R E P O R T   R E S U M E S

ED 020 776   24   PS 000 928

AN ANALYSIS OF A CLASS OF PROBLEM SOLVING BEHAVIOR. FINAL REPORT.

BY- ANDERSON, RICHARD C.
ILLINOIS UNIV., URBANA

REPORT NUMBER BR-5-0314   PUB DATE JAN 68
REPORT NUMBER CRP-2942
CONTRACT OEC-5-10-299
EDRS PRICE MF-$0.75 HC-$7.60 188P.

DESCRIPTORS- *COGNITIVE ABILITY, COGNITIVE DEVELOPMENT, COGNITIVE PROCESSES, *DISCOVERY PROCESSES, *CONCEPT FORMATION, *LOGICAL THINKING, TASK PERFORMANCE, PROBLEM SOLVING, *ELEMENTARY SCHOOL STUDENTS, TASK ANALYSIS, FACTOR ANALYSIS, TEACHING SKILLS, TRAINING TECHNIQUES,

A CONCEPT ATTAINMENT TASK IN WHICH CHILDREN ARE ASKED TO IDENTIFY A CONCEPT THROUGH DETERMINING WHICH OF A SET OF OBJECTS ARE OR ARE NOT INSTANCES OF IT WAS USED TO INVESTIGATE THEIR PROBLEM-SOLVING BEHAVIOR. OF PARTICULAR INTEREST WAS THE QUESTION OF TO WHAT EXTENT THEY USED OR LEARNED THE STRATEGY OF VARYING ONE FACTOR WHILE HOLDING ALL OTHERS CONSTANT. IN THE FIRST OF THREE EXPERIMENTS A SAMPLE OF 144 FOURTH GRADERS WERE GIVEN 24 APTITUDE TESTS AND NINE CONCEPT ATTAINMENT TASKS. OF SEVEN APTITUDE FACTORS FOUND, ONLY TWO, CALLED SPATIAL ORIENTATION AND FIGURAL ADAPTIVE FLEXIBILITY, SHOWED SIGNIFICANT CORRELATIONS WITH PROBLEM-SOLVING PERFORMANCE. THE SECOND EXPERIMENT INVESTIGATED THE EFFECT OF THE PRESENCE OR ABSENCE OF ACTUAL OBJECTS ON CHILDREN'S PERFORMANCE. A SAMPLE OF 24 THIRD OR FOURTH GRADERS WERE EACH GIVEN THREE PROBLEMS TO SOLVE. OF FOUR GROUPS, ONLY THE ONE NEVER SHOWN ANY OBJECTS PERFORMED SIGNIFICANTLY MORE POORLY. THE THIRD EXPERIMENT WAS INTENDED TO COMPARE TWO METHODS OF TEACHING THE DESIRED PROBLEM-SOLVING STRATEGY, A PART-TASK METHOD WHICH ATTEMPTED TO TEACH COMPONENT SKILLS AND A WHOLE-TASK METHOD WHICH ATTEMPTED TO TEACH THE TOTAL SKILL. TWO GROUPS OF 18 FIRST GRADERS RECEIVED ONE OR THE OTHER FORM OF TRAINING, AFTER WHICH THEIR PERFORMANCE WAS COMPARED WITH A GROUP OF UNTRAINED FIRST GRADERS. ON RETENTION PROBLEMS THE PART-TASK GROUP WAS SIGNIFICANTLY BETTER THAN THE WHOLE-TASK GROUP, WHICH WAS IN TURN BETTER THAN THE UNTRAINED GROUP. (DR)

ERIC

BR-5-0314-
P.A. 24

ED020776

FINAL REPORT

Project No. 2942
Contract No. OE 5-10-299

AN ANALYSIS OF A CLASS OF PROBLEM SOLVING BEHAVIOR

January 1968

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

Office of Education
Bureau of Research

PS 000928

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FINAL REPORT

Project No. 2942

Contract No. OE 5-10-299

AN ANALYSIS OF A CLASS OF PROBLEM SOLVING BEHAVIOR

Richard C. Anderson

University of Illinois
Urbana, Illinois

January 1968

The research reported herein was performed pursuant to a contract 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

PS 000928

Acknowledgements

The author is deeply indebted to the many who assisted in this research as follows: Rivka Yavin, Jessica Daniel, Marianne Roderick, Carla Cucciatti, Diane Fuglsang, Julianne Boone, Heath Tuttle, Richard Willis, Donald Cunningham, Walden Lewis, Ciro Farina, Merrily Faust, Orpha Duell, Jean Behrens, Elsa Roy, Rick Blomme, John Guthrie, James Rubovits, Gerald Faust, and Phillip Zediker. Appreciation is extended to the staff of the cooperating schools: Bennett School, Hawthorne School, Lincoln School, Mattoon, Illinois; Prairie School, Wiley School, Urbana, Illinois; Tolono School, Tolono, Illinois; Sadorus School, Sadorus, Illinois; and South Side School, Champaign, Illinois. Special thanks are due to Bonnie Anderson who carried the main burden of producing materials and reports.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

ERIC

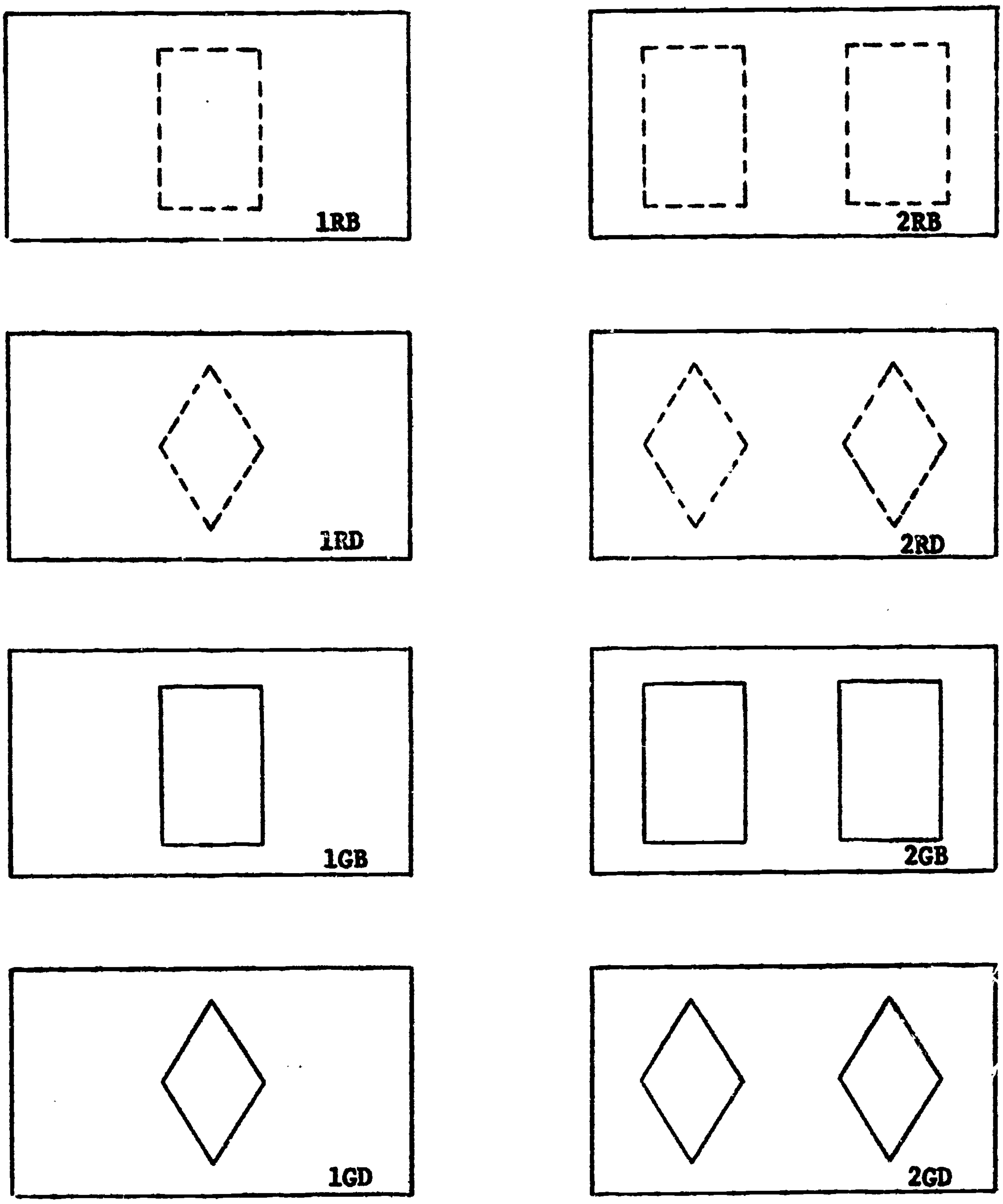
CHAPTER 1

An Analysis of a Concept Attainment Task

Most of the research described in these pages involved a single kind of problem. The subject--a first, second, third, or fourth grader--was confronted with an instance of a phenomenon. His task was to discover the necessary and sufficient conditions for the occurrence of the phenomenon. This he did by making a series of observations, or experimental tests, in which he varied features of the situation to determine which were associated with the presence of the phenomenon and which were not. When he was satisfied that he knew, he described the necessary and sufficient conditions in words. The kind of task which has just been described is often called a concept attainment problem. In this language it is said that the subject is shown a positive or "focus " instance; that he selects instances; receives feedback as to whether the instances are positive (show the concept) or negative (do not show the concept); and, as soon as he can, names the concept.

Description of the class of problems

A concrete example will serve to illustrate the type of problem presented to the child and the sort of behavior expected from him. The materials consisted of eight 2-1/2 X 3-1/4 in. cards taped in an orderly arrangement on a 16 X 18 in. Masonite panel. The cards had figures inscribed upon them that varied with respect to number (one figure or two figures), color (red or green) and form (box or diamond). Fig. 1.1



------- Red outline

——— Green outline

Fig. 1.1. The card array.

contains a representation of the card array. The coded description of each card (e.g. 1RD, 2GB) was not included in the materials the child saw. The following instructions were read to the child:

> I will think of a secret and point to one card that shows my secret. You pick cards to figure out my secret. Each time you pick a card I will tell you whether or not it shows the secret. As soon as you are sure you know, tell me the secret.

The "secret" was one of the 27 different concepts which can be formed with the card array (and the other materials described later), disallowing relational and disjunctive concepts. There is one 0-dimensional concept (<u>all the cards</u>) and there are six 1-dimensional concepts, twelve 2-dimensional concepts, and eight 3-dimensional concepts. The 27 possible concepts and the positive instances of each are listed in Fig. 1.2.

The problem might involve any one of the 27 possible concepts. In no case did the child know which aspects nor how many aspects of the stimulus were discriminative. He began only with a positive or "focus" instance. It was not enough that the child say the "right answer." To be counted as having solved the problem, the child had to choose a set of cards and state a concept such that the set of cards logically implied the concept he stated and no other.

Compare the following two protocols. The only difference between them is with respect to event #4.

<u>Child A</u>

1. The experimenter reads instructions then points to 1RD indicating that it "shows the secret."

2. The child points to 1GD.

3. The experimenter says "No, that (indicating 1GD) does <u>not</u> show my secret."

Concept name	Positive instances
All the Cards	All the Cards
One	1RB, 1RD, 1GB, 1GD
Two	2RB, 2RD, 2GB, 2GD
Red	1RB, 1RD, 2RB, 2RD
Green	1GB, 1GD, 2GB, 2GD
Box	1RB, 1GB, 2RB, 2GB
Diamond	1RD, 1GD, 2RD, 2GD
One Red	1RB, 1RD
One Green	1GB, 1GD
One Box	1RB, 1GB
One Diamond	1RD, 1GD
Two Red	2RB, 2RD
Two Green	2GB, 2GD
Two Box	2RB, 2GB
Two Diamond	2RD, 2GD
Red Box	1RB, 2RB
Red Diamond	1RD, 2RD
Green Box	1GB, 2GB
Green Diamond	1GD, 2GD
One Red Box	1RB
One Red Diamond	1RD
One Green Box	1GB
One Green Diamond	1GD
Two Red Box	2RB
Two Red Diamond	2RD
Two Green Box	2GB
Two Green Diamond	2GD

Fig. 1.2. Possible concepts with the card array materials.

4. The child points to 1GB.

5. The experimenter says "No, ..."

6. The child points to 2RD.

7. The experimenter says "Yes, ..."

8. The child says "The secret is Red Diamond."

Child B

1. The experimenter reads instructions then points to 1RD indicating that it "shows the secret."

2. The child points to 1GD.

3. The experimenter says "No, that (indicating 1GD) does not show my secret."

4. The child points to 1RB.

5. The experimenter says "No, ..."

6. The child points to 2RD.

7. The experimenter says "Yes, ..."

8. The child says "The secret is Red Diamond."

Child B has solved the problem, that is, he has completed a series of observations which can be taken to prove that the concept is Red Diamond. Child A stated the correct concept, but the observations he made did not logically eliminate the possibility that the concept was Red.

Rationale for the study of a class of problems

Generic definitions of problem solving are valueless since problem solving is a family resemblance term. There are several reasonable and customary usages of the term and a variety of tasks which are called problems. For this reason there will be no attempt to justify the task chosen for intensive study in the present experiments by appeal to purportedly generic definitions of problem solving.

Rather, the defense is that the task described on the preceding pages is a representative of a peculiarly important class of tasks. The task entails a pattern of inductive inquiry that has been important in the rise of modern science and also plays a role in such practical problem solving as medical diagnosis and the trouble shooting of malfunctions in complex equipment. Bruner, Goodnow, and Austin (1956), whose provocative book <u>A Study of Thinking</u> laid the groundwork for the analysis of concept attainment, employed the case of a neurologist interested in pattern vision in the monkey to illustrate the parallel between inductive inquiry in science and a task such as a card array problem. "More specifically, [the neurologist] is interested in six cortical areas and their bearing on pattern vision. He knows that, with all six areas intact, pattern vision is absent. His technique of research is extirpation. In planning his research, how shall he proceed? Destroy one area at a time? All but one at a time? In what order shall he do his successive experiments?" (p. 81)

Suppose the neurologist destroys five of the six cortical areas and finds that pattern vision is absent. This experiment has proved exactly nothing. The neurologist knows that one or more of the cortical areas is essential for pattern vision, but he knew that to begin with. Suppose, on the other hand, the five cortical areas were extirpated and pattern vision were unimpaired. If this were the case, the problem would be solved in one fell swoop. The remaining area must be the one and only area necessary for pattern vision (provided it is not true that one or another of several areas can maintain adequate pattern vision; in other words that pattern vision is absent only when each involved area is destroyed).

An alternative strategy is to extirpate one cortical area. If, as a consequence, pattern vision is absent, the experiment has proved that the involved cortical area is essential. If pattern vision is unimpaired, then the area which was extirpated is irrelevant to pattern vision. Unlike the strategy of destroying five areas at once, the strategy of destroying just one guarantees that useful information will be obtained regardless of the outcome of the experiment. On the other hand, it would be impossible to solve the whole problem with just one experiment using the latter strategy.

Bruner and his associates have made much of "risk regulation" as a function served by a problem solving strategy. In the process they have underestimated the risk of loss of information involved in a "gambling strategy." While the neurologist who destroyed five areas of the cortex might be lucky enough to discover the one cortical area essential for pattern vision, then again he might not be. So long as in reality exactly one cortical area was necessary for pattern vision, the strategy of extirpating some set of five areas at a time would be as efficient--no more, no less--as extirpating one at a time. Suppose, however, that two independent cortical areas have to be intact for pattern vision to be present. If this were the case, the problem could <u>never</u> be solved by destroying sets of five areas at a time. Pattern vision would always be absent and there would be no way to tell which cortical areas were involved. If only four areas were destroyed at a time the same problem would arise should it happen that three of the cortical areas were necessary for pattern vision. Even supposing that exactly two of the six cortical areas were involved in pattern vision, it would be inefficient to locate them by destroying four areas at a time. There are 15 different combinations of six things taken two at a time. Nothing will be learned from any of the experiments in which pattern vision is absent because there will be no way to tell which cortical area is responsible. Thus it could take as many as 15 experiments to find the single case in which four areas have been destroyed but pattern vision remains intact. It would take six experiments at most to discover the cortical areas that control pattern vision using the strategy of extirpating one area at a time and, unlike the strategy of destroying several areas at once, this strategy would work successfully no matter how many areas were actually involved.

The notion developed in the preceding paragraphs is that the optimum strategy for solving concept attainment problems is to vary or manipulate each factor in succession while holding all other factors constant. This, of course, is the classical strategy of experimental science. It has been called the Method of Difference by John Stuart Mill, who was among the first to analyze inductive methods rigorously. Stebbing (1888, p. 120) wrote the following about the Method of Difference in _Analysis of Mr. Mill's System of Logic_:

> "_Canon_--If an instance in which the phenomenon in question occurs and an instance in which it does not occur have every other circumstance in common save one, that one occurring only in the former,--that circumstance in which alone the two instances differ, is the effect or the cause of a necessary part of the cause, of the phenomenon.

The principle is that of comparing an instance of the occurrence of a phenomenon with a similar instance in which it does not occur, to discover in what they differ.

Remarks:--

1. This method is more particularly a method of artificial experiment (its ordinary use being to compare the condition of things before, with those after, an experiment), because--

2. It is commonly employed to determine the effects of given causes; and because--

3. The instances which it requires are rigid and definite--they must be exactly alike, except that in one the phenomenon must be present and in the other absent.

4. If this method is inapplicable, it is usually because artificial experiment is impracticable.

5. It is the only method, of direct experience, by which laws of causation can be proved.

6. If the instances fulfill exactly the requirements of the Canon, this method is perfectly rigorous in its proof."

The method of varying each factor in succession while holding all other factors constant is the most efficient strategy there is for attaining concepts. A concept can always be attained using this strategy in exactly as many trials (experiments, observations) as there are variables (factors, stimulus dimensions) to be processed. Any other strategy will on the average (assuming there are no predictable constraints on the nature of the concept to be attained) require more trials than there are variables. The superior efficiency of the Method of Difference follows from the well-known theorem of information theory that the most efficient information processing procedure is the one that reduces the set of possible concepts by exactly half at each step. The strategy of studying each factor holding all other factors constant is the only strategy for attaining conjunctive concepts that reduces the set of possible concepts by exactly half at each experiment. Consider the pattern vision illustration. Each of six cortical areas either is or is not relevant to pattern vision. Thus, there are 2^6 combinations, or concepts, each representing the relevance of one or more areas and the irrelevance of the rest.[1] One area is extirpated. If pattern vision is impaired the half of the combinations which do not involve the extirpated area can be eliminated. On the other hand, if pattern vision is intact, the half of the combinations which do involve the extirpated area can be discounted. In either event, 2^5 combinations will remain possible. The next experiment involving the next cortical area will reduce the number to 2^4, the next to 2^3, and so on until the problem is solved. As already explained a strategy that entails simultaneous variation on more than one factor at a time will sometimes reduce the possible concepts by more than half but that more frequently less than half will be eliminated and that such a strategy is on the average less efficient than the strategy of varying only one factor at a time.

Superficially, the complex designs science frequently employs may seem to violate the maxim that simultaneous variation of several factors is inefficient. Consider the factorial experiment of social scientists for instance. Here the counterbalancing makes it possible to contrast

---

[1]Actually there are 2^6-1 combinations, or concepts, since the case in which none of the areas is relevant can be eliminated on the basis of the givens of the problem.

ERIC

the levels of a single factor. The counterbalancing is the logical equivalent of holding all other factors constant.

The purpose of this section has been to construct a rationale for the intensive study of a kind of problem. The chief argument has been that the concept attainment task is a representative of a large and socially-important class of problems. Other representatives of the class include certain forms of scientific inquiry, medical diagnosis, and trouble shooting. Another conclusion is that, despite the inference permitted by previous investigators, there is an optimum strategy for the solution of concept attainment problems.

Limits of the analogy between the concept attainment task and a larger class of problems

Much of the really critical genetic research during the last century was completed with sweet peas, fruit flies, and bread mold. These organisms were employed--instead of, say, oak trees and elephants, or people--mainly because they are convenient to study. It was an article of faith that the hereditary mechanisms underlying, for instance, the presence of curly wings or straight wings in the fruit fly would shed light on the genetics of living things in general.

The concept attainment task is intended as a *Drosophila* more easy to study than other representatives of the class of problems. Thus far the attempt has been made to show that the essential logic of a concept attainment problem and a problem in scientific inquiry, if not identical, have much in common. However, there are differences.

The first difference is that the person working at a concept attainment task begins with a well-defined system whereas defining the system to be investigated is the first problem in problem solving more generally. The "system" is the set of interacting variables. A system is "well-defined" for a person when he can enumerate the potentially relevant variables (e.g. list the areas of the cortex that might control pattern vision), name the values each variable may take on (e.g. answer the question "What are the different colors these cards show?"); find or create the exemplifying instances given the name of any concept (e.g. respond to the request "Find the compound, pinate leaves."); name the concept given any complete set of exemplifying instances (e.g. answer the question "These cards (1RB, 1RD) show the secret; what is the secret?").

A second difference is that in the concept attainment task, at least the form employed in the present research, the instances are physically arrayed in front of the subject. In contrast, the scientist who is trying to "attain the concept" of the necessary and sufficient conditions for the occurrence of a phenomenon seldom has all the possible instances available within his visual field from the start. He must create or search out the instances. Let us be perfectly clear. An "instance" is the combination of one value from each of the factors in the system. A card showing 2RD is an instance in a system of cards with figures inscribed upon them that vary with respect to form, color, and number. An organism with two specified areas of the cortex extirpated and four others intact is an instance in a system involving the extirpation or nonextirpation of six cortical areas. The pattern vision problem could be made just like the card array problem if, in advance of checking impairment of vision, each of 64 organisms were prepared with one of the 64 combinations of destruction and nondestruction of the six areas of the cortex. These instances could then be arrayed in front of the investigator. Or, the transformation could be made the other way around. The card problem could be made like the pattern vision problem if the person were given pencils with green and red lead and templates for constructing a box or a diamond.

In the concept attainment task--unlike otherwise similar representatives of the general class of problems--the instances are physically arrayed in front of the problem solver. According to Piaget, this should make a difference for children since they are said to be unable to "conceive combinatorial possibilities." Translating this into operational language, it refers to the behavior of producing or creating instances which will meet the logical requirements for problem solution. Anderson (1965) found that children who received training did very well on problems in which the instances were physically arrayed in front of them but less well on problems in which the instances had to be created. It would be interesting to find out if this occurred because of lack of skill in producing instances or for some other reason. One of the experiments reported later investigated this issue.

A frame of reference

It is popularly said that people "employ strategies" to solve problems. Terms such as "strategy," "plan," "decision," and "hypothesis" are redolent of volition, conscious ego control, and awareness. The customary language of those involved in problem solving research tends to make a person visualize problem solving as a series of deliberate decisions of which the problem solver is aware. It may be that behavior that is accompanied by awareness and a sense of volitional control is different from or even more adaptive than other behavior. The evidence seems to indicate, for example, that people whose verbal behavior is modified by a social reinforcer will later be able to describe the class of responses that was reinforced (Spielberger and DeNike, 1966). Nonetheless, the causal role, if any, that is played by "awareness" is obscure. If awareness means anything useful it refers to implicit verbal and, perhaps, nonverbal behavior. The person "says something to himself." This response has stimulus propertie which could in part control subsequent behavior. "Listening to one's self" say something should be approximately equivalent to hearing someone else say the same thing.

It hardly seems inevitable that implicit verbal and nonverbal behavior is of overriding importance in problem solving. The controlling variables may operate directly on publicly observable behavior without mediating chains of implicit responses. This is a position which, for several reasons, investigators in the area of problem solving are disinclined to believe. One reason is to be found in the prevailing language of discourse.

It is often said that a subject "applies a rule." Usually such an ascription is entirely metaphorical. If a rock is thrown into the air, there is a one-to-one correspondence between a set of points in the air that mark the flight of the rock and the set of points on a graph that satisfy some equation. It would not be said that the rock was "applying" the equation nor that there was a mechanism in the rock programed according to the equation. The state of physics being what it is, these ascriptions are not likely to be made. Such ascriptions are made by psychologists and educators, especially when the subject of discussion is problem solving. The difficulty lies in the confusion between the literal and metaphorical senses of the phrase to "apply a rule."

A person is "applying a rule" only when he completes a specific chain of verbal responses. For instance a person faced with the problem below could be said to "apply a multiplication algorithm" if he ran off the following chain: "four times four is sixteen (writes 6)--carry one--four times one is four, plus one is five (writes 5 to the left of 6)--four times three is twelve (writes 12 to the left of 5)."

$$\begin{array}{r} 314 \\ \underline{\times\ 4} \end{array}$$

If no one observed a chain of responses such as this it still might be imagined that the person applied this multiplication algorithm. However it would be entirely possible that he applied some other algorithm (e.g. "four times 300 is 1200--four times 10 is 40, 1240--four times 4 is 16, 1256"). And it is possible, though perhaps not likely, that the stimulus controlled the response directly, that the person got the answer without applying an algorithm. Notice that the larger multiplier consists of the first three digits of π . Some people have learned the products of π and single-digit numbers much like they learned the multiplication tables.

ERIC

In other cases it is completely misleading to say that a person is "applying a rule." There is, for instance, no useful sense in which a baseball outfielder who catches a flyball after a long run can be said to be applying the laws of physics. If he were applying laws of physics he would be substituting estimated values into equations in order to determine the course he should take to intercept the ball. He does no such thing. His behavior is under more direct stimulus control. It is permissible to say that the outfielder behaves as though he were applying rules of physics. However the person who allows himself to drop the "as though" runs the risk of entering epistomological quicksand.

Those who study language seem fascinated by the paradox they have created for themselves: children apply rules of grammar yet children know no rules of grammar, indeed, really no one but linguists do. Herein lies another example of confusion between the literal and metaphorical senses of the phrase "to apply a rule." A native speaker of a language seldom applies rules of grammar when he utters well-formed sentences in ordinary conversation. The same person may apply grammatical rules some of the time when writing a manuscript, and he almost certainly does when he later edits the manuscript.

To the present author's way of thinking, talk about "strategies," the "application of rules," and the "testing of hypotheses" is not objectionable primarily because it is speculative or because it refers to purported events which are unobservable. The big problem with constructs such as strategy is that they beg the most interesting questions.

Problem solving behavior can be understood in the following terms. The more or less organized sequences or patterns of behavior, popularly called strategies, that seem to comprise much problem-solving behavior at a molar level of description, are hypothetically under stimulus control of two kinds. First, especially on the part of the good problem solver, behavior is under discriminative control of relevant task and problem cues, those present at the beginning and, depending upon the kind of problem, those cues that emerge during the course of work on the problem. Second, problem-solving behavior is partly controlled by response-produced stimuli, including, perhaps, intraverbals or other forms of mediation. Reinforcers of several kinds--for instance, we might suppose, what Polya (1954) calls "signs of progress"--act to increase the likelihood of certain forms of the behavior that is emitted during the problem-solving episode. If there is an orderly relationship between the contingencies of reinforcement operating during problem solving and the criteria for problem solution, then the problem-solving episode is a sort of microscopic natural-selection process. Presumably, three processes--discriminative control of problem-solving behavior by task cues, control of behavior by response-produced cues, and selection of forms of behavior by reinforcement--are sufficient to account for the coordinated, purposeful character of some problem-solving behavior that has seemed so salient to many students of problem solving.

