have been reported. The data concerning the SI tests and the problem-solving criteria have been described and subjected to four different types of analyses. The results of the regression analyses and the three multivariate correlational analyses were reported. In the next chapter these results will be discussed and conclusions will be drawn.

DISCUSSION AND CONCLUSIONS

In this chapter, the results of the two studies will be discussed in terms of the purposes of the project. Several alternative explanations for these results will be explored. Conclusions will then be drawn, and recommendations for further study will be made.

The Reliability Study

The purpose of this study was to evaluate the problem-solving criteria which had been designed and developed to more adequately sample problem-solving behavior. These instruments were evaluated in terms of their stability over time, or test-retest reliability, and in terms of content sampling or alternate forms reliability. It is evident from the results reported in the previous chapter that none of the problem-solving measures exhibited high levels of reliability of either type. Two-week test-retest reliabilities ranged from a low of .07 to a high of .7[illegible] with most falling in the range .[illegible] to .30 to .35. This indicates that, at best, the total solutions from the Verbal Maze #1 contained 26% error variance, while most measures contained more than 50% error variance.

The results for the evaluation of content sampling or alternate form reliability were even poorer. Only the number of errors for the conjunctive concept and the measures from the Verbal Maze problems were significantly different from zero. The reliability coefficients for these measures ranged from .34 to .82. Thus percent of error variance in the measures predicted by this method ranged from 100% to 18%.

The results of the two different methods of determining reliability indicate that only two of the measures exhibited consistent reliabilities. Total solutions from the Verbal Maze problems and the number of errors in the concept identifiction problem had reliabilities of about .55 and .76 respectively across the two measures. Yet, these results leave much to be desired since a considerable proportion of the score variance was still error variance.

Several possible explanations exist for these results in the reliability study. Low reliabilities may be due to task difficulty, differential learning, content sampling difficulties, or inadequate measure validity.

Test-retest and alternate form correlations may both be affected by task difficulty. If the tasks were too difficult for Ss, their performance on the measures would be essentially meaningless and consist mostly of random error. This condition did not seem to characterize the concept identification tasks and the Verbal Maze problem, since the reported means indicated that most Ss solved these problems. However, this may have been true of the simplified problems. Test administrators did report that Ss generally needed considerable help with the format and directions in order to solve the problem. In addition the results reported in the previous chapter indicated that less than half the Ss successfully solved the problems, and a number of illegal finishes were recorded. Test administrators also reported that the disjunctive concept problems were extremely difficult for the Ss. Thus it would appear that in spite of the efforts of the developers, these instruments were quite difficult for fifth grade students. Consequently, a contributing factor to the low reliabilities of these instruments may have been the extreme difficulty of the problems.

Another factor which could affect both types of reliabilities is differential learning occurring between successive administration of the tasks. If each S learns the same amount while solving the problem this should be reflected in a change in the means across tests, but a high correlation since each S changes by the same amount. However, if differential learning takes place, i.e. the amount of learning varies across Ss, then both the means and the correlation will be affected since both the performance and the relative ranking of Ss will change. Learning apparently did take place for practically all of the problem-solving measure in the test-retest condition since the reported means generally changed across testing sessions. For the alternate forms, learning was evident primarily for the Verbal Maze problems. If differential learning did occur, it could account forthe low test-retest correlations of most of the measures.

Another factor that could affect the alternate forms reliability is the content that each instrument sampled. The basic assumption behind alternate forms reliability measurement is that the two forms measure the same thing. Forms should, thus, only differ in superficial characteristics. In the original developmental effort, an attempt was made to construct parallel forms for each of the types of instruments. For the simulated problems essentially the same problem maps were used for both the single and multiple solution forms. Problem goals were essentially the same; to find something missing by searching a variety of places, and to collect a given amount in a minimum time in order to obtain something. In the concept tasks the forms were exactly the same. Only the target concept was changed in terms of the relevant dimensions. The Verbal Maze problems used the same problem maps for both forms. Only the names were changed from one form to the other. However, some possibly important differences did exist. *Free the Dream* contained no provisions for quitting, while *The New Bike* did. The saliency of the relevant concept dimensions may have differed between target concepts. If these differences were, in fact, important or the differences between forms which were designated as superficial, were not superficial, then the constructed parallel forms may have not been alternate forms. The large number of non-significant alternate forms correlations support the contention that the constructed parallel forms were not, in fact, alternate forms.