<u>A functional-behavioral analysis of a problem-solving skill</u>

A rough distinction will be made between a functional and a behavioral analysis. A functional analysis deals with molar processes and skill components. The emphasis in a functional analysis is on the task and on the processes which must be assumed in order for the task to be accomplished. The behavioral analysis attempts to detail the specific chains of responses and indicate the controlling variables at every step. The analysis will begin at a functional level and progress to a behavioral one.

The person attempting to solve a concept attainment problem will engage in two easily-distinguishable kinds of observable behavior. He will select instances and he will name concepts. In other words he will engage in instance-selection behavior and in concept-naming, or conclusion-drawing, behavior. Optimum behavior will entail the selection of as many instances as there are variables in the system and then the statement of the concept. There are three variables--form, color, and number--in the card array task. Thus, every card array problem can be solved after the selection of three instances. A problem has been solved when the instances selected imply the concept stated and no other. The question is what must be assumed in order for a person to make an optimum solution to any concept attainment problem. While a logical relationship defines the solution of a problem, there is no need to assume that the successful problem solver "applies logic." Perhaps he could solve problems by applying logic, but he need not. The distinction intended is the one developed in the preceding section.

Let us begin by analyzing conclusion-drawing behavior independently from instance-selection behavior. Assume that a subject receives the following four-card array instances in serial order. The + and - indicate whether the instance "shows the secret" or not.

1GB+
1RB+
2GB-
1GD+

These instances imply One and only One. What process has to be assumed in order to explain how a person could draw this conclusion, and the correct conclusion from any well-formed set of instances? Perhaps it is unfortunate that there is not a single answer to this question. There is an embarrassingly large variety of processes that could result in perfect conclusion-drawing behavior.

Of historic interest is what may be called an "information-theoretic model," reminiscent of Hovland (1952). The person makes a list of the possible concepts. The list consists of all the combinations that can be formed from the values of the initial positive instance. For the above example, the list of possible concepts is as follows:

All the Cards
One
Green
Box
One Green
One Box
Green Box
One Green Box

As each new instance appears the person deletes half of the concepts in the list. If the new instance is positive, the concepts containing the feature which was varied are eliminated. If the new instance is negative, the concepts are deleted that do not name the feature of the original positive instance which failed to appear in the new instance. Remember we are talking about conclusion-drawing behavior under the assumption that somehow an appropriate series of instances is available; otherwise, without additional assumptions, the process that has just been described would not always lead to a correct solution.

The information-theoretic model is beyond doubt a poor representation of people's conclusion-drawing behavior on concept attainment problems. To use the term of Bruner and his associates, the process involves too much "cognitive strain." It seems improbable that the neurologist enumerates for himself the 64 possible combinations of areas of the cortex which could control pattern vision.

Even if he used a paper and pencil to reduce "cognitive strain" the neurologist is unlikely to go to the trouble to list the possible concepts, because listing the concepts is probably not a component of his problem-solving behavior. A second conclusion-drawing process may more closely approximate the behavior of the neurologist. After each new instance the person draws a partial conclusion. If the new instance is positive, the conclusion is that the factor which was varied is irrelevant. If the new instance is negative, the conclusion is that the feature of the original positive instance which failed to appear in the new instance is a part of the concept. When the complete series of instances has been processed, the partial conclusions are combined to give the final conclusion. With the illustrative card array problem this process would work as follows:

Instances	Conclusions
1GB+	
1RB+	Color irrelevant
2GB-	One necessary
1GD+	Form irrelevant
——→	One

This will be called the Method of Difference because it most closely resembles the procedure outlined and named by John Stuart Mill. In their most self-conscious moments scientists probably can be said to be applying the Method of Difference when they work on a concept attainment problem.

A third approach will be called the "common elements" method. The conclusion is drawn by naming the elements, or features, which the positive instances have in common. Under this procedure the conclusion is not drawn until a complete set of instances is available. With respect to the illustrative set of card instances, all of the positive cards show one; there is no other common feature. Thus the conclusion is One.

A fine grained analysis of the common elements method will be attempted since a training procedure described later was based on this method. The conclusion-drawing phase begins when there is a complete, "well-formed" set of instances available. The discriminative stimulus (S^D) marking this state of affairs is a byproduct of the instance-selection process which will be described later. In the first phase of the conclusion-drawing process the problem solver scans the positive instances. One of the discriminative stimuli (S^Ds) for the scanning response is whatever cue has identified an instance as positive. In the case of the card array task, the S^D consists of the words of the experimenter ("Yes, that shows my secret."). Remember that the stimuli arrive serially in time, hence, the S^D will not be temporally proximate to the response it presumably controls, and therefore some form of memory for positive instances will be required. The matter of memory will be treated when instance-selection behavior is considered. The problem solver scans features of the instances. It is assumed that the preproblem history of the problem solver has established each feature as a reliable S^D for a naming response. During the problem he must get S^Ds for <u>which</u> features to name. These S^Ds come from the check on whether features are common to all positive instances. The assumption is that the problem solver does not simultaneously scan all features, that he probably scans them in some serial order. The question of <u>what</u> the order is will be considered when instance-selection behavior is discussed. Since, by assumption, the final conclusion is not determined at an instant but is instead built up in stages, memory is once again implicated. Probably the problem solver verbally encodes the parts of the conclusion at the moment the S^Ds for each of them is available. That is, he says the partial conclusion to himself or outloud; the S^D evokes the implicit or explicit verbalization.

We turn now to an analysis of instance-selection behavior. Each instance selected must have the property that it is identical to the focus instance or other positive instances in every respect save one. The entire set of instances should have the property that each of the stimulus dimensions is varied once and only once. The basic process involved in selecting instances is a compare-and-contrast operation.

A discrimination is easier if there is immediately available a standard for comparison. Herein lies the reason simultaneous discrimination is generally easier than successive discrimination. The person can use the positive instance singled out by the experimenter at the beginning of a concept attainment problem as a standard of comparison. In the broader context the standard of comparison is given by the "definition of the problem" ("with all six areas intact pattern vision is present, with all six destroyed it is absent;" "at 60 mph the front end begins to shake"). The technique is to compare each prospective instance with the standard. Assume the task involves $\underline{k}$ stimulus dimensions. The prospective instance is an appropriate one if it is the same as the standard in k-1 ways and different in the final way. Bruner, Goodnow and Austin (1956) called this procedure the "conservative focusing strategy." This is the origin of the term "focus instance" for the initial positive instance identified by the experimenter.

In what order will the stimulus dimensions be varied (manipulated, observed)? Logically speaking, order makes no difference. But a first instance must be chosen, then a second, and so on until $\underline{k}$ instances have been selected. Notice that, whatever the first choice, there is a new constraint on the second and later choices. The problem solver must avoid varying a dimension already varied at a previous choice. With each new choice there are fewer "logical degrees of freedom." The statements here about logic refer not to logic which the problem solver is applying, but to the logical relationship, from a third party point of view, between the stimulus situation and the problem solver's behavior. Some aspects of the choice of behavior are "free" in the sense that they need not be, in fact could not be, determined by the S^Ds which embody logical constraints. The "free" aspects of choice are probably controlled by the discriminability and "salience" of the features of the stimulus. There are data on the salience of stimulus dimensions from the concept learning literature which could be brought to bear to predict order of choice. Unfortunately a clean prediction is not possible, since it is not clear whether the most salient dimension will be the one the problem solver manipulates first or the one he manipulates last. Which it is might depend in complicated ways upon the conclusion-drawing process.

Somehow or other as he chooses a series of instances the problem solver must keep track of the stimulus dimensions processed and remaining to be processed. There are several ways this could be done. Each of them involves some form of memory. One way is to verbally encode the dimensions processed (or dimensions remaining to be processed, or both). Before each choice, the problem solver recalls these dimensions. That is, he gives himself an implicit or explicit autoclitic, a self-tact of the form "I have already tried color." If this is what is going on, the completion of each instance selection cycle must provide the S^D for the "review tact."

Another procedure to keep track of progress through the problem is to remember the instances themselves, or perhaps only the positive instances, depending upon the type of conclusion-drawing process.

As long as the instances are physically present this procedure need not involve much verbal encoding, perhaps only the "marking" of each instance in terms of its spatial position in the array. To find out where he has been and where he should go, the problem solver scans the instances (including the focus instance) already selected. If the same feature appears in every instance a dimension is thereby identified which remains to be processed. For instance assume that cards represented as follows have been chosen:

1GB
1RB
2GB

These cards all contain boxes, so form is a dimension which remains to be manipulated. The presence of a box in each previous instance provides the S^D for the selection next of an instance which does not contain boxes.

There is a somewhat more elegant procedure for keeping track of progress through the problem which does involve verbal encoding. If what was called the Method of Difference, or any other procedure which entails partial conclusions after each instance, is employed, the problem solver need only rehearse the partial conclusions (which he will probably need to do anyway in order to avoid forgetting them) in order to remind himself of the stimulus dimensions already processed. This integrated procedure is more elegant in the same sense that a machine with fewer moving parts is good. It is more economical of the problem solver's time and less prone to failure from forgetting, interference, or faulty discrimination.

Thus far several different kinds of conclusion-drawing behavior and several different kinds of instance-selection behavior have been described. The analysis is by no means exhaustive. There are many variations of the procedures described. There are others which were not described and undoubtedly still others which the author has not thought of. Anyone of many processes could produce perfect solutions to any and all concept attainment problems. That is to say, there are an indefinite number of processes that could produce identical protocols of observable responses (instance-selection responses,concept-naming responses) for any concept attainment problem, and with the further restriction that each protocol constitutes an optimumly efficient solution to the problem.

Ostensibly much of the research on problem solving has been concerned with discovering the processes entialed when people solve one or another kind of problem. The "process tracing" study is premised on the naive assumption that people exhibit their processes when they solve a problem. Exhibited are strings of responses. As the preceding analysis has demonstrated, there are many processes which could produce even identical strings of responses. It is not possible directly to see the process that causes the string of observable responses to be run off as it is. This remark is not made under the assumption that implicit responses occurring within the organism are the really important events. Rather the process is to be understood in terms of controlling variables, external as well as, perhaps, hypothesized internal events.

ERIC

Intended here is more than the relatively superficial observation that many things are possible and that, therefore, conceptualizations should be mistrusted. Not only are different processes logically possible, but it seems entirely probable that the process is different for different people and, indeed, different for the same person on various occasions. Consider a mundane parallel. People have various door-opening responses in their repertoires: A person may exert force on the door with his hip, push the bottom of the door with his foot, lift a latch with his left hand and then pull, or turn a knob in a clockwise motion with his right hand and then push, or step on a mat concealing a switch that activates a motor that moves the door. While these responses may vary in strength, the details of the stimulus situation are more important than response strength in determining which behavior sequence will occur. With respect to the concept attainment problem, there is every reason to believe that many people--especially professional problem solvers such as physicians, auto mechanics, scientists, and TV repairmen--have repertoires of alternative behavior sequences. To reiterate once again, the differences among these sequences can be very subtle, so subtle that a record of the easily-distinguishable overt responses will not reveal the differences in the behavior nor in the controlling variables.

The author concludes that there are weak grounds to justify the process tracing approach predominant among those who engage in problem solving research. This monograph represents a different approach. Instead of attempting to discover the process by which people solve problems the approach is to teach more or less optimum problem solving behavior. This approach is justified on at least two grounds. First it is exercise in social engineering of potential practical value, since the skill to be arranged is of demonstrable importance. Second, it may in several ways contribute to knowledge about complex processes in general.

ERIC

CHAPTER 2

Procedures for Presenting and Scoring Problems

In all of the research reported in this monograph problems were administered individually by specially trained experimenters (Es). The nine Es who participated in one or more experiments were research assistants who ranged in age from 22 to 45. Five Es were graduate students in educational psychology, two were housewives, one was a summer undergraduate NSF fellow, and one was a third-year law student supplementing his income with part-time work. There were seven women and two men. All the Es were college graduates. Only one had had extensive experience with children prior to participating in this research.

Each E received 15 or more hours of experience training and presenting problems to children before participating in an experiment. The author monitored one or two hours of each E's preexperimental performance.

Problems (and/or training) were presented in such reasonably private and distraction-free spaces as the cooperating schools were able to provide. These included nurses' offices, storage closets, corridors, a kitchen, an auditorium stage, widely separated stations in cafeteria and gymnasia, and in one instance a janitor's closet.

The research entailed three sets of materials. The card array was described in detail in Chapter I. The other two sets of materials were the cowboy game and the pencil collection. While the three sets of materials involved different stimulus dimensions and different "story lines" were employed with them, an identical problem solving task could be created with each set of materials. It is the task detailed with respect to the card array in Chapter I.

The cowboy game consisted of eight plastic cowboys and accessories imbedded in an orderly arrangement in a 12 X 13 in. plaster of Paris base. The game was placed on a table such that the cowboys faced the child. The cowboys were either standing or riding horses, either wearing a hat or bareheaded, and either carrying a gun (rifle) or not. These characteristics were coded as follows:

Characteristic	Code
Riding	R
Standing	S
Gun	G
No gun	N
Hat	H
Bareheaded	B

An instance can be described by listing the appropriate codes. For instance RNB refers to a cowboy riding a horse who has neither a gun nor a hat. A schematic representation of the cowboy game appears in Fig. 2.1.

A story line was employed with the cowboys. The E said:

> In this game you've got to figure out which kind of cowboy is a friend of the sheriff. I will point to one cowboy who is a friend of the sheriff. You point to other cowboys. Each time you point to a cowboy, I will tell you whether he is a friend of the sheriff. As soon as you are sure you know, tell me which kind of cowboy is a friend of the sheriff.

ERIC

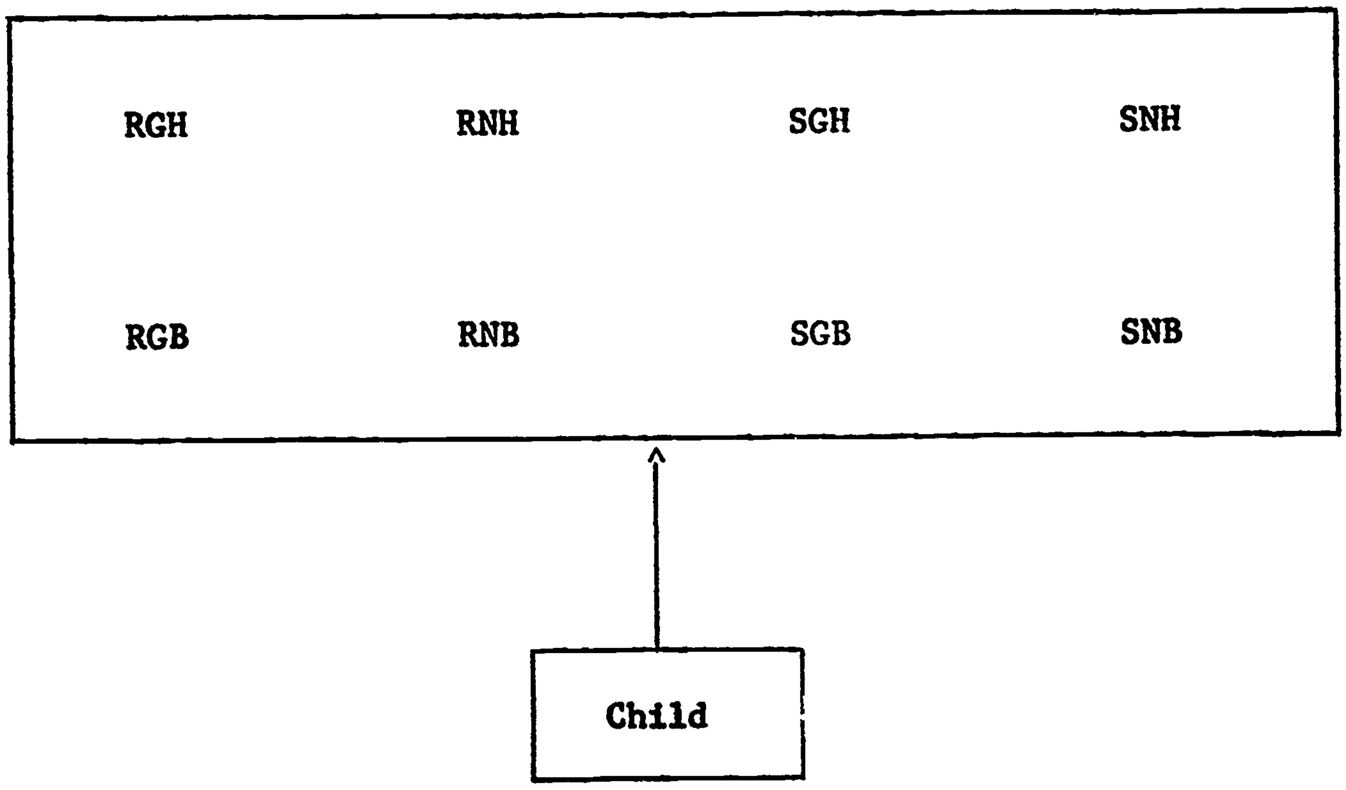


Fig. 2.1. Schematic representation of the cowboy game.

The final set of materials consisted of eight, hexonanal No. 2 pencils. The pencils displayed three attributes: length (6 or 2 in.), presence or absence of eraser, and sharpened or unsharpened. These features were coded as follows:

Characteristic	Code
Long	L
Short	S
Sharpened (point)	P
Unsharpened	U
Eraser	E
No eraser	N

An instance can be described with these codes. For example SUE stands for a short, unsharpened pencil with an eraser. The pencils were placed in a disorderly fashion on a table in front of the child. After each problem, _E_ shuffled the pencils. In all but one experiment to be described in later chapters, the child was allowed to handle the pencils during the course of a problem, for instance, to sort them into groups.

At the beginning of every experimental session in which problems were presented, the child first solved a series of simpler problems designed to provide warmup, to furnish a high initial frequency of reinforcement, and to make sure that the concepts the child would subsequently have to attain were in his repertoire. In this orientation exercise, the _E_ named concepts while the child pointed to all of the positive instances of the concept. For instance, if _E_ were to say, "The secret is long pencils with no erasers," the child would be expected to point to LUN and LPN. The concepts treated in this manner always included the ones which would have to be attained in order to solve problems later in the session as well as fillers. If the child made an error, either of omission or commission, he was prompted to make the correct response and, in addition, any item upon which an error occurred was repeated after an interval until the child made an unprompted correct response.

Administration of the problem

The E's task was to provide the orientation exercise described above, to read the instructions (or quote them verbatim without reading), to indicate the focus instance, to give the child feedback as to whether each instance selected was positive or negative, to respond with a standardized remark to one or another contingency which might arise during the course of the problem, to write a protocol of the problem, and to determine when the problem should be terminated and either the next problem begun or the session terminated.

If the child stopped trying to solve the problem, the E attempted to keep him going with one of a series of "motivational prompts." Specifically, if the child made no task-relevant responses for a period of 10-15 sec., the E said "What are you going to do now?" If 10-15 sec. more elapsed in which there were still no task-relevant responses, the E then said "See if you can figure out my secret by picking more (cards, pencils, cowboys)." When this failed to work, the E exhorted the child to continue using whatever language he believed appropriate.

If the child stated an incorrect conclusion with which he was apparently satisfied, the E attempted to elicit further progress with a "challenge prompt." The first challenge prompt was, "Are you sure that's my secret?" If the child insisted that he was sure or did nothing for 10-15 sec., the E then said, "Maybe that's not my secret." Finally, if the child still persisted, the E usually said, "Maybe the secret could be something else. Try to find out for sure what the secret is by picking more...(cards, pencils, cowboys)." Other wordings were provided for special contingencies. When the child accumulated enough information logically to solve the problem but did not volunteer a conclusion, the E said "As soon as you are sure you know, tell me the secret." When the child seemed to be trying to communicate the concept by pointing out the set of positive instances, the E said "Try to say the secret in words instead of pointing." The language of all the prompts and the contingencies for their use were prescribed. The Es were permitted no unstandardized remarks. A summary of the allowable experimenter statements appears in Fig. 2.2.

There was a three-fold stopping rule. The problem was terminated (1) if the child solved the problem, that is, if he selected a set of instances and stated a concept such that the instances implied the concept and no other; (2) if the child refused to continue despite allowable experimenter prompts; or (3) if the child selected six instances without solving the problem. If the child selected six instances without solving the problem but had enough information to

ERIC

Category	Code	Statement[a]
Motivational prompts		
Slight motivational prompt	M1	"What are you going to do now?"
Moderate motivational prompt	M2	"See if you can figure out my secret."
Strong motivational prompt	M3	Encourage the child to continue in your own words.
Challenge prompts		
Slight challenge prompt	P1	"Are you sure that's my secret?"
Moderate challenge prompt	P2	"Maybe that's not my secret."
Strong challenge prompt	P3	"That's <u>not</u> my secret. Try to figure out my secret by..." or "<u>Maybe</u> the secret could be something else. Try to find out <u>for sure</u> what the secret is by..." or "That's not the secret. We don't have secrets like that in this game. It's one of the secrets we had in the first game. Try to figure out the secret by..."
Eliciting prompts		
Slight eliciting prompt	E_1	"As soon as you are sure you know, tell me the secret."
Strong eliciting prompt	E_2	"I can't let you pick any more ... Can you tell me the secret now?"
Special categories		
Refuses to continue	Refuses to continue	
Put in words	PW	"Try to say the secret in words instead of pointing."

[a]In the case of cowboy problems, the statements should read "the kind of cowboy who is a friend of the sheriff" in place of "my secret."

Fig. 2.2. Allowable experimenter statements.

do so, the E said "I can't let you pick any more...(cards, pencils, cowboys). Can you tell me the secret now?" The Es were instructed to continue the problem when they were not sure whether a child had solved a problem, since, should it happen that the child had solved the problem, the extraneous portion of the record could be eliminated.

The E wrote a protocol for each problem, indicating the concept and focus instance and recording in coded form the instances the child selected, the statements the child made, and the statements E made in the sequence in which these events occurred. Figs. 2.3A and 2.3B contain sample protocols. The numbered entries represent instances the child selected. The + or - indicates whether the instance "showed the secret" or not. Entries preceded by a V (for verbalization) represent the child's concept naming behavior. All other entries represent the behavior of E as per Fig. 2.2.

The dots under the focus instance code are a form of bookkeeping to help E keep track of what the child has proved or failed to prove so that E will know when the problem can be terminated and when eliciting prompts should be given. Nonetheless, there is a minor experimenter error evidenced by the protocol in Fig. 2.3B. The slight eliciting prompt (E1) should have been given after the third instance rather than the fourth. No doubt the E was fooled by the third instance which, through logically a poor choice, happened in this case to eliminate both presence or absence of an eraser and presence or absence of a sharp point from consideration.

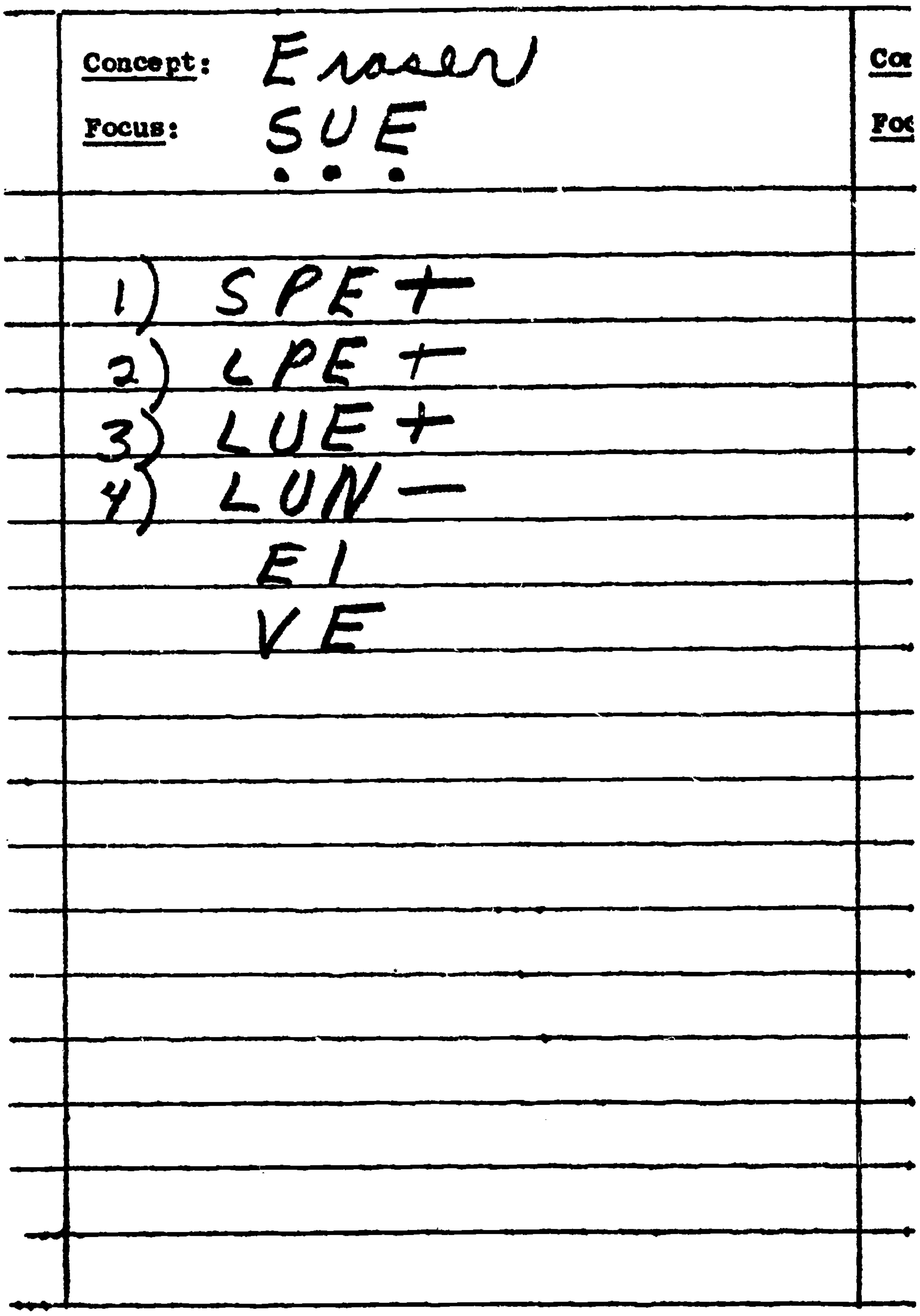


Fig. 2.3A. Sample problem solving protocol involving the pencil collection.

Concept: Long

Focus: LPE

1) SUE −
2) SPE −
3) LUN +
4) LUE +

E1

VALL

P1

P2

5) LPN +
6) SPE −

E2

VL

Con

Foc

Fig. 2.3B. Sample problem solving protocol involving the pencil collection.

Immediately following a day's administration of problems, the Es exchanged protocols to check them for legibility, completeness, accuracy of subject and problem identification, and accuracy of coding.

Scoring the problem

The instances selected by a child can be classified into four mutually exclusive and exhaustive categories. They may be inefficient, redundant, repetitious, or properly chosen. An inefficient choice is so called because on the average it yields less than the maximum attainable amount of information. Operationally, an inefficient choice is an instance which differs along more than one stimulus dimension from the focus instance or other positive instance. A choice is redundant if the outcome, whether the instance is positive or negative, is implied by previous instances. Redundant choices yield no new information. For example, consider the following instances:

1GD+
2GD+
1GB+

At this point 2GB would be a redundant choice since it must be +. Another f[illegible] of redundancy can be seen in the following example:

1RD+
1GD+
1GB-

A choice of either 2GB, 1RB, or 2RB would be redundant because previous instances prove that an instance will be - unless it contains D. In short, a choice is redundant (1) if the instance contains an as yet untried combination of variables known to be irrelevant and the instance is otherwise the same as a previous positive instance or (2) if the instance does not include all of the features previously shown to be relevant.

A repetitious choice exactly duplicates a previous instance.

Some redundant choices are also inefficient (e.g. 2GB and 2RB in the example above). The distinction between the two categories is maintained by the arbitrary convention that an instance is categorized as inefficient only if it is not redundant.

The analysis of concept naming or conclusion stating behavior depends upon the analysis of instance selection behavior. A statement of the concept is either valid, consistent, inconsistent, or "miscellaneous." These categories are exhaustive of the unambiguous concept-naming behavior of subjects. Only unqualified declarative statements about the relevance or irrelevance of variables or unqualified conclusions about the concept are categorized. Any nonconjunctive statement is categorized as miscellaneous. Otherwise, statements are categorized, as the category names suggest, according to whether they are consistent or inconsistent with the instances that have been chosen up to that point. Consistent statements are an inappropriate form of verbal behavior since, when a statement is scored as consistent, there always remains at least one other possible concept.

When a subject makes a valid statement, he has by definition solved the problem; he has stated the one and only correct concept and he has selected a set of instances that proves that it is the correct concept.

ERIC

Sequence is the essence of this system of content analysis. A choice which is inefficient at one point might be proper at another and redundant or repetitious later. Similarly a statement that is consistent at one point in time might be inconsistent later while a statement which is valid might only have been consistent at an earlier point.

The content analysis procedure that has been outlined is entirely objective; there is no need for subjective "judgement" to classify behavior appropriately. Indeed, a computer program was developed to score protocols.

Various indices of the adequacy of problem solving performance based on the content analysis are possible. The following overall summary index has proved useful: Performance is summarized on a three point scale. a) The S solves the problem and neither makes any inefficient, redundant, or repetitious choices of instances nor states any inconsistent, consistent or miscellaneous conclusions (2 points). b) The S solves the problem but makes one or more inappropriate choice or states one or more inappropriate conclusion (1 point). c) The S fails to solve the problem (0 points). Other performance measures were also employed and these will be described as the occasion demands.

There are several advantages to computer processing. In previous research (Anderson, 1964, 1965) protocols were scored by human raters. This is a rather difficult task. People make mistakes. While inter-rater reliability coefficients have been quite high, around .90 (Anderson, 1965, p. 81), this means there is only 80 percent "true score" variance. Computer processing should reduce noise from this source to zero. Second, a number of hours of training is required before a person can score proficiently. Third, the scoring of the protocols is a laborious, time consuming task. After warmup, a proficient rater can score a problem in 5 min. The card-input IBM 1620 computer with disk file upon which the content analysis was completed scored a protocol in about 3 sec. The research reported herein involved over 2,500 protocols, counting neither protocols collected in the course of the development of training procedures nor the training of Es. Thus, a considerable saving in time was realized. (At least, the computer programer's time was returned with a dividend.) Fourth, computer processing obviated the necessity for two manual transformations of the data and manual adding of category totals, both operations in which errors could be introduced. The human rater makes a score sheet for each protocol. Then he makes a summary ledger sheet. If the analysis is large or complex, cards are punched from the ledger sheets. If computer processing is employed, the keypuncher works directly from the protocol. Note that the keypuncher need not make any substantive decisions. Finally, it was possible using the computer to detect and eliminate errors introduced by the E or the keypuncher.

An elaborate series of checks was included within the content analysis program. The accuracy of the concept and focus instance codes was externally verified. From this point on, the computer served as its own proofreader. Each line of the protocol was checked against the concept. When an impossible case was detected the computer printed

ERIC

subject and problem identification and then a message of the form "Trial X (meaning instance) was miscoded by the experimenter or the keypuncher." These cases were manually checked, the error was corrected, and then they were resubmitted to the computer. The computer also checked to see whether the stopping rules were observed. The computer classified cases in which the stopping rules were violated in an "illegal termination" category. These cases were manually checked, corrected, and resubmitted. Finally, all of the problems in which the "right answer" was stated but the problem was not solved were manually checked.

One experiment, to be described in detail later, involved 848 protocols. On the first pass through the machine, the computer rejected 21 protocols, or about 2.5% of the total. A keypunch error had been made in 11 cases. In the remaining 10 cases the fault was in the protocol; in other words, there was an experimenter error. These protocols were examined and a decision was made as to what the codes should have been. The smallest change was made in the protocol which would make it internally consistent. Herein lies the only aspect of the analysis in which subjective judgement was required. Otherwise the analysis was as objective as, say, the scoring of a multiple choice test.

The computer program is printed in Appendix I.

CHAPTER 3

Individual Differences and Problem Solving[1]

This chapter will approach problem-solving from the perspective of the psychologist, especially the applied psychologist, interested in learning or training. The question to be answered is what is there in the literature on individual differences, substantive or methodological, of worth to the person concerned with developing training procedures that facilitate problem-solving performance. Much of the research on problem-solving has been concerned with individual differences, including, for example, such factors as rigidity, availability of function, kind of strategy employed, and cognitive style. Despite some provocative ideas and data, there are grounds for arguing, as others have argued before (e.g., Duncan, 1959; Schulz, 1960) that as a group "process-tracing" studies of individual differences within the context of problem-solving have not led to a substantial corpus of knowledge. Partly for this reason, the emphasis will be upon individual differences in problem-solving in terms of dimensions established in factor analyses of aptitude tests.

In what follows, there will be, first, a brief sketch of a theoretical orientation toward problem-solving and aptitudes and then, second, a consideration of several ways in which the research on aptitude factors may bear upon

---

[1]Reprinted with the permission of the publisher from Robert M. Gagne´, Editor, Learning and Individual Differences. Columbus, Ohio: Merrill Books, 1967, pp. 66-89.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Charles E. Merrill Publishing Co.
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

investigations of the learning of complex skills, such as problem-solving skills. No consideration will be given to such customary topics as increasing the precision of experiments by statistically controlling individual differences, or selecting persons likely to succeed if presented with training. Instead, there will be an attempt to see whether research involving aptitude factors could have value for the development and improvement of training procedures.

Theoretical Orientation

The author believes that an S-R orientation offers the most fruitful approach to problem-solving. The view outlined in the following passage has much in common with the conceptions of others with a similar orientation such as Cofer (1957), Kendler and Kendler (1962), Maltzman (1955, 1960), and Skinner (1957). First of all, problem-solving behavior is not fundamentally different from other varieties of behavior. In Duncan's (1959, p. 425) words, "Problem-solving in human adults is a name for a diverse class of performances which differs, if it differs at all, only in degree from other classes of learning and performance..." The belief is that problem-solving behavior, like other kinds of complex behavior, can be conceived most profitably in terms of colligations of elemental stimulus-response relationships. The more or less organized sequences or patterns of behavior, popularly called "strategies", that seem to comprise much problem-solving behavior, at a molar level of description, are hypothetically under stimulus control of two kinds. First, especially on the part of the good problem-solver, behavior is under discriminative control of relevant task and problem cues, those present at the beginning and, depending upon the kind of problem, those cues that emerge during the course of work on the problem. Second, problem-solving behavior is partly controlled by response-produced stimuli, including, perhaps, intraverbals or other forms of mediation. Reinforcers of several kinds -- for instance, we might suppose, what Polya (1954) calls "signs of progress" -- act to increase the likelihood of certain forms of the behavior that is emitted during the problem-solving episode. If it happens that there is an orderly relationship between the contingencies of reinforcement operating during problem-solving and the criteria for problem solution, then the problem-solving episode is a sort of microscopic natural selection process. Presumably, the three processes that have been listed -- discriminative control of problem-solving behavior by task cues, control of behavior by response-produced cues, and selection of forms of behavior by reinforcement -- are sufficient to account for the coordinated, purposeful character of some problem-solving behavior that has seemed so salient to many students of problem-solving.

Scores on "intelligence" tests can also be given an S-R interpretation; however, it is different from the prevailing one. The concept of trait is one of the key prevailing ideas underlying the psychology of individual differences. Certain tests are said to measure special types of traits called aptitudes or abilities. But it is not necessary to infer a trait behind a distribution of scores on a test. Nor is one required to believe that the results of factor analytic investigations represent "vectors of the mind". Aptitude factors can be regarded as representing independent classes of similar or related behavior. Parenthetically, it should be recognized that trait psychology plays a large part in the "figurative model" (Tatsuoka, 1965) that underlies the mathematics of test theory and the theory of factor analysis. It would be an interesting and maybe valuable exercise to see what sort of individual differences model could be constructed on behaviorist assumptions.

Aptitude Factors and Task Analysis

Tasks can be described in terms of loadings on reference factors (or correlations with either factor scores or univocal reference tests). The pattern of loadings indicates the kinds of skills that are important for success on the task and indicates their relative importance. Possibly descriptions in these terms could be useful to the training psychologist undertaking a task analysis. Such an approach has often been advocated by factor analysts, though, to the author's knowledge, few if any psychologists concerned with training have tried seriously to use the procedure for this purpose.

One of the leading advocates of task analysis by factor analysis, J. P. Guilford, and his associates have completed an investigation in the area of problem-solving that illustrates the method (Merrifield, Guilford, Christensen, and Frick, 1962). Three problem-solving tests were invented. These were included in a battery composed of reference tests for established factors and new tests to measure hypothetical factors. The results were these: No separate problem-solving factor appeared. Each of the problem-solving tests showed a sensible pattern of loadings on established factors. Finally, almost all of the true variance of the problem-solving tests was accounted for. The study supported the notion that complex problem-solving behavior is not wholly different from simpler behavior, but rather can be analysed in terms of more elemental patterns of behavior (though, in this case, behavior that is still rather complex).

A second study from Guilford's laboratory (Frick & Guilford, 1957) will be described, since it involves the water-jar problem, a traditional favorite. A group version of the water-jar test was included in a battery of tests that were intercorrelated and factor analysed. The water-jar test is supposed to measure rigidity-flexibility. As a matter of fact, the test, at least the version employed by Frick and Guilford, showed a loading of .42 on the factor General Reasoning, a loading of .45 on the factor Logical Evaluation, but a loading of only .18 on the factor Adaptive Flexibility. This is striking data. Apparently, less than 4% of the variance of the water-jar test is a function of rigidity-flexibility. A person who developed a training procedure to facilitate performance on the water-jar test by somehow increasing flexibility or eliciting flexible behavior might well come to grief, simply because flexibility is not very important for the solution of the water-jar problem.

Considering the results with the water-jar problem, one cannot help wondering whether performance on other popular problems, such as the hatrack problem or the two-string problem, involves the behaviors each is commonly believed to involve. The two-string problem, in particular, has been used in a number of training studies, studies notable chiefly for being inconsistent with one another (see Maltzman, Belloni, & Fishbein, 1964, p. 6). One of the difficulties is that the problem is psychometrically inadequate (Ray, 1955; Duncan, 1959; Anderson & Anderson, 1963). Could it also be the case that the behavior the training procedures have been designed to arrange or evoke plays a relatively unimportant role in the solution of the two-string problem? Perhaps such a state of affairs would be revealed by an analysis of the two-string problem like the analysis Frick and Guilford made of the water-jar problem.

Conry (1965) and Walsh (1963) have studied the relationship between aptitude measures and performance on the type of problem-solving task in which the subject selects instances until he can name correctly a concept. Both investigations employed the technique of canonical correlation. The present author feels that this technique is inappropriate when the goal is to understand the criterion task in terms of stable, meaningful dimensions. Nor does the method of including the task to be explained in a factor analysis with the reference tests for established factors seem entirely satisfactory since then there is a confounding of explanans and explanandum. In any event, the Conry (1965) data and the Walsh (1963) data were re-analysed in this way: First, the correlations among the aptitude reference tests were factored. Then, loadings were obtained for the measures of problem-solving performance by extension.

Table 3.1 contains the re-analysis of Conry's data, which was obtained from 94 females enrolled in a psychology course. The first factor, General Reasoning, is the only one that shows really promising relationships with the problem-solving measures. On the whole, the aptitude factors explain disappointingly little of the variance of the problem-solving measures, perhaps for one reason, because the reliabilities of the latter measures were low. Conry was not able to obtain estimates of the reliability of these measures, but some indication can be gleaned from the correlations between the two problems that were employed. The between-problems correlations were .23 and .41 for time to criterion and cards to criterion, respectively. Evidently, the reliability of any of these measures alone is rather low.

The re-analysis of Walsh's (1963) data appears in Table 3.2. Her study was conducted with 53 third graders from a suburban community. The problem-solving measure is a composite that gives equal weight to success or failure on a problem and an estimate of efficiency involving the instances the child selected. Walsh experienced some difficulty with her battery of aptitude tests. Many of the subtests of the California Test of Mental Maturity failed to discriminate well. As a result, these subtests distributed over several weak, hard-to-interpret factors. On the other hand, most of the subtests from the Solving Puzzles Test discriminated very well indeed, maybe part of the reason why these tended to load highly together on the first factor. The one thing the Walsh data does suggest is that Originality

Table 3.1

Varimax Factor Matrix for the Conry (1965) Data[a]

Tests	I General Reasoning	II Verbal Comprehension	III Memory Span	IV Spatial Scanning	V Perceptual Speed	VI Rote Memory	h^2
	Factors						
Predictors							
Object-Number	10	-30	63	10	10	79	76
First and Last Names	-03	24	-01	-05	07	84	77
Vocabulary	16	85	07	05	13	03	77
Advanced Vocabulary	15	90	06	-01	09	-05	85
Logical Reasoning	62	30	-12	-05	05	10	51
Inference	41	36	-03	27	-33	14	51
Nonsense Syllogism	76	09	10	-02	-07	-04	60
Auditory Number Span	-01	04	90	-01	07	-03	81
Auditory Letter Span	-05	07	87	06	07	06	78
Ship Destination	56	-02	-03	45	31	-09	62
Necessary Arithmetic Operations	68	12	-03	13	25	08	57
Finding A's	05	12	19	02	71	15	58
Number Comparisons	-22	12	-32	43	35	23	52
Locations	57	-13	-02	19	51	-12	66
Letter Sets	18	09	-01	17	59	12	44
Map Planning	13	-06	17	84	-08	08	77
Maze Tracing	10	10	-08	69	37	-09	65
Variance	22%	18%	16%	16%	15%	13%	
Criteria[b]							
Time to criterion -- 1st problem	-35	-08	17	-25	-10	03	(23)
Time to criterion -- 2nd problem	-23	-10	-07	-13	-23	-03	(14)
Cards to criterion -- 1st problem	-09	-06	-03	-12	-01	11	(04)
Cards to criterion -- 2nd problem	-21	-12	-20	-03	-09	-03	(11)

[a]Decimal points omitted. Computations began with Conry's (1965) correlation matrix in which the coefficients were rounded to two places. The intercorrelations were analysed by the principal components method with units in the diagonal. Factors with eigenvalues $\geq$ 1.00 were rotated. The factors that were rotated accounted for 66% of the total variance.

[b]Loadings obtained by extension.

Table 3.2

Varimax Factor Matrix for the Walsh (1963) Data[a]

Tests	I Fluency/ Flexibility	II Reasoning I	III Originality	IV Verbal Comprehension	V Reasoning II	VI Number	VII Right-Left Specific	h^2
Predictors								
Immediate recall (California Test of Mental Maturity 1)	10	55	29	00	58	-05	05	74
Delayed recall (CTMM 2)	15	-25	-10	79	25	-11	12	81
Sensing Right and Left (CTMM 3)	09	-02	07	02	14	10	82	72
Manipulation of areas (CTMM 4)	25	60	-16	-17	04	13	28	57
Opposites (CTMM 5)	09	-10	06	08	53	63	-12	72
Similarities (CTMM 6)	07	72	17	09	07	-09	-13	59
Analogies (CTMM 7)	15	11	-07	-06	75	14	19	67
Inference (CTMM 8)	42	05	63	21	17	24	-04	71
Number concepts (CTMM 9)	52	44	07	15	17	28	-04	60
Numerical quantity (CTMM 10)	25	08	08	-17	02	80	28	83
Verbal concepts (CTMM 11)	-06	17	15	79	-25	-00	-07	74
Finding hidden words -- flexibility (Kaya Solving Puzzles Test, Part 1)	67	46	19	-07	-03	21	-13	76
Consequence -- fluency (SPT 2)	77	11	01	13	07	02	43	81
Uses -- fluency (SPT 3)	82	09	10	07	08	07	13	74
Homophones -- organization (SPT 4)	70	03	25	21	14	20	23	71
Stories -- organization (SPT 5)	58	-08	21	-38	31	-34	03	75
Match stick puzzles -- flexibility (SPT 6)	-09	40	61	-16	-10	10	33	70
Finding the rule -- flexibility (SPT 7)	71	18	-02	-07	10	12	-19	69
Doodles -- fluency (SPT 8)	63	-15	28	-39	-02	-03	67	66
Originality (Unusual responses on SPT 2, 3, 5, 8)	33	04	77	-00	01	-06	-02	72
Variance	30%	14%	13%	12%	11%	10%	9%	
Criteria[b]								
Card array problems	-13	13	40	06	06	20	-05	(25)
Switching problems	-11	07	05	-20	-01	28	-06	(14)
Chemical problems	21	-07	35	03	17	15	02	(22)

[a] Decimal points omitted. Computations began with Walsh's (1963) raw data. Intercorrelations were factored by the principal components method with units in the diagonal. Factors with eigenvalues $\geq$ 1.00 were rotated. The factors rotated accounted for 71% of the total variance.

[b] Loadings obtained by extension.

may be a factor in the performance of children on the kind of problem in which the subject selects instances until he can name a concept.

Having now reviewed some of the data bearing upon the factor structure of problem-solving tasks, let us take a closer look at the question of the value, to the psychologist interested in developing a training procedure, of a description of a task in terms of loadings on reference factors. Of course, one could not get very far by merely pondering the names of the factors with which a task is saturated. An intimate acquaintance with the skills involved in performance on the tests that define a factor is surely a pre-requisite for making much sense of a task in terms of factor loadings. Still, even with this proviso, the result of such analysis is likely to be a recipe of the form, for instance, two parts Verbal Comprehension, one part General Reasoning, a dash of Ideational Fluency, and a pinch of Semantic Spontaneous Flexibility.

Description in terms of factor loadings seems to provide a sort of <u>taxonomic</u> task analysis. But it does not tell how to put the ingredients in the recipe together in order to get a desired, complex skill. The training psychologist needs a <u>functional</u> task analysis that deals with intratask relationships among component skills and the order in which component skills are to be performed.

Thus far the presumption has been that an adequate factor analytic investigation can at least give, at its own peculiar level of discourse, a complete list of the essential component skills, so to speak, the necessary "ingredients" of a complex skill. However, the fact that a task loads on several factors may very well mean that the criterion task involves alternative component skills rather than that each of the skills represented by the several factors is important. Referring to the factor description of the water-jar problem, it may be with respect to any single individual that the skill involved in performance on the problem is General Reasoning or Logical Evaluation or Adaptive Flexibility. The factor recipe does not seem to tell the cook which ingredients are essential and which can be substituted for one another.

A description of a task in terms of factor loadings can tell what characteristics differentiate the performance of individuals on the task. But the factors that differentiate individuals are not necessarily in one-to-one correspondence with the factors that are relevant to performance on a task. The relevant behaviors that all of the members of a sample manifest or manifest in a sufficient degree cannot be revealed by correlational methods. By the same line of argument, the relative importance of various components of a complex task is not necessarily to be found in the magnitude of the loadings of the task on reference factors.

Finally, there is the problem of naturalistic bias in analysing tasks in terms of factor loadings. Such a procedure indicates which skills differentiate the performance of persons who, usually, have no special preparation for the task at hand and who work at the task for a short period of time. It seems reasonable to suppose that especially qualified persons -- mathematicians, operations researchers, psychologists employing functional-behavioral methods of analysis, depending upon the kind of problem -- with ample time for study can develop approaches to the solution of classes of problems that are not inferior to those inferred from the behavior emitted by students from Psychology 100 when briefly confronted with the task. Perhaps it is idle to inquire -- with either multivariate investigations or with process-tracing studies -- whether or to what degree the components of "ideal strategies" are exhibited in the behavior of college sophmores or Air Force recruits. When a conception of an "ideal strategy" is available it often should be possible to develop training procedures that systematically arrange the required skills. If and when a high level of proficiency can be produced, the task will no longer be called a problem and it may even be said that the task does not require "real thinking" or that it can be performed by "rote". But, then, what is the point of this game anyway?

The conclusion is that under some circumstances description of a complex task in terms of loadings on reference factors might have some value as an adjunct to task analysis. However, from the perspective of the psychologist concerned with training, it is inconceivable that factor descriptions could bear the principal burden of task analysis.

Aptitude Factors and the Effects of Training

This section will examine investigations of the relationships of factor structure and training. This was a popular topic in the 1930's. In recent years there are Fleishman's (e.g., Fleishman & Hempel, 1954; Fleishman, 1957) studies of the changes with practice in the factor structure of complex psychomotor tasks. Stake (1961) has completed the most substantial, recent investigation examining conceptual learning, an investigation involving training on 12 tasks and an extensive battery of reference tests.

Though each has employed factor analysis, there have been distinct differences in methodology among these studies, partly as a function of the particular issue that interested the investigator. In terms of the interests of this paper, Fleishman's method is among the most heuristic. He broke performance up into intervals and as a result was able to study the changing characteristics of performance interval by interval as training progressed. Fleishman's work indicates that there are systematic "changes in the patterns of abilities contributing to proficiency on complex tasks as training continues and proficiency increases" (1957, p. 271). In other words, qualitatively different skills, not merely different degrees of the same skill, seemed to be involved as proficiency on a task increases. There is a task-specific factor, the importance of which usually increases as training continues. At the same time the amount of variance explained by reference tests usually shows a regular decline as training progresses. It is not known whether this latter fact is inevitable or whether it merely reflects the current state of the art of testing and the psychology of individual differences. Of special interest is Fleishman's finding that tests of "higher processes" are important mainly in the early stages of performance. These diminish in importance as training advances.

Though there may be some, to the author's knowledge, none of the studies of the effects of training and factor structure in the so-called "cognitive domain" have employed exactly the same methodology as Fleishman. There have been some interesting investigations though. A study of Woodrow's (1938) is especially provocative. Subjects practiced about ten minutes a day on each of seven tasks for 39 days. Three scores were obtained for each task, initial performance, final performance, and gain, that is the difference between initial and final performance. The intercorrelations of these scores, scores on two intelligence tests, and scores on several tests given just once were factored using the centroid method. The discussion that follows has reference to the varimax (Kaiser, 1958) rotation of Woodrow's centroid matrix which appears in Table 3.3.

There is an extremely interesting generalization that emerges from Woodrow's data: Terminal performance was related to both initial performance -- that is, entering behavior -- and gain or improvement. However, and this is the fascinating point, initial performance and gain were remarkably independent of one another on all but one of these tasks (spot patterns). Notice that terminal performance for any task tends to load on two factors, one of which is saturated with initial performance and the other of which exhibits a high loading on gain or improvement with practice. Final performance showed a more pronounced relationship with initial performance in some cases but with gain in others. On the one task that could be said to involve problem-solving, the anagrams task, final performance was much more highly related to initial performance than to gain. Speaking generally, one would suppose terminal performance on many tasks employed in laboratory

Table 3.3

Varimax Rotation of Woodrow's (1938) Centroid Matrix

			I	II	III	IV	V	VI	VII	VIII	IX
Horizontal adding	Initial	1	27	32	09	42	-01	35	30	15	-15
	Final	2	12	22	10	30	06	27	80	16	-11
	Gain	3	-02	03	09	11	12	10	80	08	04
Substitution	Initial	4	06	75	16	-05	08	03	08	12	-08
	Final	5	-06	53	18	-06	24	72	09	19	[illegible]
	Gain	6	-08	05	07	-07	23	70	10	13	06
Spots	Initial	7	47	35	49	14	-12	04	05	-09	09
	Final	8	20	19	92	04	-12	11	07	-02	05
	Gain	9	-13	05	68	-03	-03	06	10	14	03
Multiple cancellation	Initial	10	07	72	05	29	21	-04	-13	-05	05
	Final	11	18	48	-04	05	71	26	07	-16	27
	Gain	12	33	07	-04	-12	87	26	15	-08	-08
Relative percent length	Initial	13	14	27	48	17	-02	-52	-01	38	16
	Final	14	09	17	24	03	06	-19	00	[illegible]	85
	Gain	15	-06	-05	-21	-10	-08	27	-05	[illegible]	54
Speed, gates	Initial	16	31	53	-12	13	-08	-05	38	-44	-00
	Final	17	30	69	05	-11	06	10	24	29	15
	Gain	18	-04	05	22	-28	-07	19	-02	73	-04
Anagrams	Initial	19	26	07	07	90	-04	-07	-02	-11	-01
	Final	20	23	08	08	72	-02	-14	30	-21	-03
	Gain	21	-02	01	05	-05	-01	-05	35	-17	-00
Artificial language (A, before practice)		22	76	16	-03	12	06	-05	02	02	-01
Artificial language (B, after practice)		23	70	16	27	12	25	01	-01	-07	-17
Form analogies, A		24	73	08	11	13	09	19	03	21	17
Form analogies, B		25	63	26	26	-10	06	-03	11	08	26
Verbal analogies, A		26	86	-00	-05	16	-04	-08	09	-17	-03
Verbal analogies, B		27	83	08	-09	09	-10	-12	-05	-07	-04
Thorndike CAVD		28	69	-09	05	36	25	09	07	18	23
Average 6 Otis Form A and 6 Otis Form B		29	90	04	13	13	18	06	-04	07	04
Categories		30	79	13	00	-09	14	-20	02	-26	-24
Cancellation, 3-digit		31	12	54	-13	07	56	11	-00	26	-12
Arithmetical problems		32	28	-10	49	24	23	-11	11	12	-11
Speed, making crosses		33	10	66	-01	06	00	11	14	-31	14

Note: -- Decimal points have been omitted.

studies of learning would have a relatively slight relation to initial performance, indeed, such tasks are frequently selected precisely because they de-emphasize the role of entering behavior. On the other hand, it is to be expected that initial performance will exhibit a hefty association with final performance on many tasks that confront children in school; for culturally relevant tasks, a brief period of training weighs small in proportion to a long history of learning.

Notice that on Woodrow's first factor (see Table 3), which can be called g, there is a tendency for final performance on most of the tasks to have a smaller loading than initial performance. Though the differences are not impressive, the trend in these loadings is consistent with the view that the skills represented in intelligence tests are less important at the end of training than at the beginning, a result not unlike Fleishman's.

The main conclusion that can be gleaned from the investigations of the relationships between factor structure and training is that aptitude tests are very truly measures of entering behavior. As Woodrow (see 1946) repeatedly argued and rather dramatically demonstrated, these tests do not measure "ability to learn."[2] In the case of complex behavior and complex training procedures -- let us say, an auto-instructional program to teach

---

[2] Stake (1961) who has identified ability to learn with speed of learning rather than amount of improvement, found some weak speed of learning factors. It is not inconceivable that additional research with existing tests, or research with new tests might reveal some associations with learning (either speed or amount of improvement) but, still, the conclusion that for the most part aptitude tests measure entering behavior seems sound.

a school subject -- it is probably true that relevant, previously learned behavior "enters" at various points during instruction, not just at the beginning.