A final factor generated by the test characteristics is the relevance of the scoring procedures. That is, are the measures derived from the instruments measuring significant aspects of the S's problem-solving behavior, or are the scoring procedures yielding valid measures? Unfortunately, the only way to determine this for the problem-solving criteria, at present, is to evaluate their face validity. For the simulated problems, the money/points earned and days used appear to have face validity since better problem-solvers should earn more money/points in fewer days. The information score which reflects the amount of information used by the problem-solver appears to be valid because the better problem-solver should utilize more information in solving the problem. The number of steps also appears to be valid because the better problem-solver should take fewer steps to solve the problem. There may be some question about the strategy score and inconsistency score, however. The strategy score reflects the conformance of the S's pattern of choices with the sure strategy. But this sure strategy may not correspond to the S's conception of a best strategy, since the common assumption is that the best strategy involves the fewest steps. The sure strategy guarantees that the problem will be solved in the minimum of days, rather than in the fewest steps. Thus Ss may not be even attempting to follow a sure strategy which causes the conformance measure to lose meaning. The inconsistency score supposedly reflects the logical errors S make in following directions in the problem. However this measure may reflect errors from two sources: misunderstanding given directions and accidental random errors due to such sources as turning to the wrong page. This second source is not systematic, and therefore the inconsistency score may reflect random as well as systematic variance.

For the concept problems, the number of errors and focusing score are recognized measures of concept identification often appearing in the research literature. However, the scanning score may be less valid in this type of situation, though it is a recognized (Bourne, et al., 1972) concept identification strategy. It is a complex strategy, requiring the S to carry all possible concepts generated from the focus example in his head, and eliminate specific possibilities as he looks at each exemplar. It seems probable that Ss in a single concept identification task, when they are not familiar with the type of task, would not employ such a complex strategy. Thus the scanning score may, in most cases, only reflect accidental sequences of exemplars rather than a conscious strategy.

All the Verbal Maze measures appear to possess face validity. Total time is a measure of how efficient the problem-solver is. Total trials is a measure of the errors the S makes in solving the problem. And total solutions reflects the S's ability to find a number of different solutions.

Therefore, from the standpoint of validity, it would appear that the simulated problems and the concept identification tasks could be somewhat affected by the validity of derived measures. The Verbal Mazes, on the other hand, do not seem to be affected by this consideration. This lack of validity may help to explain some of the low reliabilities for the simulated problems and the concept identification tasks.

From the above discussion it is evident that a number of conditions existed which could adversely affect the reliability coefficients for the problem-solving criteria. Since the reliabilities were quite low, it is reasonable to assume that some of these factors were active in lowering the reliability of the different measures. Regardless of the causes of the decrements in reliability, the important outcome of this study was that the problem-solving criteria exhibited consistently low reliabilities. These low reliabilities will adversely affect all attempts to investigate the problem-solving behavior sampled by these measures, as will be seen in the following sections.

The Problem-Solving Abilities Study

The primary outcome of this study was that few relationships were evident between this particular set of problem-solving criteria and the set of tests representing a number of Structure of Intellect (Guilford, 1967) factors. However, the relationships that were evident will be discussed in the following sections.

Regression Analyses.

The regression analyses were the most fruitful in terms of relationships revealed. These analyses were designed to determine the predictability of the problem-solving criteria from the SI test. Each of the problem-solving criteria was significant correlated with at least one of the SI tests. Multiple correlations were not great, the greatest being .47, and thus large amounts of variance in the criterion variables were left unexplained. Optimum predictor batteries ranged from one to five SI tests.

Three SI tests were significantly related to all of The New Bike measures except the inconsistency score. These were, in order of importance, New Uses, Learned Information, and Related Alternatives. New Uses represented the SI ability of convergent production of semantic transformations. Learned Information represented the SI factor of memory for semantic systems, and Related

Alternatives represented the SI ability of memory for semantic implications. New Uses appears to be logically related to performance on this instrument in that the decision-making involves changing and integrating the given information in order to arrive at the best decision, convergent production of semantic transformations. The logical relationship of Learned Information and Related Alternatives to performance on this instrument seems to lie in the necessity for the S to recall the structure and meaning of information given in previous steps in order to make decisions in the present step. This appears to be the same abilities as described by memory for semantic systems (structure) and implications (meaning).