Gagné (Gagné, 1962; Gagné and Paradise, 1961) has also identified aptitude tests with entering behavior; however, his analysis suggests that the behavior reflected in basic aptitude tests "enters" -- that is to say, is relevant -- mainly during the initial stages of instruction, and is decreasingly relevant thereafter. The present author's suspicion is that, even when a complex skill is analysed recursively into a hierarchy of subskills using the technique that Gagné has expounded, various previously learned skills will become more or less relevant in a shifting pattern as learning progresses. Such shifting patterns were revealed in Fleishman's studies, even though the psychomotor tasks employed in these studies were relatively less complex than the academic tasks that Gagné talks about, and the training procedures used more homogeneous than the instruction required to teach academic skills.

Earlier, it was suggested that there is a naturalistic bias in task analyses in terms of factor loadings. There appears to be another facet to this naturalistic bias. Training, some of the time at least, perhaps most of the time, has the effect of changing the pattern of factor loadings. The magnitude of the change presumably depends upon the effectiveness and the extent of training and the degree to which previously acquired skills are relevant to the task at hand. Nonetheless, one can speculate, the <u>direction</u> of the effects of training is generally toward a shift in factor pattern, in other words, a change in the relevance of previously-learned skills. If this is true, then factor analyses of measures on naive subjects could yield only an imperfect description of the component skills involved in highly proficient performance following a period of training.

We are accustomed to speak of an "increase in proficiency" on a task, but such an expression permits the inference that the same skills comprise novice-level performance and master-level performance, that the two levels and intervening levels differ only in quantitative degree. There are both analytic and empirical grounds for resisting such an inference. It is probable that different levels of proficiency on complex tasks, if not simple, homogeneous tasks, usually involve qualitatively different skills. Similarly, even if by some external standard of proficiency, two groups -- one with training of a certain sort, the other without -- happen to perform at the same level, it cannot be assumed that the same skills are involved. In fact, it is very likely that the skills are different.

To put the whole matter another way, behaviors bearing a certain task or function name can show markedly different topographies. From some frames of reference, these differences in topography are unimportant. When the issue is judging what an organism will be able to do in the future on more complex tasks, based on what it can do now; or the issue is planning a training procedure that will most efficiently build upon the organism's current repertoire of behavior; or the issue is analysing a complex skill into simpler, component skills, topography can be very critical, indeed.

Interactions of Aptitude Factors and Training Variables

Stolurow (1965) has recently reviewed studies showing interactions between ability and techniques employed in self-instructional programs. The evidence reviewed by Stolurow suggests that programs featuring knowledge of results, overt responding, and immediate feedback make a difference with low ability students, but that high ability students do just as well on programs without these features. Other studies seem to indicate that programs allowing "self-direction" are superior to linear programs for bright students while slow students do as well (badly) with either kind. It should be noted that frequently the apparent interactions, usually in the form of correlations of different magnitudes between an ability measure and a criterion performance measure, are discovered in *ex post facto* analyses. Still, considering the fact that most psychologists concerned with learning and training have not been very concerned with individual differences, the facts available suggest that rather often there are interactions between ability and training variables.

Studies involving several training conditions in which scores were obtained on a differentiated battery of aptitude tests are scarce (though see Dick, 1963). The whole thrust of the analysis presented on the preceding pages is to suggest that interactions between aptitude factors and training variables are likely to be the rule rather than the exception, at least for complex tasks and complex training procedures, such as programs to teach academic skills.

Of what worth to the training psychologist is it to know that there is an interaction between training conditions and an ability measure? Let it be assumed that the interaction is manifested in differences among training conditions in the degree of correlation between the ability measure and the criterion performance measure. Is a high correlation a favorable sign? According to some educators, the answer is yes. It is a part of the educational folklore that there is an increase in the variance of the achievement scores of students of the good teacher, since he stimulates bright pupils to forge ahead to a greater degree than does the mediocre teacher. The popular answer of persons in programed instruction is no, a high correlation between an ability measure and a performance measure is not a favorable sign. The reasoning is that with effective instruction even the dull student will achieve a high level of performance. The fact of the matter is that in and of itself the magnitude of the correlation between a measure of ability and performance is neither favorable nor unfavorable. It all depends on the means. When the mean level of achievement under a certain training condition is higher than the level under a standard condition, that is good; when the mean of the training condition is lower that is bad. The difference in the correlation of the ability measure and the achievement measure under the training condition and under the standard condition and the difference in mean achievement under the two conditions can, provided several assumptions hold, suggest the kind of student who has profited most or least from training. If mean achievement is higher and the correlation between achievement and ability is higher, then probably the dull student did poorly as usual while the bright student did especially well. This is the case many educators who deal with conventional forms of instruction feel is typical (desirable?). If mean achievement is higher but the correlation of achievement and ability is lower under the training condition than under the standard condition, then probably the low-ability student profited especially and the high-ability student did as well as usual. This latter case is often seen in programed instruction studies.

The remarks that have just been set forth apply as well to general ability tests and differentiated aptitude measures. A second aspect of the issue has particular reference to differentiated aptitudes. There are those who say that if following a certain training regimen, achievement scores show a relatively high correlation with some special ability, this proves that the instruction requires reasoning, creativity, or whatever (see Braund & Heath, 1965) and, indeed, it does. It proves that the regimen makes salient certain skills. But it does not prove that this training procedure is to be preferred to some other. It is entirely possible that another training procedure that led to lower correlations of measures of reasoning or creativity with final performance, would result in higher levels of performance on problem-solving tasks involving reasoning or creativity.

Several of the foregoing points are illustrated by the data in Table 4. Thirty first-grade children received training designed to teach them a problem-solving skill, the skill of attaining a concept or solving a problem by varying each factor in succession while holding all other factors constant. The training used the techniques of programed instruction, though it took the form of a script used by a human teacher (with one child at a time) rather than a self-instructional text or a teaching machine program. In the initial phase of training, the program was divided into seven units called "games" that were designed to teach components of the desired terminal behavior. The first three games arranged "conclusion-drawing behavior" and brought this behavior under the appropriate stimulus control. The remaining four games arranged "instance-selection behavior" and integrated conclusion drawing with instance selection. The second phase of training consisted of an abbreviated form of the program with each of five widely different, additional tasks. The purpose of the second phase of instruction was to vanish the control of particular task and problem characteristics and bring the behavior under the control of the relevant abstract or schematic attributes of tasks and problems.

Following training, the training group and a control group that received no treatment of any kind, were presented with a series of problems, some from tasks that had been employed during training in order to assess retention, and some from tasks new to all children to assess transfer. The problems were exactly like those employed in the Walsh and the Conry studies previously discussed. For each problem, the child began with a stimulus configuration which was, he was told or shown, a positive instance of a concept (e.g., it "shows the secret"). The child was instructed to select or create instances (e.g., "pick cards," "mix chemicals") until he could say the concept. The child was scored as having solved a problem if the instances he selected implied the concept he stated and no other. A further description of the tasks and the method, as well as the main results of the study, can be found elsewhere (Anderson, 1961).

Four hundred and eight first graders, including the 60 children who were presented the problem-solving tasks, completed the California Test of Mental Maturity (Long Form, 1957 Edition). Subtest scores were intercorrelated and factor analysed by the principal components method. Four factors were extracted and rotated using the varimax (Kaiser, 1958) procedure (see Table 3.5). Factor scores were computed for the 60 children who received problems. Each task was represented by a composite problem-solving measure, the same measure as was used with the Walsh data previously discussed.[3]

Table 3.4 contains the correlations of the aptitude factors with the problem-solving measures. It should be noted in passing that the sample was rather homogeneous with respect to IQ and SES. Perhaps stronger relationships would have been observed with a more heterogeneous sample. Though there is detail in the table that looks interesting, remarks will be limited to two observations. First, notice that over-all the aptitude measures account

---

[3]The composite score consisted of the sum of the standardized scores for number of unnecessary trials (reflected) and number of solutions. The measures were standardized separately for the training group and the control group. The median correlation of unnecessary trials and solutions for the various tasks was .58, a coefficient high enough to support the belief that the two scores represented the same thing.

Table 3.4

Correlations of Problem-Solving Performance and Aptitude Factor Scores for Groups of First Graders that Did and Did Not Receive Training (N=30 in Each Group)

Task	Factor score: I Memory Training	I Memory Control	II Reasoning Training	II Reasoning Control	III Number/Space Training	III Number/Space Control	IV Verbal Training	IV Verbal Control	"h^2" Training	"h^2" Control	Reliability[a] Training	Reliability[a] Control
Retention												
*Cowboys	35	-01	04	19	-17	35	34	19	27	19	56	56
*Pegboard	39	03	41	37	42	42	24	38	55	46	64	54
Transfer												
*Screws	14	-04	47	-13	14	-05	22	05	31	02	55	47
*Suburbia	28	-04	08	12	28	00	33	49	27	26	45	62
Pendulum	06	05	22	-07	03	13	00	-06	05	03	--	--
Chemical	08	22	54	-01	37	18	07	24	44	14	60	13
*Pencils	14	12	18	35	12	08	51	16	33	17	55	60
Mean	22	05	29	12	18	17	25	21	33	19	53	50

Note: -- Decimal points have been omitted. Training group - control group correlation differences $\geq .26$ are significant at the .05 level (two-tailed test).

*Task on which the mean performance of the training group is significantly higher ($P < .01$) than the mean performance of the control group.

[a]First problem -- second problem correlation extended with the Spearman-Brown formula. Only one pendulum problem was presented.

Table 3.5

Varimax Factor Matrix for the California Test of Mental Maturity with a Sample of First Graders (N = 408)

		I	II	III	IV	h^2
Immediate recall	1	47	00	19	-07	27
Delayed recall	2	66	-01	-11	01	45
Sensing right and left	3	05	-15	60	12	40
Manipulation of areas	4	-17	62	03	01	41
Opposites	5	02	36	23	-07	20
Similarities	6	07	57	-17	12	37
Analogies	7	25	31	-08	-10	18
Inference	8	25	12	18	-51	37
Number concepts	9	-12	17	26	57	43
Numerical quantity	10	-01	07	63	-05	41
Verbal concepts	11	40	-04	-05	60	52
Variance		25%	25%	25%	25%	

[a]Decimal points in the body of the table omitted. Intercorrelations were factored by the principal components method with the highest correlation of each test as the communality estimate. The four factors that were rotated accounted for 96% of the common variance.

for more of the variance of the training group than of the control group. Since the level of performance of the training group was higher than that of the control and since, except for the chemical problems, the reliability of the problem-solving measure was about the same for the two groups, the correlations suggest that this is a case in which training benefited the bright student most. To look at the issue in another way, the entering behaviors or skills represented by the aptitude factors were engaged or made salient by the training procedure. Second, the training seemed to change the pattern of relevant abilities somewhat. Memory was a more prominent factor in the problem-solving of those who received training than those who did not. Look particularly at the correlations of memory with performance on cowboys task and the pegboard task, tasks that were employed during training and upon which most of the children reached a rather stringent criterion of mastery.

Whenever an aptitude measure shows a marked correlation with a terminal performance measure, one type of subject is doing less well than another. This information could be used to improve the training procedure. One alternative is to develop different training procedures for the different kinds of people. There are a variety of forms that such an approach could take, ranging from wholly discrete programs to a collection of training segments that could be assembled according to a formula based on a vector of scores for each student (see Stolurow & Davis, 1965).

Summary

To pull together and summarize what has been said in this paper, first from the perspective of the psychologist concerned with developing training procedures to arrange problem-solving skills, a description of tasks in terms of loadings on aptitude reference factors might sometimes prove to be a useful adjunct to other task analysis methods. But there are several reasons why such a technique could not comprise the principal method of task analysis. Second, studies of the relationship between training and factor structure seem to indicate that aptitude tests are primarily measures of entering behavior, for the most part unrelated to how much improvement will result from training. When the task or the training procedure is at all complex, a shifting pattern of relationships between aptitudes and performance on the training task is likely to appear. The shifting pattern indicates the changing relevance as training progresses of the previously-learned skills represented by the aptitude factors. Third, interactions, more expressly correlations, between aptitude measures and performance after training contain information that can be employed to improve training by matching the kind of training to the kind of persom in one of several possible ways or modifying the training procedure so that those with low aptitude scores achieve better. Nonetheless, there is good reason to be suspicious of a judgment about a training procedure when that judgment is based on a correlation coefficient alone. Finally, such data on aptitude factors and problem-solving behavior as are available are consistent with the belief that problem-solving is not essentially different from other varieties of behavior.

The information that achievement correlates highly with an aptitude measure might be used in a second way. Under some circumstances, it might be possible to modify a training procedure so that those with the low aptitude score have a better chance of succeeding. If one is speaking of differentiated aptitude factors, and this had been the presumption in the foregoing, the aptitude factor itself might suggest the sort of modification that would improve the performance of those with low scores. For instance, if Verbal Comprehension is a factor that is strongly associated with achievement, perhaps a simplified vocabulary is indicated, or perhaps in the sections in which verbal complexity is unavoidable more trials or frames are needed. Ordinarily, aptitude tests are used to diagnose the failings of students. The suggestion is to use aptitude tests to diagnose the failings of training procedures. It is difficult to forecast how useful aptitude tests would prove in the diagnosis of the shortcomings of training procedures. It could be that the information obtained from correlations between aptitude measures and performance measures would turn out to be just an expensive, indirect substitute for information obtainable by more direct means.

In this section, the argument has been that interactions between ability measures and training conditions are probably the rule rather than the exception. But, caution was urged in interpreting interactions, especially those that appear in the form of correlation coefficients. Neither the assertion that the correlation between ability measures and terminal performance goes up with effective training, nor the contrary assertion is universally true. The adequacy of a training procedure must be judged by the level of performance attained. Correlations of achievement with ability measures, particularly differentiated ability measures, can suggest the <u>kind</u> of student who profits most or least from a certain training procedure. Based on this information, there are several courses of action that could improve training. When differentiated aptitude measures are employed, the aptitude measure itself might suggest the sort of modification that would improve training, though it remains to be seen whether this suggestion has any actual value.

CHAPTER 4

Task Factors and Aptitude Factors in Children's Problem Solving

The purpose of the research reported in this chapter was to explore the effects of task factors and aptitude factors on children's problem solving. With respect to task factors, the first factor was type of materials--card array, cowboy game, and pencil collection. While these materials were not systematically chosen to represent points in a conceptual space, it was important for the program of research to discover any systematic differences among the materials. The second factor was number of relevant stimulus dimensions. The concept to be attained by the child was defined by either one stimulus dimension (e.g. color), two stimulus dimensions (e.g. color and form), or three stimulus dimensions (e.g. color, form, and number). A third factor was order of presentation of materials, in other words, the effects of practice. Type of materials, number of relevant stimulus dimensions, and order of presentation were factors in a within-subjects analysis of variance design. A fourth factor, the nature of the particular stimuli defining concepts, was nested within Materials and Dimensions. Instead, for instance, of representing 1-dimensional card array problems with a single concept--say, Red--six independent subgroups got 1-dimensional card array problems involving six different concepts.

The second purpose of the research reported in this chapter was to investigate the relationships between aptitude factors and problem solving performance. All of the children in the available population were given a battery of 24 aptitude tests intended to mark 10 established aptitude factors. Scores on the tests were intercorrelated. The correlation matrix was factor analyzed. Then problem solving performance was studied in relation to the factors. From perspective of the behaviorist, aptitude factors (or factor scores) represent the operant level of skills often important for human performance on a wide range of tasks. In other words aptitude factors represent entering behavior. The work of Fleishman with complex psychomotor tasks shows that the pattern of correlations of aptitude factors with performance at various levels of proficiency can yield valuable information about the relevance of different components of entering behavior. A Fleishman type analysis was completed. The rationale for this sort of analysis, along with a discussion of its strengths and particularly its weaknesses, appears in Chapter 3.

Method

Subjects. All of the approximately 210 children in the fourth grade in three elementary schools in a self-contained, midwestern town of 19,000 completed a battery of 24 aptitude tests. A random sample of 144 of these children later received a series of nine concept attainment problems. The sample included both white and non-white and both lower-class and middle-class children.

Procedure. The aptitude tests were administered in the regular classroom by one of six experimenters. The classroom teacher assisted. Problems were administered in individual sessions in such reasonably and private spaces as the schools were able to furnish. Each subject received problems from a single experimenter. The procedure for presenting and scoring problems was detailed in Chapter 2.

<u>Design</u>. Each of 144 fourth graders received nine concept attainment problems in a design which involved four factors. They were (A) Order of presentation of materials, (B) Type of materials, (C) Number of relevant stimulus dimensions, (D) Order of presentation of problems within materials. The design can be diagrammed as follows.

	A1			A2			A3		
Group	D1	D2	D3	D1	D2	D3	D1	D2	D3
1	B_1C_1	B_1C_2	B_1C_3	B_2C_2	B_2C_3	B_2C_1	B_3C_3	B_3C_1	B_3C_2
2	B_2C_2	B_2C_3	B_2C_1	B_3C_3	B_3C_1	B_3C_2	B_1C_1	B_1C_2	B_1C_3
3	B_3C_3	B_3C_1	B_3C_2	B_1C_1	B_1C_2	B_1C_3	B_2C_2	B_2C_3	B_2C_1

Each subject moved from left to right along one of the rows in this table. In this design "Group" is a sequence factor. Order of presentation of problems within materials, which was included for reasons of experimental control and is not of intrinsic interest, was ignored in the data analysis. When this factor is collapsed, there remains a fixed-effect design involving repeated measures which Winer (1962, pp. 571-574) calls Plan 12. There were two additional orthogonal factors which were not included in the main data analysis. The first of these was experimenter, which can be thought of as a replication factor since each experimenter ran 24 subjects in a miniature version of the experiment as a whole. Preliminary one-way analyses of variance indicated no effect for experimenter, hence data were pooled in subsequent analyses. Concept subgroup was the final factor. Independent groups of subjects received unique sets of problems. Concept subgroup could be regarded as a fixed factor nested within Materials and Dimensions. The main reason for including this factor was to broaden the base for generalization. It would be presumptuous to make much of differences among materials when each set of materials was represented by only three of the 27 possible concepts which could be formed within each set. In the present experiment, each set of materials was represented by 18 concepts. Similarly, each level of Number of Relevant Stimulus Dimensions was represented by 18 concepts instead of three. It would have been possible to sample at random from among the concepts that can be formed within each set of materials. Such a strategy, at least by the classical logic of inferential statistics, would have permitted even more confidence regarding the limits of generalization. The author chose the alternative strategy of choosing concepts on the basis of an _a priori_ analysis which can be argued to guarantee that all possible kinds of concepts are represented in known proportions. This would seem to be the only sensible decision. After all we are not dealing with an indefinite number of field plots which will receive rainfall in unpredictably different quantities.

Concept subgroup could have been included as a factor in the main analysis. However, it was not included since the interactions, if any,

ERIC

of this factor with other factors would have been uninteresting and probably uninterpretable. Instead, differences among concepts classified according to materials and number of relevant stimulus dimensions were studied in separate analyses. In the main analysis, scores were pooled across concept subgroup. At first glance it may seem that under this procedure the effects of concept subgroup are being included within error. However, this is not the case. All of the factors of interest involved within-subject error terms.

It goes without saying that subjects were randomly assigned to sequence groups, concept subgroups, and experimenters. Assignment to experimenters was random with the constraint that each experimenter worked in just one school.[1] Randomization was accomplished as follows: (a) The fourth graders within each school who completed all of the aptitude tests were randomly divided into two equal-sized groups. (b) Within each group, the names of the children were listed in random order. (c) The two lists were randomly assigned to the two experimenters who worked within the school. (d) The experimenter was also provided with a randomly-ordered set of "treatment control sheets." Each such sheet specified the sequence of problems that a single subject was to receive. (e) The first child on the list who was not absent received the treatment specified on the first control sheet; the next child received the treatment specified on the second control sheet; and so on.

---

[1]As already explained, experimenter could be considered to be a replication factor. Hence, school was also a replication factor, with experimenter nested within school. While the data were not analyzed by school, the nature of the design assured that differences are not attributable to influences associated with school, whatever they may be.

<u>Aptitude tests</u>. The 24 aptitude tests were chosen to mark ten expected factors. The <u>Manual for Kit of Reference Tests for Cognitive Factors</u> (French, Ekstrom, and Price, 1963) served as the chief basis for expectations about factors, as the guideline for choosing tests, and as the primary source with respect to interpreting and naming factors. The work of Guilford and his associates was also consulted (Guilford, Christensen, and Frick, 1962; Wilson, Guilford, Christensen, and Lewis, 1954).

At least two tests were selected to mark each expected factor. There were four sources of tests. Given below is the abbreviation used to credit the source. (ETS) The <u>Kit of Reference Tests</u> developed by a variety of persons and published by the Educational Testing Service. (PMA) The <u>Primary Mental Abilities</u> battery developed by Thurstone and published by Science Research Associates. (NYU) The <u>Solving Puzzles</u> tests developed by Kaya (1960) at New York University. (TRL) Tests developed by the author and his associates in the Training Research Laboratory at the University of Illinois.

The tests are listed below by name. Included is the source credit, the factor with which the test is presumably saturated, and a brief description of the test.

1. <u>Picture-Number</u>. ETS. Associative Memory. The subject sees pictures of common objects paired with numbers. He memorizes the number which goes with each picture, then has to write down the numbers on a page that has only the pictures on it.

2. <u>First Name and Last Names</u>. ETS. Associative Memory. This is similar to the first test except that the material is all verbal. The student memorizes a list of names and then must write the first names in blank spaces preceding the last names.

3. Subtraction and Multiplication. TRL. Number Facility. This test is nothing more than a series of subtraction and multiplication problems.

4. Spatial Orientation. TRL. Spatial Orientation. This is a typical spatial orientation test in which a sample and four alternatives are provided. The subject must select the alternative which is identical to the sample except for being rotated.

5. Finding 'A's. ETS. Perceptual Speed. There is a long list of words. The subject must cross out as many words containing the letter "a" as he can within the time limit.

6. Words. PMA. Verbal Comprehension. This is a vocabulary test. The subject is required to select the word from among four alternatives which has the same meaning as the sample word.

7. Pictures. PMA. Verbal Comprehension. This is another vocabulary test. The student is shown four pictures and is asked, for example, "Which one is the dog?"

8. Number Sense. PMA. Number Facility and General Reasoning. Included are arithmetic word problems and problems in which the subject must fill in the missing number in a number series.

9. Addition. PMA. Number Facility. This consists of 30 addition problems.

10. Spatial. PMA. Spatial Orientation. The subject is required to select the one figure from among four alternatives which, when combined with the sample figure, will produce a square.

11. Figure Grouping. PMA. Induction. Four figures are presented. The student is required to select the one which is different from the other three.

12. Word Grouping. PMA. Induction. Four words are presented, three of which share some property. The subject must select the odd word.

13. Perceptual Speed. PMA. Perceptual Speed. Four pictures are presented. The student must select the two pictures which are identical.

14. Solving Puzzles 1. NYU. Flexibility. The student is required to find hidden words in a longer word.

15. Solving Puzzles 2 (first scoring). NYU. Ideational Fluency. The score is the total number of answers the subject gives to such questions as "Name all the round things you can think of."

16. Solving Puzzles 2 (second scoring). NYU. Originality. The score here is the proportion of unusual or original answers.

17. Solving Puzzles 3 (first scoring). NYU. Ideational Fluency. This is similar to test 15, except the tests involve the uses of objects, for example, "Write down all the different ways you could use a string."

18. Solving Puzzles 3 (second scoring). NYU. Originality. The score is the proportion of answers to test 17 which are unusual or original.

19. Solving Puzzles 4. NYU. Organization. The subject is required to produce a word which is simultaneously synonymous to two words.

20. Solving Puzzles 5. NYU. Organization. The student is given four words and is required to write a very short story with two or three sentences using all four words.

21. Solving Puzzles 6. NYU. Flexibility. This is the "match-stick puzzles test." There is a figure drawn of match sticks; the directions to the subject are, for example, "Take away two matches and leave three triangles."

22. Solving Puzzles 7. NYU. Flexibility. A series of words is arranged according to some rule; the student must discover and state the rule.

23. Solving Puzzles 8 (first scoring). NYU. Ideational Fluency. A drawing is shown. The score is the number of things the subject lists that he thinks the drawing could be.

24. Solving Puzzles 8 (second scoring). NYU. Originality. The same as test 23, except that the score is proportion of original answers.

Measures of skewness and kurtosis were computed. Seven distributions departed sufficiently from normality to warrant transformations. Square root transformations were made on tests 1, 2, and 21. The scores on tests 12 and 20 were squared. Arcsine transformations were made on two of the three ratios, test 16 and test 18. After these transformations, skewness and kurtosis fell within normal bounds for all tests.

A components analysis was performed. The correlation matrix, with 1's in the diagonal, was analyzed into principal components. Factors, or components, with latent roots greater than 1 were rotated using the varimax procedure.

Task factors and problem solving

Specific concepts. For this phase of the analysis there were three dependent variables: (1) Number of solutions to the problem; (2) Number of inappropriate choices of instances, which is the sum of inefficient, redundant, and repetitious choices; and (3) Number of inappropriate conclusions, which is the sum of inconsistent, consistent-but-invalid, and miscellaneous conclusions. The details of the scoring procedure, it should be mentioned again, and the logic upon which these procedures are based appear in Chapter 2.

Tables 4.1, 4.2, and 4.3 contain the means and standard deviations for the card array, cowboy game, and pencil collection, respectively. Analyses of variance were performed on each set of six means. It should be emphasized that these analyses were not independent. The same six subgroups were involved in each analysis. For what it is worth, there were three significant ($\alpha = .05$) $\underline{F}$ ratios. There were significant differences in proportion of subjects solving the problem among 2- and 3-dimensional card array concepts and 1-dimensional pencil concepts.