In the concept identification task, only the number of errors had an important amount of variance predicted by the optimum set. Three SI tests, Logical Reasoning, Unlikely Things, and Utility Test - fluency, were significantly, and negatively related to number of errors. Logical Reasoning appears to be logically related, in that semantic reasoning is necessary to interpret feedback and determine which dimensions are relevant. Unlikely Things does not appear to be logically related because of the SI factor it represents, but rather through the fact that both instruments depend on the S's ability to notice pictorial details.

Utility Test - fluency may be related to performance on the concept identification task in that it is necessary in both tasks to generate a number of possible solutions to the problem-divergent production.

A number of SI tests were also related to the total solutions measure of the Verbal Maze instrument. The significantly related tests were Remembered Relations, Unlikely Things, Logical Reasoning, Plane Flight Test, and Learned Information. The underlying cause of relationships appears to lie in the fact that the Verbal Maze is a semantic system, a relation among relations where the relation is "can talk to". Thus Remembered Relations (memory for semantic relations) and Learned Information (memory for semantic systems) may be related in that they measure the recall of semantic relations and systems. These correspond to remembering the name pairs and the connections among name pairs. Unlikely Things (evaluation of semantic systems) appears to be related in that the problem task requires the S to judge the adequacy of his solutions which are semantic systems evaluation of semantic systems. The two reasoning tests, Logical Reasoning (evaluation of semantic implications) and Plane Flight Test (cognition of semantic systems), appear to be logically related tothe task in that the task requires reasoning of the form: if A can talk to B and B can talk to C then one can pass a message from A to C. This type of reasoning encompasses both semantic and general reasoning specified by the two tests.

From the above discussion it is evident that a small number of relations existed between the SI tests and the problem-solving measures. These relations lend some weight to the assertion that some human abilities are important to problem solving. However, it is important to remember that the ability measures that were related to the problem-solving criteria predicted only a small portion of the criterion variance. Thus the relations are generally weak.

Structural analyses.

The factor analytic procedures employed in these analyses generally revealed very little about the relationships between the SI tests and the problem-solving criteria. The purpose of these analyses was to determine the underlying factors common to both the SI tests and the problem-solving criteria, and in this way to elucidate relationships between the two batteries.

Conventional Factor Analysis. The conventional factor analysis revealed five factors, none of which resembled SI factors. The first factor was a general test performance factor, similar to a general intelligence factor, which had loadings of all the SI tests and at least one measure from each of the problem-solving criteria. The second and third factors were instrument specific and represented the concept identification task and The New Bike with New Uses. The fourth factor represented divergent production of semantic products. Only the last factor demonstrated any crossover between the two batteries. However, this factor made little logical sense, and since the important factor-test correlations were barely greater than .30, the factor was judged unimportant. Thus the conventional factor analysis revealed nothing new concerning the relationships between the two batteries.

Canonical correlation analysis. The canonical correlation analysis yielded only one important canonical factor. The test-factor correlations revealed that this factor was highly similar to the general test performance factor found in the previous analysis. The only new information this analysis added was that very little of the general performance variance in the problem solving criteria was predictable from the SI tests.

Extension loadings analysis. The purpose of this analysis was to discover how the problem-solving tests fit into the SI factor structure supposedly represented by the SI tests. However, considerable difficulty was encountered in reproducing the SI factor structure. This structure was approximated only when the questionable procedure of extracting all of the variance by factors was employed. The extension loadings provided little new information. This analysis revealed only that Learned Information appeared to be related to The New Bike measures.

Possible explanations for the poor results. The poor results of all these analyses can be traced back to the correlation matrix of the SI tests and problem-solving criteria. These correlations were generally quite low, and a large number did not exceed a magnitude which would be significantly different from zero. Thus the expectation of factors across the two batteries could not be high. In turn, these low correlations can be directly attributed to the low reliabilities of the problem-solving criteria. According to classical test theory (Lord and Novick, 1968), the correlation between two tests is limited by the square root of the product of the reliabilities of the two tests. Thus the low reliabilities of the problem-solving criteria severely limit the possible magnitude of correlations between them and the SI tests. The possible explanations for these low reliabilities were discussed in the previous section. This appears to be the primary reason for the lack of evident relationships in the multivariate correlational analyses.

Several other more general factors may have also affected these analyses. In addition to those factors which could affect the reliability of the problem-solving criteria, characteristics of the specific tests and the sample may have affected the outcome.

The New Bike placed great deal of emphasis on the reading ability of the Ss. It is possible that the importance of reading ability overshadowed the effects of all other abilities. Thus the reading ability of the S may have dominated performance on the task, to the exclusion of most other abilities. However, the low correlations of The New Bike with Word Completion argue that this possibility is not highly probable.

Another factor which may be characteristic of all of the problem-solving tasks is the complex nature of the problem-solving involved. The possibility exists that the measures derived from the instruments did not suit the complexity of the tasks. If problem-solving is a complex interaction of a number of human abilities, it may not be reasonable to assume that gross measures which count events across the entire span of the problem-solving episode will reflect accurately the functioning of these abilities. All the measures derived from the problem-solving tasks sampled behavior across the entire task, so it may be that the importance of some abilities may have been masked by the gross nature of the measures.

A factor which may have affected the concept identification tasks is the nature of the task content. The S had to identify a two dimensional geometric concept. All exemplars were geometric figures. Yet none of the SI abilities represented by the SI tests dealt with figural rather than semantic content. If abilities focusing on figural content were of primary importance, few members of the SI test battery would be related to performance on the problem-solving task.

A characteristic of the sample may have also affected the relationships between the two sets of tests. Guilford's SI tests were developed primarily using samples of servicemen, college students, and high school students. The question therefore arises: Do these tests measure the same abilities among fifth grade children as they do among adults? Some evidence exists (Torrance, 1966b; Wallach & Kogan, 1965; Getzels & Jackson, 1961) that children possess divergent production abilities at a young age. However, it is open to question if such tests as Logical Reasoning, Apparatus Test, and Judging Object Adaptations elicit the same responses among adults as they do among children. It is quite possible that the vocabulary component of these tests is much more important for children than it is for adults. If this was true, then adult abilities measured by these tests would not be present in the scores of the Ss in the sample. Thus the relationships between the adult abilities and the problem-solving behaviors would not be present.

One final contention concerning problem solving by Newell and Simon (1972) may have also influenced the lack of relationships. They contend in their theory of problem solving in a human information processing system that:

> A few, and only a few, gross characteristics of the human IPS (information processing system) are invariant over task and problem solver (underlining added)

This implies that there are not a large set of human abilities which are important to all problem solving situations, but rather that the abilities employed vary from task to task and person to person. If this contention were true, then performance on the problem-solving criteria of this study would depend both on the task and the problem solver. Since all analyses were aimed at extracting common characteristics across individuals, this dependence on the individual would effectively remove the possibility of the analyses yielding such general characteristics. Thus, few, if any, relationships would be found between the SI tests and problem-solving criteria.

It would appear from the above discussion that there are several possible explanations for the lack of significant relationships between the SI tests and the problem solving criteria. One or several of them may have, in fact, been true for this study. However, concrete evidence exists for only one explanation. From the first study it was known that the reliabilities of the problem-solving criteria were quite low, and thus the correlations with the SI tests were limited. The resulting low to zero correlations effectively prevented the elucidation of many possible relationships.

Inability to reproduce a Structure of Intellect factor structure. One final question must be dealt with. Why was an SI factor structure not reproducable from the intercorrelations of the SI tests? One very simple explanation is that in order to produce a given factor it is necessary to have at least two tests which represent that factor, and they must have a higher correlation between them than each has with all other tests. This study had only one test representing each of the chosen SI factors, and thus the likelihood of producing SI factors was extremely small.

Even if the proper number of tests were present, there is some question of whether or not the SI structure could be reproduced using strictly empirical methods. To quote Horn (1970)

> The fact that subjective methods were allowed in research upon which the SI model is based invokes the uneasy feeling that perhaps - in what degree one can only surmise - what are referred to as distinct abilities are, in reality, only drawingboard vectors created by overfactoring and rotated to fill otherwise unfillable cells in the SI cube ...