Effects of order of presentation, type of materials and number of relevant stimulus dimensions. For this phase of the analysis there were four dependent variables. Summary problem solving performance, a measure discussed in Chapter 2, was included in addition to number of solutions, number of inappropriate choices, and number of inappropriate conclusions. Unlike the latter three measures, which are simply counts of "natural units," the summary performance score is a derivative measure in an arbitrary metric. Nonetheless, the summary score has the advantage of

Table 4.1

Card Array Concept Means and Standard Deviations

Concept	Proportion Solving Problems M	Proportion Solving Problems SD	Number Inappropriate Choices M	Number Inappropriate Choices SD	Number Inappropriate Conclusions M	Number Inappropriate Conclusions SD
			1-Dimensional Problems			
One	.75	.44	1.33	1.17	1.29	1.55
Red	.63	.49	2.00	1.10	1.00	1.47
Box	.88	.34	1.33	1.01	1.38	1.35
Two	.83	.38	1.54	1.18	1.17	1.01
Green	.83	.38	1.67	1.17	1.25	1.11
Diamond	.92	.28	1.25	.90	.92	1.21
F ratio	1.70		1.59		.45	
			2-Dimensional Problems			
Green Diamond	.71	.46	1.00	1.22	.75	1.03
Two Green	.58	.50	1.54	1.44	1.17	1.34
Two Diamond	.88	.34	1.33	1.49	.92	1.38
Red Box	.79	.41	1.17	1.40	.46	.78
One Red	.54	.51	2.04	1.37	1.33	1.31
One Box	.88	.34	1.46	1.10	1.13	1.36
F ratio	2.63*		1.73		1.63	
			3-Dimensional Problems			
One Red Diamond	.58	.50	1.58	1.53	.88	.90
One Green Box	.58	.50	2.00	1.35	.88	1.36
Two Red Box	.67	.48	1.67	1.69	1.17	1.24
Two Green Box	.88	.34	1.21	1.22	.25	.61
Two Red Diamond	.50	.51	1.92	1.50	.96	.81
One Green Diamond	.83	.38	1.63	1.31	1.13	1.33
F ratio	2.57*		.91		2.25	

$^{*}p < .05$

Table 4.2

Cowboy Concept Means and Standard Deviations

Concept	Proportion Solving Problems		Number Inappropriate Choices		Number Inappropriate Conclusions	
	M	SD	M	SD	M	SD
			1-Dimensional Problems			
Riding	.71	.46	1.83	1.17	.88	1.08
No Gun	.75	.44	2.04	1.40	1.00	.88
Hat	.88	.34	1.79	1.61	1.04	1.27
Standing	.71	.46	2.04	1.27	1.21	1.02
Gun	.75	.44	1.63	1.13	1.33	1.37
Bareheaded	.88	.34	1.21	1.22	1.00	1.18
F ratio	.82		1.36		.50	
			2-Dimensional Problems			
Standing Gun	.67	.48	1.96	1.40	1.00	1.02
Standing Bareheaded	.54	.51	2.42	1.32	1.25	1.22
Gun Bareheaded	.75	.44	1.92	1.28	1.58	.88
Riding No Gun	.75	.44	2.00	1.35	1.25	1.19
Riding Hat	.75	.44	1.83	1.55	.88	.80
No Gun Hat	.79	.41	1.71	1.33	1.54	1.02
F ratio	.96		.74		1.79	
			3-Dimensional Problems			
Riding Gun Hat	.75	.44	1.83	1.58	.63	.77
Riding No Gun Bareheaded	.63	.49	2.04	1.46	1.21	1.25
Standing Gun Bareheaded	.75	.44	2.13	1.42	.71	.91
Standing No Gun Bareheaded	.67	.48	1.96	1.52	1.00	1.29
Standing Gun Hat	.67	.48	2.25	1.36	.96	1.08
Riding No Gun Hat	.79	.41	1.71	1.43	.92	1.14
F ratio	.47		.43		.90	

ERIC

Table 4.3

Pencil Concept Means and Standard Deviations

Concept	Proportion Solving Problems M	Proportion Solving Problems SD	Number Inappropriate Choices M	Number Inappropriate Choices SD	Number Inappropriate Conclusions M	Number Inappropriate Conclusions SD
			1-Dimensional Problems			
Long	.92	.28	1.83	1.52	1.21	1.02
No Eraser	.58	.50	2.08	1.28	1.13	1.12
Point	.92	.28	1.42	1.47	1.29	1.20
Short	.75	.44	1.54	1.22	1.17	1.24
Eraser	.83	.38	1.50	1.64	1.29	1.33
Unsharp	.92	.28	1.54	.83	1.13	.99
F ratio	3.10*		.84		.10	
			2-Dimensional Problems			
Short Eraser	.71	.46	1.50	1.32	1.13	1.42
Short Unsharp	.42	.50	2.50	1.47	1.17	1.20
Unsharp Eraser	.75	.44	1.29	1.40	1.17	1.31
Long No Eraser	.79	.41	1.63	1.35	1.00	1.18
Long Point	.71	.46	1.58	1.44	1.25	1.36
Point No Eraser	.71	.46	1.58	1.28	1.42	1.28
F ratio	2.02		2.22		.28	
			3-Dimensional Problems			
Long Point Eraser	.79	.41	1.92	1.47	.96	1.00
Long Unsharp No Eraser	.67	.48	1.75	1.42	.58	.72
Short Point No Eraser	.67	.48	1.79	1.44	1.13	1.08
Short Unsharp No Eraser	.88	.34	1.50	1.41	.71	.86
Short Point Eraser	.75	.44	1.42	1.44	.71	.81
Long Unsharp Eraser	.71	.46	1.75	1.33	.79	1.25
F ratio	.81		.43		.99	

*$p < .05$

representing the various components of problem solving performance with a single number. An $\underline{S}$'s summary score for a problem is determined as follows. Two points are awarded if the $\underline{S}$ solves the problem and makes no inappropriate choices and states no inappropriate conclusions. One point is awarded if the $\underline{S}$ solves the problem but makes one or more inappropriate choices or states one or more inappropriate conclusions. An $\underline{S}$ receives no points if he does not solve the problem. Naturally the four measures are not independent.

Consider first the summary measure. Means and standard deviations appear in Table 4.4. The analysis of variance is described in Table 4.5. Order of presentation was the one significant effect. Subjects averaged .74 per problem on the first set of materials, .95 on the second set, and 1.06 on the final set.

Table 4.6 contains the means and standard deviations of number of solutions to the problems. Since each cell is represented by a single problem which a subject either solved or failed to solve the figures in Table 4.6 are actually proportions of $\underline{S}$s solving each problem. For this reason, incidentally, the means and standard deviations are not independent. As can be seen in Table 4.7, order of presentation and number of relevant stimulus dimensions both had a significant effect on problem soltuions. The average proportion of problems solved was .61 for the first set of materials, .77 for the second set and .83 for the third set. An average proportion of .80 subjects solved the three 1-dimensional problems while .70 solved the 2-dimensional problems and .71 solved the 3-dimensional problems.

Table 4.4

Means and SDs for the Problem Solving Summary Measure

Materials	Number of Dimensions	Order of Presentation: 1st M	1st SD	2nd M	2nd SD	3rd M	3rd SD
	1	.92	.61	.88	.49	.94	.52
Cards	2	.75	.70	1.08	.77	1.13	.67
	3	.63	.73	.98	.79	1.21	.68
	1	.69	.59	.87	.58	1.15	.50
Cowboys	2	.58	.65	.90	.59	1.06	.60
	3	.63	.67	1.10	.69	.98	.63
	1	.81	.53	.90	.52	1.02	.48
Pencils	2	.83	.78	.83	.75	1.00	.62
	3	.83	.69	1.02	.56	1.02	.79

ERIC

Table 4.5

Analysis of Problem Solving Summary Variance

Source	df	MS	F
Between Ss	143		
Group	2	.79	.76
Ss within Groups	141	1.05	
Within Ss	1152		
Order of Presentation (O)	2	11.15	32.39*
Materials (M)	2	.36	1.06
(O X M)'	2	.63	1.84
M X Ss within Groups	282	.34	
Number of Relevant Dimensions (D)	2	.09	.25
D X Group	4	.75	2.21
D X Ss within Groups	282	.34	
O X D	4	.62	1.87
M X D	4	.24	.73
(O X M)' X D	4	.38	1.16
M X D X Ss within Groups	564	.33	
Total	1295		

*p < .01

Table 4.6

Means and SDs of Number of Solutions

Materials	Number of Dimensions	1st M	1st SD	2nd M	2nd SD	3rd M	3rd SD
	1	.77	.42	.81	.39	.83	.38
Cards	2	.60	.49	.75	.44	.83	.38
	3	.48	.50	.69	.47	.85	.36
	1	.63	.49	.77	.43	.94	.24
Cowboys	2	.50	.51	.77	.42	.85	.36
	3	.52	.50	.81	.39	.79	.41
	1	.75	.44	.81	.39	.90	.31
Pencils	2	.60	.49	.63	.49	.81	.39
	3	.67	.48	.85	.36	.71	.46

Order of Presentation: 1st, 2nd, 3rd.

Table 4.7

Analysis of Variance of Number of Solutions

Source	df	MS	F
Between Ss	143		
Group	2	.48	1.12
Ss within Groups	141	.43	
Within Ss	1152		
Order of Presentation (O)	2	5.58	31.55*
Materials (M)	2	.03	.17
(O X M)'	2	.21	1.19
M X Ss within Groups	282		
Number of Relevant Dimensions (D)	2	1.27	8.49*
D X Group	4	.24	1.61
D X Ss within Groups	282	.15	
O X D	4	.27	1.92
M X D	4	.15	1.05
(O X M)' X D	4	.27	1.92
M X D X Ss within Groups	564	.14	
Total	1295		

*$p < .01$

ERIC

Table 4.8 contains means and standard deviations of number of inappropriate choices per problem while Table 4.9 summarizes the analysis of variance for this measure. Once again, order of presentation had a significant effect. The mean number of inappropriate choices per problem was 2.09 for the first set of materials, 1.57 for the second set, and 1.45 for the third set. Type of materials also made a significant difference. The mean number of inappropriate choices per problem for the three cowboy problems was 1.91. The means were 1.53 and 1.67 for the cards and pencils, respectively.

The means and standard deviations for number of inappropriate conclusions appear in Table 4.10. Table 4.11 shows that order of presentation, number of relevant stimulus dimensions, and the interaction of these two factors were significant. Table 4.12 contains the means entailed in these significant effects. The Number of Relevant Stimulus Dimensions X Group interaction was marginally significant, but the means involved in this effect will not be summarized since the result is completely uninterpretable.

Discussion. The one factor which had a strong influence was order of presentation. This factor had a significant effect on each of the four measures. Evidently children grow more proficient at solving problems as their experience with solving problems increases. Facilitation occurred solely as a function of practice. There was no training. Indeed, there was not even feedback as to the correctness of the child's responses, except insofar as feedback was indirectly involved in the procedures for administering problems. Nonetheless, problem solving performance improved steadily from the first to the third set of problems.

Table 4.8

Means and SDs of Number of Inappropriate Choices

Materials	Number of Dimensions	1st M	1st SD	2nd M	2nd SD	3rd M	3rd SD
	1	1.58	1.11	1.58	1.11	1.39	1.11
Cards	2	1.94	1.41	1.25	1.38	1.08	1.16
	3	2.27	1.28	1.56	1.50	1.16	1.33
	1	2.31	1.29	1.77	1.31	1.23	1.13
Cowboys	2	2.54	1.35	1.71	1.29	1.67	1.31
	3	2.54	1.22	1.58	1.50	1.83	1.47
	1	1.69	1.27	1.58	1.30	1.69	1.49
Pencils	2	1.89	1.52	1.67	1.42	1.48	1.27
	3	2.06	1.37	1.45	1.30	1.54	1.49

Table 4.9

Analysis of Variance of Number of Inappropriate Choices

Source	df	MS	F
Between Ss	143		
Group	2	6.44	1.28
Ss within Groups	141	5.04	
Within Ss	1152		
Order of Presentation (O)	2	50.06	35.74*
Materials (M)	2	14.95	10.68*
(O X M)'	2	1.40	1.00
M X Ss within Groups	282	1.40	
Number of Relevant Dimensions (D)	2	2.07	1.57
D X Group	4	2.28	1.73
D X Ss within Groups	282	1.32	
O X D	4	2.78	2.06
M X D	4	1.26	.93
(O X M)' X D	4	1.24	.93
M X D X Ss within Groups	564	1.35	
Total	1295		

*p < .01

Table 4.10

Means and SDs of Number of Inappropriate Conclusions

Materials	Number of Dimensions	Order of Presentation: 1st M	1st SD	2nd M	2nd SD	3rd M	3rd SD
	1	.79	1.05	1.58	1.47	1.13	1.20
Cards	2	1.20	1.32	.88	1.18	.79	1.18
	3	1.10	1.34	1.02	1.14	.50	.62
	1	1.02	1.00	1.55	1.40	.69	.78
Cowboys	2	1.50	1.17	1.25	.96	1.00	.97
	3	1.25	1.31	.75	.98	.71	.85
	1	1.23	1.15	1.56	1.24	.81	.89
Pencils	2	1.38	1.27	1.48	1.44	.71	.97
	3	.96	1.11	.81	.94	.67	.83

Table 4.11

Analysis of Variance of Number of Inappropriate Conclusions

Source	df	MS	F
Between Ss	143		
Group	2	.03	.01
Ss within Groups	141	2.52	
Within Ss	1152		
Order of Presentation (O)	2	23.86	21.39**
Materials (M)	2	.75	.67
(O X M)'	2	2.01	1.80
M X Ss within Groups	282	1.12	
Number of Relevant Dimensions (D)	2	11.05	8.79**
D X Group	4	3.18	2.53*
D X Ss within Groups	282	1.26	
O X D	4	6.79	6.70**
M X D	4	1.78	1.75
(O X M)' X D	4	1.22	1.20
M X D X Ss within Groups	564	1.01	
Total	1295		

*$p < .05$

**$p < .01$

Table 4.12

Mean Number of Inappropriate Conclusions Per Problem as a Function of Order of Presentation and Number of Relevant Stimulus Dimensions

Number of Dimensions	Order 1	2	3	M
1	1.01	1.56	.88	1.15
2	1.36	1.20	.83	1.13
3	1.10	.86	.63	.86
M	1.16	1.21	.78	

ERIC

Research with first graders reported in Chapter 6 showed that there is a definite limit to the amount of improvement which will result from unguided practice alone. While it is impossible to estimate how close to asymptote the fourth graders in the present experiment were by the end of their third set of three problems, it is worth noting that performance gain was negatively accelerated. That is, with respect to the summary measure, number of solutions, and number of inappropriate choices (but not number of inappropriate conclusions) there was substantially greater improvement from the first to the second set of problems than from the second to the third.

Tentatively, the significant effect of number of relevant stimulus dimensions on number of solutions and on number of inappropriate conclusions can be attributed to the control which positive instances exercise over conclusion drawing behavior and conclusion stating behavior. Subjects stated fewer inappropriate conclusions on 3-dimensional problems than they did on 1- and 2-dimensional problems. And, fewer 3-dimensional problems than 1- or 2-dimensional problems were solved. When an S solves a problem he must by definition state a conclusion. Thus, less conclusion-stating behavior, either appropriate or inappropriate, was emitted on 3-dimensional problems than on 1- and 2-dimensional problems. In a 3-dimensional problem formed with a set of materials representing three stimulus dimensions each of which takes on two values, there is only one positive instance, the focus instance designated by E. None of the instances chosen by S in the course of work on a 3-dimensional problem could be positive. The argument is that Ss tend to draw conclusions by naming the common elements of positive instances. When no positive instances are encountered, conclusion stating behavior is inhibited.

The effect of type of materials on number of inappropriate choices is attributable to the greater frequency of such choices with cowboy problems. The author and his associates have often observed that the story line employed with the cowboy materials interferes with problem solving behavior. A child will insist that a "friend of the sheriff" must ride a horse or must carry a rifle. Possibly the thematic characteristics of the cowboy task led to interference; though if this is the case it is not clear why number of inappropriate choices was the only measure affected.

Aptitude factors and problem solving

Interpretation of factors. The correlation matrix appears in Appendix II. Table 4.13 contains the rotated factor loadings. The seven rotated factors accounted for 63.4% of the total variance. The communalities of the individual tests and an estimate, not always an appropriate one, of the reliabilities of the tests appear in Table 4.14. The interpretation of factors is given below.

I. Verbal Comprehension. All of the tests loading highly on this factor have verbal content. It is clearly Verbal Comprehension.

II. Speed. This factor is difficult to interpret. Loaded highly are the tests which were expected to define Associative Memory and Number Facility as well as Perceptual Speed. Speed seems to be the common element. First of all, the two Perceptual Speed tests (tests #5 and #13)

Table 4.13

Varimax Factor Matrix

	I	II	III	IV	V	VI	VII
1	.17	.75	-.05	.09	.31	-.04	-.06
2	.65	.46	.21	.09	.11	-.07	-.01
3	.29	.53	.41	.07	.18	.29	.17
4	.15	.17	.25	.12	.71	-.07	-.07
5	.11	.65	.00	.22	-.07	.07	.09
6	.79	.21	.23	.11	.09	.11	.09
7	.60	-.01	.00	.00	.45	.09	.14
8	.39	.22	.41	-.03	.28	.32	.28
9	.17	.43	.24	.16	.13	.36	.24
10	.16	.23	-.02	.05	.70	.27	.13
11	.23	.09	.48	-.12	.29	.45	.06
12	.56	.36	.44	-.01	.12	.19	.07
13	.32	.42	.23	.13	.19	.35	.15
14	.58	.18	.23	.27	.19	.31	.12
15	.26	.11	.35	.68	-.10	.09	.05
16	.04	-.09	.68	.26	.12	-.06	-.02
17	.15	.23	.17	.70	.18	.11	.08
18	.02	-.21	.11	.27	.46	-.14	.43
19	.74	.06	-.00	.20	.03	.15	-.02
20	.22	.16	.62	.19	-.03	-.03	.31
21	.16	.05	-.09	.18	-.03	.84	-.08
22	.46	.13	.01	.10	-.08	.05	.60
23	.06	.12	-.01	.74	.11	.06	.18
24	-.04	.09	.15	.14	.11	.02	.81
Percent Common Variance	23.0	15.7	14.0	13.2	12.5	10.9	10.6

Table 4.14

Estimated Reliabilities and Communalities of the Twenty-four Aptitude Tests

Test Number	Procedure for Estimating Reliability*	Estimated Reliability	Communalities
1	2	.56	.71
2	2	.82	.71
3	2	.90	.69
4	1	.55	.64
5	2	.70	.51
6	1	.89	.76
7	1	.65	.59
8	1	.82	.63
9	1	.93	.50
10	3	.86	.65
11	3	.79	.60
12	3	.88	.69
13	1	.92**	.52
14	1	.94	.64
15	4	.72	.69
16	4	.37	.56
17	1	.77	.65
18	1	.54	.55
19	1	.68	.62
20	4	.79	.59
21	4	.66	.77
22	1	.71	.60
23	4	.87	.61
24	4	.51	.72

*There were four computational procedures:

(1) odd-even correlation stepped up by Spearman-Brown formula,
(2) parallel form correlation stepped up by Spearman-Brown formula,
(3) Horst formula for unequal length parts,
(4) items were intercorrelated, then r to z (Fisher) transformations were made for each unique correlation. The average z was transformed to r. The Spearman-Brown formula was then applied to the r.

**This is to be regarded as an inaccurate upper bound estimate. The misapplication of a formula by Gulliksen gives an inaccurate lower bound estimate of .25.

did load highly. Second, it is known that tests involving routine arithmetic problems sometimes do load on the Perceptual Speed factor rather than defining a separate factor. With regard to tests #1 and #2, which supposedly represented Associative Memory, these were intended for adults but were used with children in the present study. Scores were low. It is not unreasonable to suppose that those who scored high did so because they worked quickly.

III. Reasoning. This factor is called Reasoning because of the loadings of tests #3, #8, #11, and #12; however, the high loadings of tests #16 and #20 make its status ambiguous.

IV. Ideational Fluency. All three tests of ideational fluency loaded highly on Factor IV.

V. Spatial Orientation. The two tests with the highest loadings on this factor were both tests of spatial orientation. The other tests with moderate to high loadings all had figural as opposed to verbal content.

VI. Figural Adaptive Flexibility. This factor was so named because the one test with a very high loading was the match stick puzzles test.

VII. Originality. All three originality scales loaded on this factor.

Relationship of aptitude factors to problem solving performance. As a service to the cooperating schools aptitude factor scores, correctly computed (Glass and Maguire, 1966), were furnished in percentile rank form (Willis and Anderson, 1965). These scores were intercorrelated to be sure that the conversion to percentile ranks had not introduced noise. All of the entries were near zero.

Next, correlations between factor scores and problem solving performance were computed. Relationships with all of the problem solving measures were studied, but only the analysis involving the summary measure will be reported here. This is logically the single best representative of problem solving performance, the most comprehensive measure, and probably the most reliable measure. Then too, as will become apparent in a moment, nothing very striking was discovered, so there was no point to including all the figures.

The fact that different subjects received different problems and the fact that many of those who received the same problems received them in different orders were potentially serious difficulties for the correlational phase of the analysis. The situation is comparable to that which would exist if several different forms of an achievement test were given. If the data were simply pooled without regard to test form, the correlations of achievement with other variables might be attenuated. In order to gauge the extent of the attenuation, both within-group correlations and pooled-data correlations were computed.

Number of relevant dimensions was collapsed. Each S then had three scores, one for his first three problems, one for his second three problems, and one for his third three problems. For the computation of the within-group correlation, the Ss were then divided into the 18 groups defined by Sequence Group and Problem Subgroup. The variance-covariance matrix, involving the seven aptitude factor scores and the three problem solving summary scores, was computed for each group. The matrices were summed, and the within-group correlations were computed from the total matrix. Table 4.15A contains the within-group correlations between the aptitude factors and problem-solving performance. Pooled-data correlations appear in Table 4.15B. Included is the multiple correlation, R, of the aptitude factor scores with the measure of problem solving performance. The reliability estimate, rr, is the highest correlation that a given task showed with the other two tasks.

None of the correlations appearing in Tables 4.15A and 4.15B is very large. The apparent difficulty is the lack of reliability of the problem solving criterion. An alternative possibility is that the aptitude factors interact with type of materials or number of relevant stimulus dimensions, two factors which were collapsed for the correlational analysis. If, for example, a certain factor were positively correlated with performance on 1-dimensional problems but negatively correlated with performance on 3-dimensional problems the net relationship would be near zero. The data were reanalyzed by task and by number of relevant stimulus dimensions. There were no indications of interactions between these task factors and the aptitude factors.

It does seem, then, that at least one reason for small relationships was the low reliability of the problem solving criterion. Anderson (1964) and Walsh (1963), in studies reviewed in Chapter 3, reported higher reliabilities for a summary measure of problem solving than were obtained here. In both of the previous studies the child performed on two problems involving the same number of stimulus dimensions and the same materials at one sitting. Perhaps this made the difference. Also, the summary

Table 4.15A

Relationships Between Aptitude Factors and Problem Solving Performance. A. Within-Group Correlations

Problem Solving Performance	Factor								
	I	II	III	IV	V	VI	VII	R	rr
First task	.03	.03	.06	-.07	.28	.24	-.09	.39	.38
Second task	.06	.16	.00	-.04	.19	.29	.05	.39	.47
Third task	.07	.01	.11	.18	.16	.25	.19	.41	.47

Table 4.15B

Relationships Between Aptitude Factors and Problem Solving Performance. B. Pooled-data Correlations

Problem Solving Performance	Factor								
	I	II	III	IV	V	VI	VII	R	rr
First task	-.02	.01	.10	-.09	.22	.24	-.03	.35	.39
Second task	.04	.17	.04	-.04	.16	.23	.12	.36	.46
Third task	.02	.08	.09	.13	.13	.23	.22	.39	.46

ERIC

measure employed in the correlational analysis of the data from the previous studies was slightly different and, perhaps, somewhat more reliable than the one used in this study.

Two factors showed modest but significant correlations with problem solving performance, namely Spatial Orientation and Figural Adaptive Flexibility. Both of these factors entail skill at dealing with spatial or figural as opposed to verbal material. In a typical test of Spatial Orientation, the S must select the one alternative from among several which is identical to a sample except for a clockwise or counterclockwise rotation. None of the problems involved precisely this skill; however, if the analysis presented in Chapter 1 is correct, a perhaps related skill is important for solving concept attainment problems. When selecting instances and/or when drawing conclusions, the S must discriminate a set of instances or potential instances in terms of one stimulus dimension while ignoring variations along other stimulus dimensions.

Figural Adaptive Flexibility, the second factor to show moderate correlations with problem solving performance, was defined in the present study primarily by a match stick puzzles test. On this test, the S is presented with complex configurations and asked, for example, to "take away two sticks and leave three triangles." Aside from the fact that both involve spatial configurations, the match stick puzzles test and the concept attainment problems do not share obvious characteristics.

A relationship between Memory and problem solving performance had been expected. This expectation was based on the analysis presented in Chapter 1 (which received some indirect support in the experiment reported in Chapter 5). Unfortunately, a separate memory factor did not appear. The two tests which were supposed to define a memory factor coalesced with several other tests to produce a factor which was interpreted as Speed. Assuming there is some memory component to this factor, the correlations are very low and therefore not supportive of the hypothesis. Memory was expected to play a larger role in performance with the pencils than with the cards and cowboys, because the pencils are shuffled after each problem whereas the spatial arrangement remains the same during each problem for the other materials. For what it is worth, Factor II correlated .18 with pencil performance and only .05 with card and cowboy performance. A further difficulty is that what the psychometician calls Associative Memory is a very impure measure of memory since it does not exclude learning. Memory span measures, which were not included in this study, would have been more relevant to the hypothesis.

Summary and conclusions

The chief conclusions of the study reported in this chapter are as follows:

1. Given a specified set of materials and a specified number of relevant stimulus dimensions, the particular concept to be attained evidently has little effect on children's problem solving behavior. There were only three significant ($\alpha = .05$) effects in 27 one-way analyses of variance. Each analysis involved six groups each of which received a different concept. There was no apparent pattern to the three significant effects that were obtained.

2. Order of presentation of problem sets had a strong effect on all four dependent variables. Problem solving performance improved from the first to the third set of three problems.

3. One-dimensional problems were easier to solve than either 2- or 3-dimensional problems.

4. Type of materials affected only number of logically inappropriate choices of instances per problem. There were more such choices during cowboy problems than during card or pencil problems, perhaps because the "story line" employed with the cowboys led to interfering associations (see Chapter 6).

5. There were more inappropriate conclusions, indeed, more verbal behavior both appropriate and inappropriate, during 3-dimensional problems than during 1- and 2-dimensional problems. This was interpreted to mean that positive instances play a role in controlling conclusion-drawing and conclusion-stating behavior (see Chapter 1 and Chapter 5). Since none of the instances, except the focus instance, which the child encounters during a 3-dimensional problem is positive, verbal behavior is therefore inhibited.

6. Seven of the ten expected factors materialized in the components analysis of the intercorrelation among the aptitude tests. The rotated factors were named as follows: I. Verbal Comprehension, II. Speed, III. Reasoning, IV. Ideational Fluency, V. Spatial Orientation, VI. Figural Adaptive Flexibility, VII. Originality.

7. Two aptitude factors--Spatial Orientation and Figural Adaptive Flexibility--showed modest correlations with problem solving performance. Evidently skill in dealing with spatial configurations plays a role in performance on concept attainment problems.

CHAPTER 5

Effects of the Availability of Concrete Stimulus Objects on Children's Problem Solving[1]

According to Piaget a child at the "stage of concrete operations," which is believed to occur from roughly seven to eleven years of age, will be able to solve many kinds of problems so long as these involve manipulation of concrete objects and reasoning about concrete objects which are physically present. However, the child at the stage of concrete operations is believed to be unable to engage in many forms of "abstract thinking" particularly those entailed in logical operations and inductive inquiry. The latter capabilities are believed to arise at the "stage of formal operations" which, it is alleged, develops during the period from about eleven to fourteen years of age. According to Inhelder and Piaget (1958, p. 335),

> The adolescent superimposes propositional logic on the logic of classes and relations. Thus, he gradually structures a formal mechanism (reaching an equilibrium point at about 14-15 years) which is based on both the lattice structure and the group of four transformations. This new integration allows him to bring inversion and reciprocity together into a single whole. As a result, he comes to control...hypothetico-deductive reasoning and experimental proof based on the variation of a single factor with the others held constant (all other things being equal)....

Anderson (1964, 1965) developed a training procedure which successfully taught first graders to solve concept attainment problems.

---

[1]Miss Carla Cucciatti executed the research described in this chapter during her term as an Undergraduate National Science Foundation Fellow. Miss Cucciatti drafted the first report of the experiment. The research was designed by the author who also wrote this report.

Given an instance of a phenomenon the majority of the children who received training were able to discover the necessary and sufficient conditions for the occurrence of the phenomenon. In the course of a problem most of the children performed few inappropriate experiments and seldom stated false or incomplete conclusions. The behavior of the majority of the children conformed to the strategy of varying each factor in succession while holding all other factors constant. Chapter 6 contains new evidence that first graders can be taught to solve concept attainment problems at a fairly high level of proficiency. Contrary to what one would expect on the basis of Piaget's work, first graders learned "hypothetico-deductive reasoning and experimental proof based on the variation of a single factor with the others held constant."

In the earlier experiment (Anderson, 1964, 1965) the children were able to solve problems involving new materials which they had not encountered during training. However, a high level of problem solving performance was maintained only with materials consisting of collections of concrete objects physically arrayed in front of the child. Children who received training did only slightly (and not significantly) better than untrained controls on a pendulum problem and two chemical problems. One of the chemical problems will be described for illustration. Materials consisted of a four-ounce amber bottle, a watch glass, and a set of five one-ounce dropping bottles labeled A, B, C, D, and E. Each bottle contained a colorless orderless liquid. Two of the bottles contained water. The other three contained, respectively, dilute solutions of sulphuric acid, hydrogen peroxide, and potassium iodide.

The problem was to find "which bottles you have to use to get yellow." The concept to be attained was that bottles B, C, and D were necessary. The focus instance was the demonstration that when drops from each of the five bottles were placed in the amber jar and the contents were poured on the watch glass, the liquid was yellow. The child was allowed to try a maximum of ten combinations of bottles in the attempt to solve the problem.

There are several differences between the chemical problem which has just been described and problems that involve collections of concrete objects. Note first, though, that the problems are isomorphic at a certain level of discourse. For both types, the optimum strategy for solution is to study each potentially relevant variable in succession while holding all other variables constant. One salient difference is that in the chemical problem the child must create or produce the instances needed to test the relevance of factors. It is the combination of bottles that constitutes an instance in the chemical problem, whereas in, say, a cowboy problem each concrete stimulus object (e.g. the cowboy on a horse with a hat but no rifle) is an instance. The chemical problem could be made to resemble a cowboy problem if the child were provided with thirty-two bottles, appropriately labeled, each of which contained one of the thirty-two possible combinations of five ingredients each either present or absent. Conversely a cowboy problem could be made to resemble the chemical problem if the child were provided with a cowboy, a horse, a hat, and a rifle and directed to construct the combination he wished to learn about. The same transformations were illustrated earlier (p. 12-13) with the pattern vision example and the card array.

Piaget has reasoned that skill in combining materials to produce instances that are relevant for testing hypotheses entails the capacity to "conceive combinatorial possibilities." While he might be able to select appropriately from already-formed combinations arrayed in his visual field, Piaget would not expect a child at the stage of concrete operations to be able to conceive, and therefore to construct, a combination appropriate to test an hypothesis. Herein lies one possible explanation of the failure of the first graders in the original experiment to do well on the chemical problems.

An alternative explanation, which the author favors, is that the physical presence of already-formed instances facilitates problem solving performance because the child is helped to remember what he has learned. This hypothesis is based on the analysis of the concept attainment problem presented in Chapter 1 (see also Chapter 6) and on a great deal of observation of children attempting to solve problems. A child working on a card or a cowboy problem will sometimes spontaneously hold his fingers on each of the instances he selects. Occasionally a child will even use the fingers of one hand to mark positive instances and the fingers of the other hand to mark negative instances. This sort of behavior is much more frequent on pencil problems. Many children are observed to place the pencils into two piles spontaneously, one for those that "show the secret," the other for those that do not.

The evident function of marking the instances that have already been chosen is to enable the child to remember them.

At one point, in order to make pencil problems more comparable to problems arranged with other materials, children were prohibited from handling the pencils. A deterioration in performance was observed as a result. Several children were instructed not to touch the pencils. Several others were expressly directed to sort the positive instances into one pile and the negative instances into another. The former children performed better.

There are indications that a system of marking the instances in some way as they are selected facilitates performance, presumably because memory is thereby improved. On the other hand, observation of children attempting to solve problems gives little indication that choosing logically appropriate instances is difficult for seven- to eleven-year-olds. This impression is supported by an analysis of the 1296 protocols collected during the experiment reported in Chapter 4. At his first choice there are seven instances the child could choose. The procedures effectively rule out the possibility that the child will choose the eighth instance, which is the focus. In fact the focus instance was not a first choice in any of the 1296 protocols. Three of the seven instances are logically appropriate; that is, three of the instances are different from the focus instance in terms of one factor and the same in terms of all other factors. Consequently, by chance alone a child would select a logically appropriate instance at his first choice 42.5% of the time. The observed frequency of logically appropriate first choices was 93.3%. Only first choices were considered because later choices could be affected by forgetting and because calculation of the probabilities of logically appropriate choices by chance alone become complicated for later instances.

In the present experiment, children worked pencil problems either "on the board" or "in the head." Under the "on the board" condition the eight pencils were physically arrayed in front of the child, whereas they were not in view under the "in the head" condition. The second factor was whether or not the instances selected by the child were made available and sorted into piles, or "marked," to make a physical record of choices. If children in the seven to eleven year age range are unable to "conceive combinatorial possibilities," then the "on the board" condition should lead to better performance than the "in the head" condition. If, on the other hand, the chief value of having available instances in the form of concrete stimulus objects is to facilitate memory, the presence of "markers" will be the important factor. The critical comparison for differentiating the two hypotheses is between "in the head" with markers and "in the head" without markers. The memory hypothesis predicts only a slight facilitation from the availability of markers when the instances are physically arrayed in front of the child. The facilitation is expected to be only slight because the child can remember the instances by noting their position in the array; he need not verbally encode the instances to remember them (see pp. 24-25). When the problem is worked in the head, the child must verbally encode the instances to remember them unless markers are

ERIC

available. Verbal encoding is probably a more difficult process than noticing the position of an instance and verbally encoded material is probably more vulnerable to interference. Thus, according to the memory hypothesis, performance will be poor only when the problem is worked in the head and no markers are available. The Piagetian hypothesis, in contrast, predicts poor performance for in-the-head problems whether or not markers are available.

Method

Subjects. The Ss were 16 boys and 8 girls. At the time of the study, the Ss were enrolled in a summer school in a midwestern community of 30,000 in which a large state university is located. There were 15 Ss who were entering the fourth grade the next fall and 9 who were entering the third grade. The Ss were randomly assigned to conditions.

Procedure. The procedure was described in Chapter 2 but since procedure is critical for interpreting the present experiment it will be summarized again here. The materials consisted of eight pencils displaying three attributes: length (6 in. or 2 in.); sharpened or unsharpened; and presence or absence of an eraser. The female E worked individually with Ss. An S's participation in the experiment was completed in a single session of from 15 to 40 min. in length.

The Ss received an orientation task prior to attempting problems. In this task all of the pencils were arrayed on a table in front of the S. The E named a concept and the S was required to point to all of the positive instances of the concept. A list of 24 concepts--consisting of six 1-dimensional concepts, six 2-dimensional concepts, six 3-dimensional, and six repetitions of the 0-dimensional concept--were treated in this manner. Included were the concepts that would later have to be attained in order to solve problems. After each concept, E shuffled the pencils. When S failed to identify the positive instances of a concept, he was prompted to make a correct response and the concept was repeated after an interval until an unprompted correct response was made.

Each S received a 1-dimensional, a 2-dimensional, and a 3-dimensional problem. The specific concepts were randomly selected for each S independently of other Ss from among the six concepts of each type included in the orientation task. The three problems were presented in counterbalanced orders according to a Latin square. Before each problem, E read the standard instructions modified so that the S was directed to name the pencil he wished to learn about instead of pointing to it. The Ss were not permitted to touch the pencils under any of the conditions. The E exhibited one pencil and indicated that it "showed the secret." The S named pencils and attempted to state the concept. A series of standardized prompts, described in Chapter 2, was used to keep S performing. A problem was terminated when the child refused to continue, when he selected six instances without solving the problem, or when he solved the problem by selecting a series of instances and stating a concept such that the instances implied the concept and no other. The E wrote protocols which were scored in the manner detailed in Chapter 2. There were three dependent variables: number of solutions, number of

ERIC

inappropriate choices, and number of inappropriate conclusions.[2]

Design. Half of the Ss received the On the Board condition in which the pencils were arrayed on a table in front of the S. The pencils were shuffled before each problem under this condition. The remainder of the Ss received the In the Head condition in which the pencils were concealed from the child's view during the course of the problem. The focus instance, however, was placed on the table where it remained throughout the problem. Within both the On the Board and In the Head conditions, half of the Ss received Markers. That is, as the child named a pencil, the E selected the pencil and placed it in a designated position on the table. Positive instances, including the focus instance, were placed near the S's left hand. Negative instances were placed near his right hand. Under the No Marker condition, the pencils were not placed into piles in the manner just described. The treatments entailed under each of the conditions are summarized in Fig. 5.1.

The experimental conditions involved between-subject factors. Included also were two within-subject factors, number of relevant stimulus dimensions, and order of presentation. An incomplete analysis of variance was calculated since degrees of freedom were at a premium. Higher order interactions were pooled with error variance. The Dimension X Order interaction was also pooled with error since this interaction was thoroughly studied in the experiment reported in Chapter 4.

---

[2]The summary measure had not been developed when these data were analyzed and the data has not yet been reanalyzed.

Condition	Description of Treatment
On the Board-Marker	The eight pencils are arrayed in a disorderly fashion in front of the S. The E places each positive instance named by the S near S's left hand, and each negative instance near S's right hand.
On the Board-No Marker	The eight pencils are arrayed in a disorderly fashion in front of the S. At the beginning of the problem E places the focus instance near S's left hand, otherwise the arrangement of the pencils is not disturbed during the course of the problem.
In the Head-Marker	At the beginning of the problem E places the focus instance on the table near S's left hand. The E selects each pencil named by S from a box hidden from S's view. The E places these pencils in piles, one for positive instances, near S's left hand, and one for negative instances, near S's right hand.
In the Head-No Marker	At the beginning of the problem, E places the focus instance on the table near S's left hand. No other pencils are displayed during the course of the problem.

Fig. 5.1. Description of treatment under each experimental condition.

Results

Table 5.1 contains the means and standard deviations per problem for the three dependent variables as a function of experimental condition. Analyses of variance are summarized in Tables 5.2, 5.3 and 5.4.

There was a significant interaction between the factors defining the experimental conditions with respect to number of inappropriate choices. As can be seen in Table 5.1, the interaction appeared because of the large number of inappropriate choices under the In the Head-No Marker condition. While no other effects involving the experimental conditions were significant in the analyses of variance, the preplanned comparisons of the In the Head-Marker condition and In the Head-No Marker condition were made nonetheless using two-tailed t tests ($\alpha = .05$, $df = 10$). There were significant differences in the direction expected on the basis of the memory hypothesis for number of solutions ($t = 2.91$) and number of inappropriate choices of instances ($t = 3.06$) but not for number of inappropriate conclusions ($t = .33$).

There were significant differences in number of inappropriate conclusions as a function of the number of relevant stimulus dimensions defining the concept to be attained. The means per problem were 1.42, 1.25, and .71 for the 1-, 20, and 3-dimensional problems, respectively. This result parallels the result obtained in the large experiment reported in Chapter 4.

Table 5.1

Means and SDs Per Problem for Each Measure

Condition	Number of Solutions M	Number of Solutions SD	Number of Inappropriate Choices M	Number of Inappropriate Choices SD	Number of Inappropriate Conclusions M	Number of Inappropriate Conclusions SD
On the Board-Markers	.50	.35	2.44	.78	1.28	.77
On the Board-No Markers	.44	.40	2.33	1.17	1.06	.77
In the Head-Markers	.50	.23	2.39	.90	.94	.68
In the Head-No Markers	.11	.17	3.94	.85	1.06	.49

ERIC

Table 5.2

Analysis of Variance of Number of Solutions

Source	df	MS	F
Between Ss			
On the Board vs In the Head (P)	1	.50	1.70
Markers vs No Markers (M)	1	.89	3.02
P X M	1	.50	1.70
Ss within groups	20	.29	
Within Ss			
Dimensions (D)	2	.10	.45
D X P	2	.04	.20
D X M	2	.35	1.61
Order (O)	2	.10	.45
O X P	2	.04	.20
O X M	2	.18	.84
Residual	36	.21	
Total	71		

Table 5.3

Analysis of Variance of Number of Logically Inappropriate Choices of Instances

Source	df	MS	F
Between Ss			
On the Board vs In the Head (P)	1	10.89	4.11
Markers vs No Markers (M)	1	9.39	3.54*
P X M	1	12.50	4.72*
Ss within groups	20	2.65	
Within Ss			
Dimensions (D)	2	.72	.69
D X P	2	.06	.05
D X M	2	1.06	1.00
Order (O)	2	1.35	1.28
O X P	2	.18	.17
O X M	2	1.01	.96
Residual	36	1.05	
Total	71		

*$p < .05$

Table 5.4
Analysis of Variance of Number of Logically
Inappropriate Conclusions

Source	df	MS	F
Between Ss			
On the Board vs In the Head (P)	1	.50	.35
Markers vs No Markers (M)	1	.06	.04
P X M	1	.50	.35
Ss within groups	20	1.42	
Within Ss			
Dimensions (D)	2	4.04	4.21*
D X P	2	.29	.30
D X M	2	.51	.54
Order (O)	2	3.04	3.17
O X P	2	.04	.04
O X M	2	.43	.45
Residual	36	.96	
Total			

*$p < .05$

Discussion

The results of this experiment clearly support the hypothesis that the presence of instances in the form of concrete stimulus objects facilitates problem solving performance of eight- and nine-year-olds by serving as a memory aid. There was no support for the hypothesis that the absence of concrete objects results in a deterioration in performance because children of this age are unable to conceive combinatorial possibilities and are, therefore, unable to construct (e.g. name the elements of) the instances logically required to solve problems. These conclusions were further supported in a subsidiary analysis of the data.

A rationale was presented in Chapter 2 for the subdivision of total number of inappropriate choices of instances into inefficient instances, redundant instances and repetitious instances. An inefficient instance on the average yields less than the maximum attainable amount of information. Operationally, an inefficient instance is different from the focus instance, or other positive instance, along more than one dimension. A redundant instance logically allows only a conclusion implied by previously-chosen instances while a repetitious instance exactly duplicates a previous instance.

Inefficient and redundant choices will occur when behavior is not under the control of the stimuli embodying the logical constraints in the problem. These logical constraints entail the "conception" of the instances the outcome of which, positive or negative, will control conclusion-stating behavior. In other words the problem solver must make responses that produce stimuli that evoke correct conclusions. Inefficient and redundant choices, which cannot logically lead to correct conclusions, should occur mainly because the problem solver cannot "conceive" appropriate instances, though late in the problem they might occur because he could not remember his previous choices. Presumably the only reason a problem solver would repeat an instance would be if he could not remember that he had previously chosen the instance or, if he remembered that he had chosen it, that he could not remember whether it did or did not "show the secret." If the conception-of-combinatorial-possibilities hypothesis is correct, the In the Head conditions should lead to more inefficient and redundant choices. The memory hypothesis leads to the expectation that the No Marker conditions, particularly the In the Head-No Marker condition, will lead to a marked increase in number of repetitious choices and, perhaps, a slight increase in number of inefficient and redundant choices.

Table 5.5 contains mean inefficient and redundant choices and mean repetitious choices by experimental condition. The differences among conditions in inefficient and redundant choices were not significant. Table 5.6 contains the analysis of variance for number of repetitious choices. Both factors and the interaction were significant. Comparisons (α = .05) using the Newman-Keuls procedure showed that there were more repetitious choices under the In the Head-No Marker condition than under any of the other three conditions and that there were no differences among the latter conditions.

Table 5.5

Mean Inefficient and Redundant Choices and Mean Repetitious Choices per Problem by Experimental Condition

Condition	Inefficient and Redundant Choices	Repetitious Choices
On the Board-Markers	2.33	.11
On the Board-No Markers	1.94	.39
In the Head-Markers	1.94	.44
In the Head-No Markers	2.28	1.67

ERIC

Table 5.6

Analysis of Variance of Repetitious Choices

Source	df	MS	F
On the Board vs In the Head (P)	1	35.04	16.85**
Markers vs No Markers	1	30.38	14.61**
P X M	1	12.04	5.79*
Ss within groups	20	2.08	
Total	23		

*$p < .05$

**$p < .01$

ERIC

It may be reasonably argued that the present experiment did not provide a strong test of the hypothesis that children are unable to conceive combinatorial possibilities. It could be maintained that the fact that the pencils were progressively displayed under the In the Head-Marker made the capacity to conceive combinatorial possibilities relatively unimportant whereas both hypotheses predict poor performance under the In the Head-No Markers condition. The telling counter-argument has to do with the nature of the poor performance under this condition. The data suggest that the poor performance under the In the Head-No Markers condition is attributable to memory failure rather than failure in logic.

A more general reservation is that the children had considerable experience with all of the pencils during the orientation task such that during problem solving it was not necessary for them to "conceive" entirely new combinations but merely to recall combinations--that is, pencils--which they had seen earlier. None of the data from the present experiment permit a rebuttal to this argument.

In summary, the data strongly suggest that one function of instances in the form of concrete stimulus objects is to make a physical record of the problem solver's responses and accompanying stimuli associated with these responses. The performance of eight- and nine-year-olds on concept attainment problems seems to be facilitated by the presence of concrete stimulus objects because the objects help the child remember what he has already done and what he has already learned.

The data do not support the Piagetian hypothesis that instances in the form of concrete stimulus objects facilitate the problem solving performance of children because they otherwise would not be able to conceive the instances logically needed to solve concept attainment problems; however, there remains the possibility that this theory would receive support under conditions r [illegible] included in the present experiment.

ERIC

CHAPTER 6

Part Versus Whole Task Procedures for Teaching a Problem-Solving Skill to Children[1]

Educators who have been influenced by the programed instruction movement take it as self-evident that the best way to teach a complex skill is to analyze it into component subskills and subconcepts, then teach each of these in turn. Cast in different language such an approach is a part-task method, to be contrasted with the whole-task method in which the student is required to perform the terminal behavior as best he can from the very beginning of training. Surprising as it may seem to those who have been influenced by the conceptions of programed instruction, the research on complex skill training has frequently shown whole methods to be superior to part methods.

The terms "part" and "whole" will be used in this chapter as shorthand words for talking about the issue of how lengthy and complicated a segment of a task the student should be required to attempt during instruction, especially during the initial stages of instruction. Part methods result in low initial error rates and fast progress, at least at the beginning of instruction, but when account is taken of the time to combine the parts, typically the advantage for the part method has been negligible at best. In the case of rote materials, practice on later parts (sublists) produces interference with earlier parts. This

[1]A version of this chapter will appear in the Journal of Educational Psychology.

interference must be overcome during the combination stage. With respect to complex skills and structured, meaningful material, there are coordinations and interrelationships among the subskills and subconcepts that cannot be acquired from training with the components alone. Herein lies one apparent reason that whole-task training has frequently proved superior to part training in the case of complex skills.

Whether a procedure which emphasizes lengthy task segments will prove superior to a procedure that begins with short task segments will surely be heavily dependent upon the manner in which the training procedures are developed. If entering behavior is underestimated and the steps in the part procedure are more finely granulated and more numerous than necessary, it may be less efficient than a procedure in which larger task segments are emphasized. Further, particularly when the task analysis underlying the part procedure is incomplete, subjects who receive the "part" or small-step procedure may fail to learn coordinations among component skills and concepts, whereas subjects trained from the beginning on larger segments of the task may induce these coordinations.

There may also be characteristics of tasks which systematically interact with the length and complexity of the instructional units into which the task is divided. Naylor and Briggs (1963) have suggested that the relative effectiveness of part and whole methods is a function of the level of complexity and the degree of organization of the task. Task "complexity" is said to refer to "demands on information-processing and memory-storage capacities" while task "organization" is said to refer to the nature and extent of the interrelationships among task dimensions. As organization increases, whole methods are predicted to be increasingly superior to part methods. For a highly organized task, an increase in complexity (difficulty) is predicted to result in greater superiority for the whole method. The part method is predicted to be superior to the whole method only in the case in which the task is both complex and unorganized. Naylor and Briggs (1963) completed an experiment, using what may be called concept learning tasks, in which it was found that the whole method was much better than the progressive-part method on a highly organized task regardless of task complexity. On an unorganized task, the whole method was slightly better when task complexity was low and very slightly worse than the progressive-part method when task complexity was high.

The experiment reported herein involved a comparison of a small-step, programed part-task method and a whole-task method for teaching children a complex problem solving skill, the skill of varying each factor in succession while holding all other factors constant. This skill is a classical strategy of experimental science and it is applicable to a large and important class of problems (Bruner, Goodnow, and Austin, 1956, pp. 81-125). The strategy was analyzed in detail in Chapter 1.

Relative to the tasks employed by Naylor and Briggs (1963) the problem-solving task used in the present experiment would have to be regarded as highly organized. Furthermore, the task is very difficult (complex) for seven-year-olds (Anderson, 1965), the age of the subjects

(Ss) in the present study. According to Inhelder and Piaget (1958, p. 335) people do not normally acquire the skill taught in the present experiment until 14-15 years of age. For these reasons, if Naylor and Briggs (1963) are correct, this should be a case in which the whole method is vastly better than the part method.

Method

Materials. Three sets of materials were employed. These materials were described in detail in Chapters 1 and 2. The description will be merely summarized here. The card array consisted of eight 2-1/2 X 3-1/4 in. cards taped in an orderly arrangement on a 16 X 18 in. Masonite panel. The cards had figures inscribed upon them that varied with respect to number (one figure or two figures), color (red or green), and form (rectangle or diamond). The cowboy game consisted of eight toy plastic cowboys and accessories imbedded in an orderly arrangement in a 12 X 13 in. plaster of paris base. The cowboys were either standing or riding horses, either with or without hats, and either with or without rifles. The final set of materials consisted of eight yellow, hexonanal number two pencils. The pencils displayed three attributes: length (6 or 2 in.), presence or absence of eraser, and sharpened or unsharpened. The pencils were arrayed in a disorderly fashion on a table in front of the S. While these three sets of materials involved different stimulus dimensions and different "story lines" were employed with them, an identical problem solving task could be created with each set of materials.

<u>Part-task training</u>. Three programs, one for each of the three sets of materials, were developed to teach children to apply the technique of varying each factor in succession while holding all others constant. These programs had their origins in a free wheeling, loosely-programed training procedure which, nonetheless, achieved considerable success with bright first graders (Anderson, 1965). A modification of this procedure, which more nearly resembled a program, was first developed for the card array. This 133-frame version was tried with 10 second graders from a lower middle class, urban school. Six of the 10 reached an <u>a priori</u> criterion of correct performance on four out of the last five frames in the program. One child stalled completely, failing to complete the program. Three other children completed the program but failed to reach the criterion. Among those who completed the program, the error rates ranged from 6.0% to 31.6% and the mean was 17.6%. The mean time to complete the program was 66 minutes. A substantial revision of the card array program was completed in which 49 frames were added, a number modified, and none eliminated. The revised program was tried with 10 second graders from a school in a rural community. On the last 10 frames of the program, upon which terminal behavior was required, seven children made no errors, one made one error, one made two errors, and one made three errors. The error rates over the entire program ranged from .5% to 18.7% and the mean was 8.7%. The mean time for completion was 73 minutes. A revision was made on the basis of these data. Eighteen frames were added to make a total of 200 and several other frames were modified. The card array program was not used again until the experiment described in this chapter.

The pencil program consisted of a literal translation of the card array program. In each frame, a pencil word or symbol was substituted for every card array word or symbol. No other modifications were made. Prior to the experiment, the pencil program was run with 10 naive second graders from a school in a rural community. The error rates ranged from 1.5% to 11.0% with a mean of 6.9%. The mean time to complete the program was 104 minutes. On the last 10 frames, six children made no errors, three made one error, and one made two errors. The cowboy program was also created by a literal translation of the card array program; however the cowboy program was not used with any children prior to the experiment.

The final form of the programs embodied an analysis of the total problem-solving strategy into two major subskills. The first section of each program was designed to teach appropriate conclusion-drawing behavior. E began by naming concepts while S pointed to all of the positive instances of the concepts. Then the roles were reversed; E pointed to all of the positive instances of concepts and S named the concepts. Next, E pointed to sets of instances, some positive and some negative, in such a way that each set defined a concept; S named the concepts. For example, E might point to a long sharpened pencil with no eraser and a short sharpened pencil with no eraser, indicating that each of these "showed the secret," then a long sharpened pencil with an eraser and a long unsharpened pencil without an eraser, indicating that the latter two did not "show the secret." If the child responded "the secret is sharpened with no eraser," he answered correctly. When S could correctly name seven out of eight consecutive concepts given a set of defining instances (a criterion he was required to meet before proceeding), he was then judged to have acquired a satisfactory approximation of the conclusion-drawing skill.

The second component in the total problem-solving skill is the skill of selecting appropriate instances. To begin the section of the program teaching this skill, E pointed to an instance. S was required to pick an instance which was different from E's instance in a specified way but the same in every other way. For example, S might be instructed to "pick a pencil just the same as mine except that it is a different length." After several frames involving the stimulus dimensions of a task taken one at a time, S was then required to pick three instances, each of which differed in exactly one respect from the instance designated by E. When S reached a criterion of seven out of eight consecutive correct selections of sets of three instances, he had mastered the skill of selecting instances.

The final section of each version of the part-task program taught the child to integrate the conclusion-drawing skill with the instance-selection skill. In this section, the child selected a set of instances; using the child's instance, E defined a concept; finally, the child named the concept. Next, E began indicating whether each instance was positive or negative as soon as the child selected it, instead of waiting until he had selected the entire set. This latter procedure was the same as with the terminal problems, except that when the child selected an inappropriate instance he was corrected before being told whether the instance was positive or negative. The child was not corrected on the 20 terminal frames, each of which entailed a problem presented using the same procedures as were used for test problems.

A standard correction procedure, not expressly described within the program, was implemented by E whenever S made an error. E created a new problem similar to the one upon which the error was made, told S the answer to the new problem, and then presented the original problem a second time. This procedure almost always prompted the correct response.

Whole-task training. Whole-task programs were developed for the card array, the cowboy game, and the pencil collection. The three versions of the program were equivalent in the sense that a systematic substitution of words and symbols would permit the literal translation of one version into another. The first section of the whole-task program was identical to the first section of the part-task program. In this section E named concepts and S pointed to all of the positive instances of the concepts. Thereafter, S received terminal problems. E designated a focus instance. S's task was to choose instances until he could name the concept. Whenever S chose an instance E indicated whether the instance "showed the secret." With one exception, the procedures for presenting problems within the whole-training program were the same as the procedures for administering test problems to be described in the next section. The exception was that when S selected six instances during a training problem without solving the problem E told the child the correct concept. Feedback of this sort was not given during the last 20 training problems nor during test problems.

The part-task and whole-task programs were equated in terms of the total number of task-relevant overt responses required under the assumption of error-free performance. This measure resembles measures such as number of trials that can be applied to simple tasks. For example, a child who behaves ideally on a terminal problem will select three instances and state a conclusion, a total of four distinguishable overt responses. Each of the versions of the part-task program required a total of 360 task-relevant, overt responses whereas each of the versions of the whole-task program required 364 such responses. The first 28 and the last 80 responses (20 terminal problems) were the same for both programs. In between, those who received the part-task program were led to make a progression of 252 responses designed to teach them a conclusion-drawing skill, an instance-selection skill, and to integrate the two, as detailed earlier. The middle section of the whole-task program, on the other hand, contained 64 terminal problems which could have been solved with 256 overt responses. It should be emphasized that these calculations are based on the assumption of error-free performance. Of course, errors were made. Based on data collected during the experiment, the typical S who received the part-task program made an estimated 410 overt, task-relevant responses while the typical S who received whole-task training made about 620 such responses.