When this type of objection is combined with the possibility that some SI tests may not measure the same things in children as they do in adults, the possibility of reproducing SI factors using test performances of fifth graders seems even more remote.

Summary and Conclusions

The project encompassed by this thesis developed a set of problem-solving tasks which were believed to be more in line with a set of characteristics for the ideal problem-solving task. However, the reliability study revealed that these tasks failed one important criterion - they were not reliable measures of the behaviors they were sampling. In spite of this, these problem-solving tasks were administered to a large group of children along with 17 human ability measures as specified by Guilford's (1967) Structure of Intellect model. The analysis of these results revealed few relationships between the SI tests and the problem-solving criteria. The few relationships that existed indicated that memory abilities were important to two of the three tasks, a measure of convergent production of semantic transformations was related to performance on the simulated problems, and reasoning and evaluation abilities were related to performance on the concept identification and Verbal Maze tasks. A primary stumbling block in investigating possible relationships was the low reliability of the criteria, though some other factors may have been operative.

The primary conclusion that must be drawn from this project is that the unreliability of the problem-solving criteria severely limited the study and essentially prevented the study from fulfilling many of its purposes. The question of what human abilities are important to problem solving thus remains open. The basic design of this study is still capable of providing information regarding this question, but a number of improvements need to be made. First, the reliability of the problem-solving criteria need to be raised to acceptable levels. Second, an even wider sample of problem-solving behaviors need to be included as marker tests. Third, the population of interest should be adult problem-solvers instead of children. With these improvements, it is quite possible that important contributions could still be made to our knowledge concerning human problem solving.

IMPLICATIONS FOR EDUCATION

The goal of Phase III of this project was to construct an empirically based model for instruction in human problem solving. This model was to build on the results of Phase II combined with guidelines and principles derived from the research literature with respect to problem solving. That is, Phase II was to have established with some reliability, which of the 17 specific SI abilities were related to several different aspects of problem solving. These aspects were reflected in the different measures derived from the problem-solving criteria. It was assumed that these relationships between the specific abilities and the problem-solving measures would serve as preliminary evidence that enhancing the appropriate specific abilities would lead to more efficient and accurate problem solving. The model for instruction in problem solving would then have involved a set of empirically based principles for the nurturance of these important human abilities. Since it was evident that even if the second phase were successful the model could not be complete based solely on this project's results, relevant principles derivable from the problem-solving literature were also to be included.

However, as is obvious from the preceding chapters, Phase II was essentially unsuccessful in establishing a structure of specific SI abilities which were related to the various aspects of problem solving reflected in the measures derived from the problem-solving criteria. Thus, the dilemma of constructing an inadequate model or refraining from model construction was presented. Since the evidence in the problem-solving literature concerning the importance of different specific human abilities was fragmentary and incomplete, a model based solely on the existing literature was judged to be inadequate for instructional purposes. This lack of definite empirical evidence from the research literature, together with the lack of results from Phase II of the project, led to the decision that no useful model could be constructed at this time. That is, no useful and empirically based model for instruction in specific cognitive aptitudes could be constructed from evidence presently available.

LIST OF REFERENCES

Ausubel, D. P. & Robinson, F. G. School Learning: An Introduction to Educational Psychology, New York: Holt, Rinehart & Winston, 1969.

Bourne, L. E., Ekstrand, B. R., & Dominowski, R. L. The psychology of thinking. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1971.

Bunderson, C. V. Transfer of mental abilities at different stages of practice in the solution of concept problems. Research Bulletin RB-67-20, Princeton, N.J.: Educational Testing Service, 1967.

Cattell, R. B. (Ed.), Handbook of multivariate experimental psychology. Chicago, Ill.: Rand McNally & Co., 1966, (a).

Cattell, R. B. The screen test for the number of factors. Multivariate Behavioral Research, 1966, 1(12), 245-250, (b).

Covington, M. V. New directions in the appraisal of creative thinking. Journal of Educational Measurement, in press.

Cronbach, L. J. The meaning of problems. In J. M. Seidman (Ed.), Readings in educational psychology, Boston: Houghton Mifflin, 1955.

Crutchfield, R. S., & Covington, M. V. Programmed instruction and creativity. Programmed Instruction, 1965, 4(1-2), 8-10.