<u>Procedure</u>. The training and the test problems were presented by three female graduate assistants, each of whom had had 15 or more hours experience training and testing children prior to the experiment. The author monitored one to two hours of each <u>E</u>'s preexperimental training and testing performance. Several staff conferences were held in which the letter and spirit of the procedures were detailed, ambiguities resolved, and difficult problems discussed. In addition, each <u>E</u> had a ten-page manual giving an overview of the experiment and summarizing training procedures, and a seven-page manual setting forth the procedures for administering test problems.

Each child was trained and tested by a single <u>E</u>. One-third of the <u>S</u>s under each treatment in the experiment were run by each <u>E</u>. Training and testing sessions were scheduled to be 20 min. in length. Unless the child was sick or some other circumstances such as a special school program intervened, the child received three sessions a week until he completed the training and the testing. Most sessions were conducted at three widely separated stations in a large general purpose room in the cooperating elementary school.

For both part-task training and whole-task training there was a mimeographed copy of the program for each child. The child did not read the program. Rather, the program was a script that guided the behavior of <u>E</u>. Except as otherwise indicated, <u>E</u> adhered closely to the program, which described the stimulus <u>S</u> was to see, contained the verbatim language <u>E</u> was to use, and indicated the response or responses <u>S</u> was to give.

Under both training methods E made generous use of social reinforcement. The frequency and contingencies of reinforcement were not expressly indicated within the programs, but instead were under the extemporaneous control of E. Particularly with respect to the whole-training procedure, which was quite aversive for some Ss (at the beginning of training, especially) E was coached to maintain a pleasant, nonjudgmental posture in the face of poor performance, and to find every opportunity to reinforce. Overall, E probably gave supplementary social reinforcement (in addition to feedback) for about every third correct response or chain of correct responses, except when S was doing poorly, in which case every correct response was reinforced.

Presentation and scoring of test problems. The procedure was described in detail in Chapter 2 and will not be repeated here. An S solved a problem when he selected a series of instances and stated a conclusion such that the instances implied the conclusion and no other. Performance on terminal problems was scored on a three point scale as follows: a) S solves the problem and neither makes any logically inappropriate choices of instances nor states any incorrect conclusions (2 points); b) S solves the problem but makes one or more inappropriate choices or states one or more incorrect conclusions (1 point); c) S fails to solve the problem (0 points). The test protocols were punched on cards and then scored on an IBM 1620 computer using a program written for this purpose. The terminal problems included within the training programs were scored on the same three point scale as the test problems. However, problems presented during training were scored on the spot by E and only an abbreviated protocol was written. The measure reported is percent of possible score. For example if an S solved a problem but made one or more inappropriate choices of instances his score would be 50% on that problem.

Design and subjects. There were three treatment conditions. One group (P) received part-task training. Another group (W) received whole-task training. A control group (C) received no treatment. Every S in the former two groups received training with two sets of materials. One-sixth of the Ss in each of the training groups was assigned to one of the six possible permutations of the three sets of training materials taken two at a time. About 48 hours after completing training, Ss in the training groups received eight test problems to assess retention. The retention problems involved the second set of materials with which S received training. Each control S received the retention problems during his first experimental session. One-third of the control Ss received problems involving each of the three sets of materials.

About 48 hours after receiving the retention problems Ss received a series of eight test problems to assess transfer of training. The transfer problems entailed the set of materials which S had not encountered during training. One-third of the control Ss received transfer problems involving each of the three sets of materials, a different set than was encountered during the retention problems.

With respect to both the retention and transfer problems, S received two 0-dimensional problems, two 1-dimensional problems, two 2-dimensional problems, and two 3-dimensional problems. The order of presentation of problems was randomized for each S independently of other Ss.

The Ss were 53 second-semester first graders from a predominantly middle-class school located in a new housing development on the outskirts of a Midwestern city of 30,000. These Ss were randomly selected from among all of the first graders in the school and randomly assigned to experimental conditions. Since the study was conducted over a three month period, the experimental conditions were scheduled in a predetermined random order. There were 18 Ss in Group W and in Group C but only 17 in Group P. There was to have been an eighteenth S in this group; however, the last S to be run (who appeared to be making normal progress) had to be dropped because of the impending end of the school year. He was replaced by a dummy case at the cell mean to balance the design for statistical purposes.

Classroom teachers administered the California Test of Mental Maturity (Long Form, 1963 Revision). Unfortunately one teacher found it necessary to terminate the examination in the middle of a subtest because of inattention and disorderly behavior, so it is not possible to report IQs or MAs. Raw score (not including Delayed Memory subtest score) means and standard deviations were 64.6 and 8.2 for Group P, 69.3 and 7.4 for Group W, and 66.4 and 7.0 for Group C ($F = 1.77$, $df = 2/52$, $p > .05$).

Results

Acquisition. Table 1 contains mean training times. Table 2 contains mean percent of possible score on the last 12 training problems. Table 3 presents the analyses of variance for these measures. As can be seen, Group P performed better on the problems whereas Group W completed training in a shorter period of time. Based on the estimates of number of responses made during training, which were described earlier, and the training times that appear in Table 1, it is estimated that during training Group P made relevant, overt responses at the rate of about 3.7 per min. The rate for Group W is estimated to have been about 7.9 per min.

The mean training times for Group P were considerably higher than the times obtained during preexperimental development of the programs. Part of the discrepancy was no doubt due to the fact that second graders were used in most of the developmental work while first graders were employed in the experiment. Also at two points in the versions of program used in the experiment S had to reach a criterion before proceeding. These criteria were not part of the preexperimental procedure.

It might be argued that if Group W had been allowed as much training time as Group P it would have performed as well on the terminal problems. Fig. 1 pictures performance over blocks of 12 problems for Group W. Notice that performance reaches an asymptote by the sixth or seventh block on the first task. Consequently, it seems highly improbable that further practice would have improved the performance of Group W very much.

Table 6.1

Mean Training Time in Minutes

	First Task		Second Task	
Materials	Part Training	Whole Training	Part Training	Whole Training
Cards	96.4	76.3	96.2	53.0
Cowboys	144.9	93.2	81.7	65.3
Pencils	135.9	98.2	107.7	85.2
All materials	125.7	89.2	95.2	67.8

Note.--The SDs (estimated from MS error terms) were 25.9 for the first task and 24.4 for the second task.

ERIC

Table 6.2

Mean Percent of Possible Score on the Last Twelve Training Problems

	First Task		Second Task	
Materials	Part Training	Whole Training	Part Training	Whole Training
Cards	88.9%	56.9%	63.9%	57.6%
Cowboys	49.3	54.2	85.4	42.4
Pencils	84.0	43.7	81.2	63.2
All materials	74.1	51.6	76.9	54.4

Note.--The SDs (estimated from MS error terms) were 17.3 for the first task and 18.0 for the second task.

ERIC

Table 6.3

Analysis of Variance for Acquisition Data

		Training Time				Terminal Problems			
		First Task		Second Task		First Task		Second Task	
Source	df	MS	F	MS	F	MS	F	MS	F
Treatment (T)	1	11,953.78	17.81**	6,724.00	11.34**	261.36	15.14**	261.36	13.93**
Materials (M)	2	4,023.11	5.99**	2,006.09	3.38*	78.09	4.52*	24.25	1.29
T X M 2	2	755.45	1.13	592.58	1.00	99.70	5.78**	61.03	3.25
Ss w. groups	30	671.33		593.19		17.26		18.76	

*$p < .05$

**$p < .01$

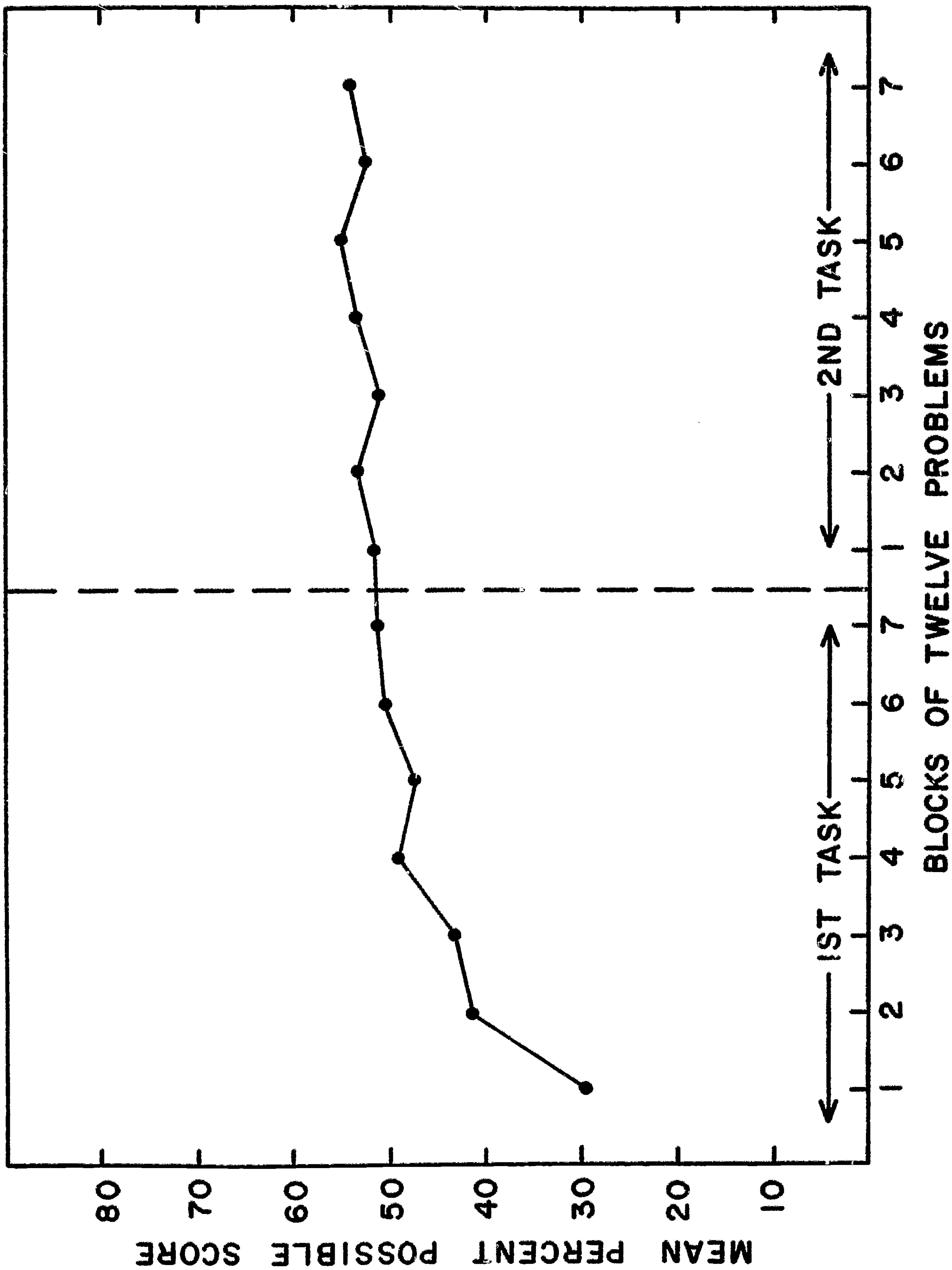


Fig. 6.1. Mean percent of possible score on blocks of twelve training problems within the whole training group.

There were significant differences between materials on the first training task due in large part to the relatively poor performance with the cowboys. We have observed that the story line employed with the cowboys tends to interfere with the problem-solving of some children, who insist that the "friends of the sheriff" must have rifles or must ride horses. Other investigators have made similar observations (Bruner, _et al_, 1956, p. 111). Evidently by the time he reached the second task _S_ had learned enough so that he was not distracted by the story line.

There were significant materials effects on training time due to the pencils. The pencils were handled by _S_ and shuffled by _E_ after each problem. These manipulations took time.

Retention and transfer. Table 4 presents the analyses of variance for the retention and transfer problems while Table 5 contains the means for the two kinds of problems. In both cases there were insignificant differences among treatments. Comparisons (α = .01) using the Newman-Keuls procedure indicated that on retention problems Group P was superior to both the other groups and Group W was superior to Group C. On the transfer problems Groups P and W were not significantly different but both were superior to Group C.

There were significant differences in the difficulty of the transfer problems according to the number of relevant stimulus dimensions the problems entailed. The means were 44.9%, 49.1%, 42.6% and 32.9% for the 0-, 1-, 2-, and 3-dimensional problems, respectively. There was also

Table 6.4

Analysis of Variance for Retention and Transfer Data

		Retention Problems		Transfer Problems	
Source	df	MS	F	MS	F
Between Ss					
Treatment (T)	2	103.39	60.55**	26.54	11.80**
Materials (M)	2	1.01	.59	2.06	.91
T X M	4	5.42	3.18*	3.10	1.38
Ss w. groups	45	1.71		2.25	
Within Ss					
Dimensions (D)	3	1.07	2.40	4.08	6.00**
T X D	6	.64	1.44	.81	1.19
M X D	6	.30	.68	2.12	3.12**
T X M X D	12	.62	1.39	.62	.92
Ss w. groups X D	135	.45		.68	

*$p < .05$

**$p < .01$

ERIC

Table 6.5

Mean Percent of Possible Score on Retention and Transfer Problems

	Part Training	Whole Training	No Training
Retention problems			
Cards	62.5%	57.3%	10.4%
Cowboys	86.5	35.4	18.8
Pencils	76.1	55.2	16.7
All materials	75.0	49.3	15.3
Transfer problems			
Cards	44.8	58.3	30.2
Cowboys	49.0	41.7	21.9
Pencils	65.6	46.9	22.9
All materials	53.1	49.0	25.0

Note.--The SDs (estimated from MS error terms) were 16.3 for the retention problems and 18.8 for the transfer problems when scores were pooled across dimensions.

a significant Materials X Dimensions interaction for the transfer problems, due primarily, for reasons which are not clear to the author, to the relatively great difficulty of the 0-dimensional card problems.

Discussion

The results indicate some limits to the generality of the rule proposed by Naylor and Briggs that whole training will be superior to part training for "highly organized" tasks. Of course it may be that there are characteristics of tasks, such as the amount and nature of its organization, which systematically interact with the length and complexity of the responses required from the subject at various stages during training, but this is evidently a matter about which there is much to be learned. The present author is pessimistic about the liklihood that broad generalizations concerning method-task interactions will emerge in the near future. Too much depends upon the specific features of the methods and the details of implementation of these features.

Both informal observation and the objective data suggest that in the present study Ss who received whole training did not acquire a systematic instance-selection skill, or, at least, that they did not acquire the same skill as was acquired by most Ss who received part training. The whole group made a mean of 1.88 logically inappropriate choices of instances per retention problem, whereas the mean was 1.08 for the part group ($t = 3.19$, $df = 33$, $p < .01$). Furthermore, there is reason to believe that there were qualitative differences in the conclusion-drawing behavior of the two groups. Most Ss in the part group learned to draw conclusions based on a restricted set of instances containing only the minimal information logically necessary to solve the problem. Ss in the whole group seldom gave conclusions until they had selected a larger-than-logically-necessary set of instances. The typical S in Group W rapidly and, seemingly, haphazardly selected instances until a conclusion occurred to him. Most Ss in Group P, on the other hand, selected instances slowly and their behavior usually conformed to the method of varying each factor in succession while holding all other factors constant. At the point at which just enough information was available logically to solve the problem, the typical S who received part training usually offered a conclusion. The marked differences between Groups P and W in rate of response during training can be traced to the contrasting patterns of behavior typical of Ss in the two groups.

It seemed possible that children who received part training would be able to solve a high percent of the problems created with a new set of materials. Obviously this did not happen, indicating the need for a more refined method of producing generalized stimulus control (Anderson, 1965).

Group P did not do as well as would be expected on the basis of the preexperimental data, probably because of the fact that first graders participated in the experiment whereas second graders were employed during program development. Actually the part procedure was successful with the majority of the first graders. Of the 17 Ss who completed part training, 11 scored 80% or better on the retention problems and only three scored below 50%. The latter three fell at the bottom of the distribution of aptitude test scores.

Unnecessarily small and redundant steps and failure to provide for integration of subskills and concepts may be shortcomings of any particular lesson employing a small-step procedure. The guidelines for lesson development which have emerged from the programed instruction movement have not been demonstrated to guard against these shortcomings; indeed, it is possible that such deficiencies, particularly unnecessary redundancy, are endemic in currently available small-step, programed lessons. In the present instance, a small-step procedure worked relatively well. Whether a small-step, programed procedure would consistently prove best in other instances remains to be seen.

CHAPTER 7

Summary

A classical strategy of experimental science is to vary, manipulate, or study each of the potentially relevant variables in succession while holding all other variables constant. This strategy, in simple or complex form, is the optimum method for solving a large and socially important class of problems, including problems in scientific inquiry, medical diagnosis, and the trouble shooting of equipment malfunctions. One easy-to-study representative of this class of problems is the concept attainment problem. The subject (S) is shown a positive instance (i.e. an example) of a concept called the "focus instance." He then chooses instances. At each choice he is informed as to whether the instance is positive or negative. As soon as he is sure he knows, the S names the concept. This report contains a theoretical and experimental analysis of children's behavior on concept attainment problems.

A standardized procedure was developed for administering concept attainment problems to children in the six to ten year age range. In all cases an experimenter (E) worked with a single S at a time in sessions that were usually about 20 min. in length. The E read directions, indicated the focus instance, gave feedback as to whether the instances the S chose were positive or negative, gave one or another of a series of standardized prompts necessary to keep S performing, and decided when the problem should be terminated. The E wrote a protocol indicating in coded form the instances S selected, the conclusions S stated, and the behavior of E in the sequence in which these events occurred. The protocols were content analyzed using a computer program prepared for this purpose. Three basic measures of problem solving performance were employed. The first was number of problem solutions. An S solved a problem if he selected a series of instances and stated a conclusion such that the instances logically implied the conclusion and no other. The second measure was number of logically inappropriate choices of instances while the third was number of logically inappropriate conclusions. A summary measure represented problem solving performance with a single score: a) the S solved the problem and neither made inappropriate choices nor stated inappropriate conclusions, b) the S solved the problem but made one or more inappropriate choices or conclusions, c) the S failed to solve the problem.

The first experiment investigated both task factors and aptitude factors. A sample of 144 fourth graders completed nine concept attainment problems and also a battery of 24 aptitude tests. The task factors investigated were order of presentation, type of materials, and number of stimulus dimensions defining the concept to be attained. Also investigated were the difficulty of specific concepts,given the materials,and number of relevant stimulus dimensions. Experimenter and school were replication factors. The chief results and conclusions of this study were as follows.

1. Given a specified set of materials and a specified number of relevant stimulus dimensions, the particular concept to be attained

evidently has little effect on children's problem solving behavior. There were only three significant (α = .05) effects in 27 one-way analyses of variance. Each analysis involved six groups,each of which received a different concept. There was no apparent pattern to the three significant effects that were obtained.

2. Order of presentation of problem sets had a strong effect on all four dependent variables. Problem solving performance improved from the first to the third set of three problems.

3. One-dimensional problems were easier to solve than either 2- or 3-dimensional problems.

4. Type of materials affected only number of logically inappropriate choices of instances per problem. There were more such choices during cowboy problems than during card or pencil problems, perhaps because the "story line" employed with the cowboys led to interfering associations.

5. There were more inappropriate conclusions, indeed, more verbal behavior both appropriate and inappropriate, during 3-dimensional problems than during 1- and 2-dimensional problems. This was interpreted to mean that positive instances play a role in controlling conclusion-drawing and conclusion-stating behavior. Since none of the instances, except the focus instance, which the child encounterd during a 3-dimensional problem was positive, verbal behavior was therefore inhibited.

6. Seven of the ten expected factors materialized in the components analysis of the intercorrelations among the aptitude tests. The rotated

factors were named as follows: I. Verbal Comprehension, II. Speed, III. Reasoning, IV. Ideational Fluency, V. Spatial Orientation, VI. Figural Adaptive Flexibility, VII. Originality.

7. Two aptitude factors--Spatial Orientation and Figural Adaptive Flexibility--showed modest correlations with problem solving performance. Evidently skill in dealing with spatial configurations plays a role in performance on concept attainment problems.

A second experiment investigated the effects of concrete stimulus objects on children's problem solving behavior. A Piagetian hypothesis is that concrete stimulus objects are important for the problem solving performance of children in the seven to eleven year age range because children of this age are unable to "conceive combinatorial possibilities." Consequently, if instances in the form of concrete stimulus objects are not available the child will be unable to select or construct instances logically appropriate to solve concept attainment problems. An alternative hypothesis, based on the analysis and data of this project, is that concrete stimulus objects are chiefly important as a memory aid; the objects help the child to remember what he has already learned. A sample of 24 third and fourth graders worked three pencil problems either "on the board," in which the pencils were arrayed in front of the child, or "in the head," in which the pencils were concealed from the child's view. Under each condition, half of the Ss received Markers. That is, each pencil named by S was selected by E and placed in view. The remainder received the No Markers condition. Consistent with the memory hypothesis, but not the Piagetian hypothesis,

ERIC

the In the Head-No Markers group showed significantly more inappropriate choices and significantly fewer solutions than the In the Head-Markers group, whereas the In the Head-Markers group was not different from the two On the Board groups. A further, fine-grained analysis suggested that the poor performance of the In the Head-No Markers group was attributable to memory failure rather than logical failure. An incidental result was that there were fewer inappropriate conclusions during 3-dimensional problems than during 1- and 2-dimensional problems, confirming the result obtained in the preceding experiment.

A third experiment compared methods of teaching children to solve concept attainment problems. Because of the programed instruction movement, it is widely believed that the optimum way to teach a complex skill is to analyze it into component skills and concepts and then teach each of these in turn. Put in other terms this is a "part-task" method to be contrasted with a "whole-task" procedure in which the learner attempts the total skill early in training. The literature suggests that the whole-task procedures are better than part-task procedures particularly when a highly organized skill is being taught. Nonetheless, there is reason to believe that a programed part-task procedure might lead to better performance than an appropriate whole-task procedure even with respect to highly organized skills, provided that the part method is based on an adequate analysis of the task that takes account of the relationships among component skills. A part-task program was developed in a lengthy process of tryout and revision. The whole-task program included the same introductory and terminal

ERIC

frames as the part-task program but the middle portion contained 84 terminal problems. These problems could be solved with 256 task-relevant responses whereas those who received the part-task program were led to make a progression of 252 overt responses designed to teach them a conclusion-drawing skill, an instance-selection skill, and the integration of the two into a total skill. Groups of 18 first graders received either part-task or whole-task training with two sets of materials. Later these two groups, and a third group which received no training, got problems involving materials included during training to assess retention and problems from materials not included during training to assess transfer. On retention problems the part-task group did significantly better than the whole-task group which was in turn significantly better than the no-training group. The mean score on the summary measure of problem solving was 75% for the part-task group, indicating that the average child who received this form of training could solve three out of four problems presented to him without making a single logically inappropriate choice of instances or a single logically inappropriate conclusion. Those who received whole-task training showed improvement but there was a rather low ceiling under this method. On the fourth block of 12 training problems whole-task Ss averaged 49% of the possible score while they averaged only 54% on the 14th block. Both the part-task and whole-task group performed significantly better on the transfer problems than did the no-training group; however, there was no difference between the former two groups. The results of this experiment confirm Anderson's (1964, 1965) earlier finding that, contrary to prominent developmental theories, first graders can develop a high level of skill at solving complex problems said to involve hypothetico-deductive reasoning. The experiment further showed some limits to the generalization that whole-task methods will be superior to part-task methods for highly organized skills.

REFERENCES

Anderson, R. C. Shaping logical behavior in six- and seven-year olds. Urbana: University of Illinois, 1964.

Anderson, R. C. Can first graders learn an advanced problem-solving skill? Journal of Educational Psychology, 1965, 56, 283-294.

Anderson, R. C., & Anderson, R. M. Transfer of originality training. Journal of Educational Psychology, 1963, 54, 300-304.

Braund, R. A., & Heath, R. W. A pilot study of a cognitive restructuring paradigm. Paper read at the Annual Meeting of the American Educational Research Association, Chicago, 1965.

Bruner, J., Goodnow, J., & Austin, G. A. Study of thinking. New York: Wiley, 1956.

Cofer, C. N. Reasoning as an associative process: III. The role of verbal responses in problem solving. Journal of Genetic Psychology, 1957, 57, 55-68.

Conry, R. F. Canonical relationships among measures of concept formation, information processing, and other hypothesized cognitive factors. Unpublished Master's thesis, University of Wisconsin, 1965.

Dick, W. Retention as a function of paired and individual use of programed instruction. Journal of Programed Instruction, 1963, 2, 2, 17-23.

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

Fleishman, E. A. A comparative study of aptitude patterns in unskilled and skilled psychomotor performance. Journal of Applied Psychology, 1957, 41, 263-272.

Fleishman, E. S., & Hempel, W. E., Jr. Changes in factor structure of a complex psychomotor test as a function of practice. Psychometrika, 1954, 18, 239-252.

French, J. W., Ekstrom, R. B., & Price, L. A. Kit of reference tests for cognitive factors. Princeton, N.J.: Educational Testing Service, 1963.

Frick, J. W., & Guilford, J. P. An analysis of a form of the water jar test. American Journal of Psychology, 1957, 70, 427-431.

Gagne´, R. M. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gagne´, R. M., & Paradise, N. E. Abilities and learning sets in knowledge acquisition. Psychological Monographs, General and Applied, 1961, 75, No. 14.

Glass, G. V., & Maguire, T. O. Abuses of factor scores. American Educational Research Journal, 1966, 3, 297-304.

Hovland, C. I. A "communication" analysis of concept learning. Psychological Review, 1952, 59, 461-472.

Inhelder, B., & Piaget, J. The growth of logical thinking. New York: Basic Books, 1958.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23, 187-200.

Kaya, E. Solving puzzles. New York: New York University, 1960.

Kendler, H. H., & Kendler, T. S. Vertical and horizontal processes in problem solving. Psychological Review, 1962, 69, 1-16.

Maltzman, I. On the training of originality. Psychological Review, 1960, 67, 229-242.

Maltzman, I. Thinking: From a behavioristic point of view. Psychological Review, 1955, 4, 275-286.

Maltzman, I., Belloni, M., & Fishbein, M. Experimental studies of associative variables in originality. Psychological Monographs, 1964, 78, (3, Whole No. 580).

Merrifield, P. 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).

Naylor, J. C., & Briggs, G. E. Effects of task complexity and task organization on the relative efficiency of part and whole training methods. Journal of Experimental Psychology, 1963, 65, 217-224.

Polya, G. Patterns of plausible inference. Princeton: Princeton University Press, 1954.

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

Schulz, R. W. Problem-solving behavior and transfer. Harvard Educational Review, 1960, 30, 61-77.

Skinner, B. F. Verbal behavior. New York: Appleton-Century-Crofts, 1957.

Spielberger, C. D., & DeNike, L. D. Descriptive behaviorism versus cognitive theory in verbal operant conditioning. Psychological Review, 1966, 73, 306-326.

Stake, R. E. Learning parameters, aptitudes, and achievements. Psychometric Monographs, 1961, (Whole No. 9).

Stebbing, M. A. Analysis of Mr. Mill's system of logic. London: Longmans, Green, 1888.