Davis, G. A. Current status of research and theory in human problem solving. Psychological Bulletin, 1966, 66(1), 36-54.

Davis, G. A., Train, A. J., & Manske, M. E. Trial and error versus "insightful" problem solving: Effects of distraction, additional response alternatives, and longer response alternatives, and longer response chains. Journal of Experimental Psychology, 1968, 76, 337-340.

Denny, J. P. The effects of anxiety and intelligence on concept formation. Journal of Experimental Psychology, 1966, 72, 596-602.

Dewey, J. *How we think*. New York; Heath, 1933.

Draper, N. R., & Smith, H. *Applied Regression Analysis*. New York: John Wiley & Sons, Inc., 1966.

Duncan, C. P. Recent research on human problem solving. *Psychological Bulletin*, 1959, *56*, 397-429.

Duncker, K. On problem solving. *Psychological Monographs*, 1945, *58*, (Whole No. 270).

Dwyer, P. F. The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika*, 1937, *2*, 173-179.

Feldhusen, J. F., Houtz, J. C., & Ringenbach, S. E. The Purdue Elementary Problem-Solving Inventory. *Psychological Reports*, 1972, *31* 891-901.

Gagne, R. M. Human problem solving: Internal and external events. In B. Kleinmuntz (Ed.) *Problem solving: Research, method, and theory*. New York: John Wiley and Sons, 1962.

Getzels, J. W. & Jackson, P. W. *Creativity and Intelligence: Explorations with gifted students*. New York; John Wiley and Sons, 1962.

Glaser, R., Damarin, D. E., & Gardner, R. M. The tab item: A technique for the measurement of proficiency in diagnostic problem-solving tasks. *Educational and Psychological Measurement*, 1954, *14*, 283-293.

Guilford, J. P. *The nature of human intelligence*. New York: McGraw-Hill, Inc., 1967.

Guilford, J. P. & Hoepfner, R. *The analysis of intelligence*. New York: McGraw-Hill, 1971.

Harman, H. H. *Modern factor analysis*. (2nd ed., Rev.) Chicago: University of Chicago Press, 1967.

Harootunian, B. & Tate, M. W. The relationship of certain selected variables to problem solving ability. *Journal of Educational Psychology*, 1960, *51*, 331-336.

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

Hoffman, H. N. A study in an aspect of concept formation with subnormal, average, and superior adolescents. Genetic Psychology Monographs, 1955, 52, 191-239.

Holt, John. How children fail. New York: Pitman, 1964.

Horn, J. L. Review of J. P. Builford's The nature of human intelligence. Psychometrika, 1970, 35, 273-279.

Hotelling, H. Relations between two sets of variates. Biometrika, 1936, 28, 321-327.

Hurley, J. R. & Cattell, R. B. The procrustes program, producing direct rotation to test a hypothesized factor structure. Behavioral Science, 1962, 7, 258-262.

John, E. R. Contributions to the study of the problem solving process. Psychological Monographs, 1957, 71, (18, whole no. 447).

Johnson, D. M. Solution of anagrams. Psychological Bulletin, 1966, 66, 371-384.

Johnson, D. M. Systematic introduction to the psychology of thinking. New York: Harper & Row, 1972.

Katona, G. Organizing and memorizing: Studies in the psychology of learning and teaching. New York: Columbia University Press, 1940.

Keisler, E. R. Teaching children to solve problems: A research goal. Journal of Research and Development in Education, 1969, 3, 3-14.

Kozol, Jonathan. Death at an early age. Boston: Houghton-Mifflin, 1967.

Laughlin, P. R. & Jordan, R. M. Selection strategies in conjunctive, disjunctive, and biconditional concept attainment. Journal of Experimental Psychology, 1967, 75, 188-193.

Lord, F. M. & Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison Wesley, 1968.

Luchins, A. S. Mechanization in problem solving: The effect of einstellung. Psychological Monographs, 1942, 54, (whole no. 248).

Lundsteen, S. W., & Michael, W. B. Validation of three tests of cognitive style in verbalization for the third and sixth grades. Educational and Psychological Measurement, 1966, 26(2), 449-461.

Maier, N. R. F. Reasoning in humans: III. The mechanisms of equivalent stimuli and reasoning. Journal of Experimental Psychology, 1945, 35, 349-360.

McCandless, B. R. Children: Behavior and development (2nd ed.), New York: Holt, Rinehart, and Winston, Inc., 1967.

McGuire, C. H., & Babbott, D. Simulation technique in the measurement of problem solving skills. Journal of Educational Measurement, 1967, 4(1), 1-10.

Mendelsohn, G. A., Griswold, B. B., & Anderson, M. L. Individual differences in anagram-solving ability. Psychological Reports, 1966, 2, 429-439.

Merrifield, J. R., Guilford, J. P., Christensen, P. R., & Frick, J. W. The role of intellectual factors in problem solving. Psychological Monographs, 1962, 76, (10, whole no. 529).

Miles, D. T. Development of a test for an experimental research program in creative problem solving. Final Report Southern Illinois University, 1968.

Mosier, C. I. A note on Dwyer: The determination of the factor loadings of a given test. Psychometrika, 1938, 3(4), 297-299.

Mosier, C. I. Determining a simple structure when loadings for certain tests are known. Psychometrika, 1939, 4(2), p. 149-162.

Miller, G. A., Galanter, E. & Pribram, K. H. Plans and the structure of behavior. New York: Holt and Co., 1960.

Nie, N. H., Bent, D. H. & Hull, C. H. SPSS: Statistical package for the social sciences. New York: McGraw-Hill, 1970.

Nattress, L. W. The branching-type written simulation exercise. Paper presented at the annual meeting of the American Psychological Association, September, 1970.

Newell, A., Shaw, J. C., & Simon, H. A. Elements of a theory of human problem-solving. Psychological Review, 1958, 65, 151-166.

Newell, A. & Simon, H. A. Human problem solving. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Osler, S. F., & Fivel, M. W. Concept attainment: I. The role of age and intelligence in concept attainment by induction. Journal of Experimental Psychology, 1961, 62, 1-8.

Ray, W. S. Complex tasks for use in human problem solving research. Psychological Bulletin, 1955, 52, 134-149.

Rimoldi, H. J. A. A technique for the study of problem-solving. Educational and Psychological Measurement, 1960, 20, 449-460.

Sheridan Psychological Services, Inc. Tests selected to represent Structure-of-Intellect Abilities. Unpublished mimeograph paper, 1972.

Silberman, C. E. Crisis in the classroom. New York: Vintage, 1970.

Stevenson, H. W., Hale, G. A., Klein, R. E. & Miller, L. K. Interrelationships and correlations in children's learning and problem solving. Society for Research in Child Development Monographs, 1968, 33(7), (Whole no. 123).

Stewart, D. & Lowe, W. A general canonical correlation index. Psychological Bulletin, 1968, 70,(3), 160-163.

Streufert, S., Kliger, S. C., Castore, C. H., & Driver, M. J. A tactical and negotiations game for the analysis of decision intergration across decision areas. Psychological Reports, 1967, 20, 155-157.

Toffler, Alvin. Future shock, New York: Bantam Books, 1970.

Torrance, E. P. Torrance tests of creative thinking. Princeton, N.J.: Personnel Press, 1966. (a)

Torrance, E. P. The Torrance tests of creative thinking: Norms and techinical manual. Princeton, N.J.: Personnel Press, 1966, (b).

Treffinger, D. J. Solving problems #1 and #2. Mimeographed test used with fifth and sixth grade students. Purdue University, 1970.

Treffinger, D. J. & Poggio, J. P. Needed research on the measurement of creativity. Journal of Creative Behavior, 1972, 6(4), 253-267.

Tyler, R. B. Expectancy for eventual success as a factor in problem solving behavior. Journal of Educational Psychology, 1958, 49, 166-172.

Wallach, M. A. & Kogan, N. Modes of thinking in young children. New York: Holt, Rinehart and Winston, Inc. 1965.

Wallas, G. The art of thought. New York: Harcourt, Brace, and World, 1926.

Werdelin, I. A synthesis of two factor analyses of problem solving in mathematics. Didakometry, 1966, 8.

Wertheimer, M. Productive thinking, (2nd ed.). New York: Harper and Row, 1959.

Whitman, R. N. Concept attainment as a function of intelligence and conceptual styles. Center for study of Cognitive Processes, Wayne State University, 1966.