Stolurow, L. M. Programed instruction and teaching machines. In P. H. Rossi and B. J. Biddle (Eds.), The impact of new media on education and society. Chicago: Aldine Publishing Co., 1965.

Stolurow, L. M., & Davis, D. J. Teaching machines and computer-based systems. Mimeo. Urbana: University of Illinois, 1965.

Tatsuoka, M. M. Mathematical models in the behavioral and social sciences. Ditto. Urbana: University of Illinois, 1965.

Walsh, N. E. The relationship between performance on certain long problems and intelligence factors. Unpublished Doctor's thesis, Rutgers University, 1963.

Willis, R. D., & Anderson, R. C. Dimensions of ability among Mattoon, Illinois fourth grade students. Mimeo. Urbana: University of Illinois, 1965.

Wilson, R. C., Guilford, J. P., Christensen, P. R., & Lewis, D. J. A factor-analytic study of creative thinking abilities. Psychometrika, 1954, 19, 297-311.

Woodrow, H. The relation between abilities and improvement with practice. Journal of Educational Psychology, 1938, 29, 215-230.

Woodrow, H. The ability to learn. Psychological Review, 1946, 53 147-158.

APPENDIX I

C O N T E N T   A N A L Y S I S   P R O G R A M
CARDS

```
      DIMENSION  ITASK(2), J(61), ITRIL1(6), ITRIL2(6), ITRIL3(6),
     1           ITRIL4(6), ITRIL5(6), ITRIL6(6), ICCEP(8),
     2           IFELD1(5), IFELD2(5), IFELD3(5), IFELD4(5),
     3           IFELD5(5), IFELD6(5), IFELD7(5), A(6), AA(6), JJ(11)
C     THE FIRST CARD MUST SAY PRINT OR PUNCH, STARTING IN COLUMN ONE
      READ 9500, I
 9500 FORMAT(1X,A1)
      IF(I-5900)9501,9503,9504
 9501 PRINT 9502
 9502 FORMAT(62HTHE FIRST CARD MUST SAY PRINT OR PUNCH, STARTING IN COLU
     1MN ONE)
      GO TO 9999
 9503 IPRINT=1
      GO TO 9505
 9504 IPRINT=2
 9505 ISKIP=1
 4002 IFIELD=0
C                    M A I N   L O O P
      DO 4100 LOOP=1,100
C     FIELDS -- USE (IFELDN(IFIELD),IFIELD=1,5)
      ALOOP=LOOP
      FLOOP=LOOP/5
      IF((ALOOP/5.0)-FLOOP)3519,6005,6000
 6000 IFIELD=1+IFIELD
      GO TO (6001,6002,6003,6004), IFIELD
 6001 READ 1, A, J
 1    FORMAT(1X,A3,A1,A3,A4,A3,A3,1X,61A1)
      IF(J(13)-7100)5058,9549,5060
 5060 GO TO (9554,6001),ISKIP
 9549 IF(LOOP-1)9551,9551,9550
 9550 GO TO (9560,9562),IPRINT
 9560 IF(ITRIAL-6)9561,9551,9551
 9561 PRINT 1002
      GO TO 9551
 9562 IF(ITRIAL-6)9563,9996,9996
 9563 IITERM=IITERM+1
 9996 IUNNEC=IINEFF+IREDUN+IREPET
      IINAPP=ICONBI+IINCON+IVM+IVPR
      IEXBEH=IP1+IP2+IP3+IM1+IM2+IM3+IE1+IE2+IPW
      PUNCH 9555,      AA,(JJ(I),I=1,5),(JJ(I),I=7,11),IEFFIC,
     1    IINEFF,IREDUN,IREPET,ICONBI,IINCON,IVM,IVPR,IP1,IP2,IP3,IM1,
     2            IM2,IM3,IE1,IE2,IPW,IRC,IITERM,ISOLVE,IUNNEC,IINAPP,
     3            IEXBEH
 9555 FORMAT(1H0,A3,A1,A3,A4,A3,A3,1X,5A1,1X,5A1,1X,22I1,I2)
C     THE FOLLOWING VARIABLES STORE TRIAL DIFFERENCES
 9551 DO 4003 I=1,6
      ICCEP(I)=0
      ITRIL4 (I)=0
 4003 ITRIL6(I)=0
      ICCEP(7)=0
```

```
      ICCEP(8)=0
      ICHANG=1
      ICONC1=0
      ICONC2=0
      ICONC3=0
      ICONC4=0
      ITRIAL=0
      IVERB4=0
      IEFFIC=0
      IINEFF=0
      IREDUN=0
      IREPET=0
      ICONBI=0
      IINCON=0
      IVM=0
      IVPR=0
      IP1=0
      IP2=0
      IP3=0
      IM1=0
      IM2=0
      IM3=0
      IE1=0
      IE2=0
      IPW=0
      IRC=0
      IITERM=0
      ISOLVE=0
      ILEFT2=8
      GO TO (9552,9554),IPRINT
 9552 PRINT 9564
 9564 FORMAT(///)
      PRINT 9553,A(3),A(6),(J(I),I=1,5),(J(I),I=7,11)
 9553 FORMAT(5X,10HID NUMBER ,A3,5X,A3,5H TASK,5X,11HCONCEPT IS ,5A1,5X,
     19HFOCUS IS ,5A1)
 9554 ISKIP=1
      DO 9997 I=1,6
 9997 AA(I)=A(I)
      DO 9998 I=1,11
 9998 JJ(I)=J(I)
      IFELD1(1) = J(17)
      IFELD2(1) = J(18)
      IFELD3(1) = J(19)
      IFELD4(1) = J(20)
      IFELD5(1) = J(21)
      IFELD6(1) = J(22)
      IFELD7(1) = J(23)
      GO TO 6006
 6002 IFELD1(2) = J(26)
      IFELD2(2) = J(27)
      IFELD3(2) = J(28)
      IFELD4(2) = J(29)
      IFELD5(2) = J(30)
      IFELD6(2) = J(31)
      IFELD7(2) = J(32)
      GO TO 6006
 6003 IFELD1(3) = J(35)
      IFELD2(3) = J(36)
```

```
      IFELD3(3) = J(37)
      IFELD4(3) = J(38)
      IFELD5(3) = J(39)
      IFELD6(3) = J(40)
      IFELD7(3) = J(41)
      GO TO 6006
 6004 IFELD1(4) = J(44)
      IFELD2(4) = J(45)
      IFELD3(4) = J(46)
      IFELD4(4) = J(47)
      IFELD5(4) = J(48)
      IFELD6(4) = J(49)
      IFELD7(4) = J(50)
      GO TO 6006
 6005 IFIELD =5
      IFELD1(5) = J(53)
      IFELD2(5) = J(54)
      IFELD3(5) = J(55)
      IFELD4(5) = J(56)
      IFELD5(5) = J(57)
      IFELD6(5) = J(58)
      IFELD7(5) = J(59)
      ICHANG=2
C     DECODE THE FIRST COLUMN OF THE CONCEPT FOR THE CARD TASK
C     (COLUMN 20)
C     AND LET US CHECK TO SEE IF THE TRIAL SIGN WAS CORRECT WITH
C     RESPECT TO THE CONCEPT
 6006 IF(J(1)-4200)9099,3002,13
 9099 ICONC4=1
      IWAS1=2
      IWAS2=2
      IWAS3=2
      GO TO 40
 3002 ICONC3=1
      GO TO 33
 13   IF(J(1)-4400)3501,3003,14
 3003 ICONC3=2
      GO TO 33
 14   IF(J(1)-4700)3501,3004,15
 3004 ICONC2=1
      GO TO 20
 15   IF(J(1)-5900)3501,3005,10
 3005 ICONC2=2
      GO TO 20
 10   IF(J(1)-7100)3501,3000,11
 3000 ICONC1=1
      GO TO 20
 11    IF(J(1)-7200)3501,3001,3501
 3001 ICONC1=2
C     DECODE THE SECOND COLUMN OF THE CONCEPT FOR THE CARD TASK
C     (COLUMN 21)
 20   IF(J(2)-4700)21,201,202
 201  ICONC2=1
      GO TO 30
 202  ICONC2=2
      GO TO 30
 21   IF(J(2)-4200)33,22,23
```

```
22      ICONC3=1
        GO TO 33
23      ICONC3=2
        GO TO 33
C       DECODE THE THIRD COLUMN OF THE CONCEPT FOR THE CARD TASK
C       (COLUMN 22)
30      IF(J(3)-4200)33,31,32
31      ICONC3=1
        GO TO 33
32      ICONC3=2
        GO TO 33
C       NOW THE CONCEPT (1-3 DIMENSIONS) IS CODED IN ONES AND TWOS.
C       NOW CHECK TO SEE WHICH DIMENSIONS WERE USED
C                   WAS THE FIRST DIMENSION USED
33      IF(ICONC1)3502, 35,34
34      IWAS1=1
        GO TO 36
35      IWAS1=2
C                   WAS THE SECOND DIMENSION USED
36      IF(ICONC2)3502, 38,37
37      IWAS2=1
        GO TO 39
38      IWAS2=2
C                   WAS THE THIRD DIMENSION USED
39      IF(ICONC3)3502, 392,391
391     IWAS3=1
        GO TO 40
392     IWAS3=2
C        DECODE THE FOCUS FOR THE CARD TASK (COLUMNS 26-28)
40      IF(J(7)-7100)3503,401,402
401     IFOCS1=1
        GO TO 41
402     IFOCS1=2
41      IF(J(8)-4700)3503,411,412
411     IFOCS2=1
        GO TO 42
412     IFOCS2=2
42      IF(J(9)-4200)3503,421,422
421     IFOCS3=1
        GO TO 100
422     IFOCS3=2
        GO TO 100
C       LOOK FOR A TRIAL NUMBER, BUT FIRST LOOK FOR A BLANK
100     IF(IFELD1(IFIELD))4102,1000,1499
1000    ISKIP=2
        GO TO (1494,1496),IPRINT
1494    IF(ITRIAL-6)1495,4001,4001
1495    PRINT 1002
1002    FORMAT(44HEXPERIMENTER TERMINATED THE PROBLEM TOO SOON)
        GO TO 4001
1496    IF(ITRIAL-6)1497,1498,1498
1497    IITERM=IITERM+1
1498    IUNNEC=IINEFF+IREDUN+IREPET
        IINAPP=ICONBI+IINCON+IVM+IVPR
        IEXBEH=IP1+IP2+IP3+IM1+IM2+IM3+IE1+IE2+IPW
        PUNCH 9555,A,            (J(I),I=1,5),(J(I),I=7,11),IEFFIC,
     1      IINEFF,IREDUN,IREPET,ICONBI,IINCON,IVM,IVPR,IP1,IP2,IP3,IM1
```

```
     2             IM2,IM3,IE1,IE2,IPW,IRC,IITERM,ISOLVE,IUNNEC,IINAPP,
     3             IEXBEH
      GO TO 4001
C     DECODE A FIELD FOR THE CARD TASK
1499  IF(IFELD1(IFIELD)-7100)106,101,1500
1500  IF(IFELD1(IFIELD)-7200)4102,101,1501
1501  IF(IFELD1(IFIELD)-7300)4102,101,1502
1502  IF(IFELD1(IFIELD)-7400)4102,101,1503
1503  IF(IFELD1(IFIELD)-7500)4102,101,1504
1504  IF(IFELD1(IFIELD)-7600)4102,101,4102
C     TRANSFER TO 101 INDICATES THIS WAS A TRIAL
101   ITRIAL=ITRIAL+1
      IF(IFELD2(IFIELD)-7100)3504,1011,1012
1011  ITRIL1 (ITRIAL)=1
      GO TO 102
1012  ITRIL1 (ITRIAL)=2
102   IF(IFELD3(IFIELD)-4700)3504,1021,1022
1021  ITRIL2 (ITRIAL)=1
      GO TO 103
1022  ITRIL2 (ITRIAL)=2
103   IF(IFELD4(IFIELD)-4200)3504,1031,1032
1031  ITRIL3 (ITRIAL)=1
      GO TO 104
1032  ITRIL3 (ITRIAL)=2
104   IF(IFELD5(IFIELD)-1000)3504,1041,1042
1041  ITRIL4 (ITRIAL)=1
      GO TO 3007
1042  ITRIL4 (ITRIAL)=2
C     NOW SET AN INDICATOR - ONE FOR PLUS, TWO FOR MINUS.
3007  ICHECK=1
3008  GO TO (3009,6007),IWAS1
3009  IF(ICONC1-ITRIL1(ITRIAL))3010,6007,3010
3010  ICHECK=2
6007  GO TO (3011,6008),IWAS2
3011  IF(ICONC2-ITRIL2(ITRIAL))3012,6008,3012
3012  ICHECK=2
6008  GO TO (3013,3015),IWAS3
3013  IF(ICONC3-ITRIL3(ITRIAL))3014,3015,3014
3014  ICHECK=2
C     NOW CHECK THE SIGN
3015  IF (ICHECK - ITRIL4 (ITRIAL)) 3016, 1050, 3016
C     IF CONTROL REACHES 3016 THEN THE TRIAL WAS MISCODED.
3016  PRINT 3017, ITRIAL
3017  FORMAT  (6HTRIAL ,I1,51H WAS MISCODED BY THE EXPERIMENTER OR THE K
     1EYPUNCHER)
      PRINT 1,A,J
      ISKIP=2
      GO TO 4001
C                          TRIAL  EVALUATION
C     THE CONCEPT TABLE IS AS FOLLOWS
C     ICCEP(1) CONTAINS NONE OF THE ATTRIBUTES (I.E. ALL ELEMENTS)
C     ICCEP(2) CONTAINS ONLY THE FIRST ATTRIBUTE
C     ICCEP(3) CONTAINS ONLY THE SECOND ATTRIBUTE
C     ICCEP(4) CONTAINS ONLY THE THIRD ATTRIBUTE
C     ICCEP(5) CONTAINS THE FIRST AND SECOND ATTRIBUTES
C     ICCEP(6) CONTAINS THE FIRST AND THIRD ATTRIBUTES
C     ICCEP(7) CONTAINS THE SECOND AND THIRD ATTRIBUTES
```

```
C     ICCEP(8) CONTAINS ALL THREE ATTRIBUTES
C     HOW MANY CONCEPTS ARE ON THE TABLE
1050  ILEFT1=0
      DO 4006 I=1,8
C     ICCEP(I)=0 FOR ON THE TABLE, AND ICCEP(I)=2 FOR OFF THE TABLE
      IF (ICCEP(I)) 3513, 4005, 4006
4005  ILEFT1=ILEFT1+1
4006  CONTINUE
C     NOW ILEFT1 CONTAINS THE NUMBER OF CONCEPTS REMAINING ON THE TABLE
C     WHICH ATTRIBUTES WERE CHANGED IN GOING FROM THE FOCUS
C     TO THE TRIAL
1120  IF(IFOCS1-ITRIL1(ITRIAL))1122,1121,1122
1121  IATBT1=1
      GO TO 1123
1122  IATBT1=2
1123  IF (IFOCS2 - ITRIL2 (ITRIAL)) 1125, 1124, 1125
1124  IATBT2=1
      GO TO 6036
1125  IATBT2=2
6036  IF (IFOCS3 - ITRIL3 (ITRIAL)) 1127, 1126, 1127
1126  IATBT3=1
      GO TO 1128
1127  IATBT3=2
C     NOW WE KNOW WHICH ATTRIBUTES WERE CHANGED IN GOING FROM THE FOCUS
C     TO THE TRIAL
C     LOOK TO SEE IF THE TRIAL WAS POSITIVE OR NEGATIVE
1128  IF(ITRIL4(ITRIAL)-1)3517,1129,1100
C     TRANSFER TO 1100 INDICATES THE TRIAL WAS NEGATIVE.
C     WE MUST NOW DELETE ALL CONCEPTS IN THE TABLE THAT DO NOT CONTAIN
C     AT LEAST ONE (OR SOME OR ALL) OF THE ATTRIBUTES CHANGED
C     DEVELOP BRANCHING FOR TWO AND THREE VALUE CHANGES
1100  JDFRNC=0
      GO TO (4271,4270),IATBT1
4270  JDFRNC=JDFRNC+1
4271  GO TO (4273,4272),IATBT2
4272  JDFRNC=JDFRNC+2
4273  GO TO (4275,4274),IATBT3
4274  JDFRNC=JDFRNC+4
C     NOW BRANCH
4275  GO TO (4276,4276,4277,4276,4278,4279,4280),JDFRNC
4277  ICCEP(1)=2
      ICCEP(4)=2
      GO TO 1135
4278  ICCEP(1)=2
      ICCEP(3)=2
      GO TO 1135
4279  ICCEP(1)=2
      ICCEP(2)=2
      GO TO 1135
4280  ICCEP(1)=2
      GO TO 1135
4276  GO TO (1102, 1101), IATBT1
1101  ICCEP(1)=2
      ICCEP(3)=2
      ICCEP(4)=2
      ICCEP(7)=2
1102  GO TO (1104, 1103), IATBT2
```

```
1103  ICCEP(1)=2
      ICCEP(2)=2
      ICCEP(4)=2
      ICCEP(6)=2
1104  GO TO (1135, 1105), IATBT3
1105  ICCEP(1)=2
      ICCEP(2)=2
      ICCEP(3)=2
      ICCEP(5)=2
      GO TO 1135
C     TRANSFER TO 1129 INDICATES THE TRIAL WAS POSITIVE.
C     WE MUST NOW DELETE ALL CONCEPTS IN THE TABLE THAT CONTAIN THE
C     ATTRIBUTE(S) CHANGED.
1129  GO TO (1131, 1130), IATBT1
1130  ICCEP(2)=2
      ICCEP(5)=2
      ICCEP(6)=2
      ICCEP(8)=2
1131  GO TO (1133, 1132), IATBT2
1132  ICCEP(3)=2
      ICCEP(5)=2
      ICCEP(7)=2
      ICCEP(8)=2
1133  GO TO (1135, 1134), IATBT3
1134  ICCEP(4)=2
      ICCEP(6)=2
      ICCEP(7)=2
      ICCEP(8)=2
C     HOW MANY CONCEPTS REMAIN ON THE TABLE
1135  ILEFT2 = 0
      DO 4008 I=1,8
      IF (ICCEP(I)) 3513, 4007, 4008
4007  ILEFT2=ILEFT2+1
4008  CONTINUE
C     NOW ILEFT2 CONTAINS THE NUMBER OF CONCEPTS REMAINING ON THE TABLE
C     AFTER THE LATEST DELETION
C     NOW WERE ANY CONCEPTS ELIMINATED BY THE TRIAL
      IF (ILEFT1-ILEFT2) 4043, 4009, 4022
C     IF CONTROL REACHES 4009 NO CONCEPTS WERE ELIMINATED
C     NOW WAS IT REDUNDANT OR REPETITIOUS (WAS THERE A DUPLICATION)
C     IT WILL HAVE DUPLICATED THE FOCUS IF IATBT 1,2, AND 3 ARE 1
4009  IF (IATBT1-1) 4044, 4010, 4014
4010  IF (IATBT2-1) 4044, 4011, 4014
4011  IF (IATBT3-1) 4044, 4012, 4014
C     (THEY ARE ALL ONE IF WE REACH 4012)
4012  GO TO (9507,9508),IPRINT
9507  PRINT 4013,ITRIAL
4013  FORMAT (6HTRIAL ,I1, 16H WAS REPETITIOUS)
9508  IREPET=IREPET+1
      GO TO 4000
C     (IT DID NOT DUPLICATE THE FOCUS, SO CHECK IT AGAINST THE
C     PRECEDING TRIALS)
4014  DO 4018 I=1, ITRIAL
C     ARE WE COMPARING THE TRIAL TO ITSELF
      IF (I-ITRIAL) 4015, 4018, 4015
C     NO
4015  IF(ITRIL1(I)-ITRIL1(ITRIAL))4018,4016,4018
```

```
4016  IF (ITRIL2(I) - ITRIL2(ITRIAL)) 4018, 4017, 4018
4017  IF(ITRIL3(I)-ITRIL3(ITRIAL))4018,4020,4018
C     IF CONTROL REACHES 4018 THEN A DUPLICATION DID NOT OCCUR
C     THEREFORE THE TRIAL WAS REDUNDANT
4018  CONTINUE
      GO TO (9509,9510),IPRINT
9509  PRINT 4019,ITRIAL
4019  FORMAT(6HTRIAL ,I1,14H WAS REDUNDANT)
9510  IREDUN=IREDUN+1
      GO TO 4000
C     IF CONTROL REACHES 4020 THEN A DUPLICATION DID OCCUR
4020  GO TO (9511,9512),IPRINT
9511  PRINT 4021,ITRIAL
4021  FORMAT(6HTRIAL ,I1,16H WAS REPETITIOUS)
9512  IREPET=IREPET+1
      GO TO 4000
C     AT LEAST ONE CONCEPT WAS ELIMINATED IF CONTROL REACHES 4022
C     COMPARE WITH THE FOCUS
4022  IDIFFR=0
      IF(IATBT1-2)4024,4023,4023
4023  IDIFFR=IDIFFR+1
4024  IF(IATBT2-2)4026,4025,4026
4025  IDIFFR=IDIFFR+1
4026  IF(IATBT3-2)4028,4027,4028
4027  IDIFFR=IDIFFR+1
C     DID MORE THAN ONE VALUE CHANGE
4028  IF(IDIFFR-1)4045,4029,4031
C     ONLY ONE VALUE CHANGED
4029  GO TO (9513,9514),IPRINT
9513  PRINT 4030,ITRIAL
4030  FORMAT(6HTRIAL ,I1,14H WAS EFFICIENT)
9514  IEFFIC=IEFFIC+1
      GO TO 4000
C     MORE THAN ONE VALUE CHANGED, SO CHECK AGAINST SUCCESSIVE POSITIVE
C     TRIALS
4031  DO 4041 I=1,ITRIAL
      IF(ITRIL4(I)-1)4054,4053,4041
4053  IDIFFR=0
C     DO NOT COMPARE A TRAIL TO ITSELF
      IF(I-ITRIAL)4032,4041,4032
4032  IF(ITRIL1(I)-ITRIL1(ITRIAL))4033,4034,4033
4033  IDIFFR=IDIFFR+1
4034  IF(ITRIL2(I)-ITRIL2(ITRIAL))4035,4036,4035
4035  IDIFFR=IDIFFR+1
4036  IF(ITRIL3(I)-ITRIL3(ITRIAL))4037,4038,4037
4037  IDIFFR=IDIFFR+1
C     NOW CHECK TO SEE IF ONLY ONE VALUE CHANGED
4038  IF(IDIFFR-1)4046,4039,4041
C     ONLY ONE VALUE CHANGED
4039  GO TO (9515, 9516),IPRINT
9515  PRINT 4040,ITRIAL
4040  FORMAT(6HTRIAL ,I1,14H WAS EFFICIENT)
9516  IEFFIC=IEFFIC+1
      GO TO 4000
4041  CONTINUE
C     IF CONTROL REACHES HERE ALL TRIALS DIFFERED BY MORE THAN ONE VALUE
C     OR ALL PRECEDING TRIALS WERE NEGATIVE
C     (THEREFORE THE TRIAL WAS INEFFICIENT)
```

```
      GO TO (9517,9518),IPRINT
9517  PRINT 4042,ITRIAL
4042  FORMAT(6HTRIAL ,I1,16H WAS INEFFICIENT)
9518  IINEFF=IINEFF+1
      GO TO 4000
C                        NON-TRIAL  EVALUATION
C                     TRANSFER TO 106 INDICATES THIS WAS NOT A TRIAL,
C                     I.E. IT IS EXPERIMENTER RESPONSES, RC, VERBAL
C                     BEHAVIOR, VM, OR VPR.
106   IF(IFELD1(IFIELD)-5900)108,6037,110
6037  GO TO (9519, 9520),IPRINT
9519  PRINT 6038
6038  FORMAT(27HSUBJECT REFUSES TO CONTINUE)
      GO TO 4001
9520  IRC=IRC+1
      IUNNEC=IINEFF+IREDUN+IREPET
      IINAPP=ICONBI+IINCON+IVM+IVPR
      IEXBEH=IP1+IP2+IP3+IM1+IM2+IM3+IE1+IE2+IPW
      PUNCH 9555,A,                 (J(I),I=1,5),(J(I),I=7,11),IEFFIC,
     1    IINEFF,IREDUN,IREPET,ICONBI,IINCON,IVM,IVPR,IP1,IP2,IP3,IM1,
     2             IM2,IM3,IE1,IE2,IPW,IRC,IITERM,ISOLVE,IUNNEC,IINAPP,
     3             IEXBEH
      GO TO 4001
C                     TRANSFER TO 108 INDICATES THIS WAS AN EXPERIMENTER
C                     RESPONSE
108   IF(IFELD1(IFIELD)-5400)130,131,132
C                     TRANSFER TO 110 INDICATES THIS WAS VERBAL BEHAVIOR,
C                     VM, OR VPR
110   IVERB1=0
      IVERB2=0
      IVERB3=0
      IF(IFELD2(IFIELD)-7100)112,4400,4401
4400  IVERB1=1
      GO TO 120
4401  IVERB1=2
      GO TO 120
C     TRANSFER TO 112 INDICATES THIS WAS VM OR VERBAL BEHAVIOR (BUT NOT
C     ONE OR TWO)
112   IF(IFELD2(IFIELD)-5700)114,4302,4304
4302  GO TO (9521, 9522),IPRINT
9521  PRINT 4303
4303  FORMAT(40HSUBJECT GIVES A VERBAL POINTING RESPONSE)
9522  IVPR=IVPR+1
      GO TO 4000
4304  IVERB2=2
      GO TO 120
C     TRANSFER TO 114 INDICATES THIS WAS VM OR VERBAL BEHAVIOR (BUT NOT
C     RED, ONE, OR TWO OR V.P.R.)
114   IF(IFELD2(IFIELD)-4700)115,4305,4306
4305  IVERB2=1
      GO TO 120
4306  GO TO (9523,9524),IPRINT
9523  PRINT 4307
4307  FORMAT(33HSUBJECT GIVES IRREGULAR STATEMENT)
9524  IVM=IVM+1
      GO TO 4000
C     TRANSFER TO 115 INDICATES THIS WAS EITHER BOXES OR DIAMONDS OR ALL
115   IF(IFELD2(IFIELD)-4200)9101,4308,4309
```

```
9101   IVERB4=1
C      IS THIS VERBAL BEHAVIOR--ALL--THE SAME AS THE CONCEPT
       IF(IVERB4-ICONC4)4260,4253,4260
4308   IVERB3=1
       GO TO 120
4309   IVERB3=2
C                      LOOK FOR SECOND DIMENSION IN VERBAL BEHAVIOR
120    IF(IFELD3(IFIELD)-7100)4209,4207,4208
4207   IVERB1=1
       GO TO 125
4208   IVERB1=2
       GO TO 125
4209   IF(IFELD3(IFIELD)-4700)4212,4210,4211
4210   IVERB2=1
       GO TO 125
4211   IVERB2=2
       GO TO 125
4212   IF(IFELD3(IFIELD)-4200)4250,4213,4214
4213   IVERB3=1
       GO TO 125
4214   IVERB3=2
       GO TO 125
C                      LOOK FOR THIRD DIMENSION IN VERBAL BEHAVIOR
125    IF(IFELD4(IFIELD)-7100)4217,4215,4216
4215   IVERB1=1
       GO TO 4250
4216   IVERB1=2
       GO TO 4250
4217   IF(IFELD4(IFIELD)-4700)4220,4218,4219
4218   IVERB2=1
       GO TO 4250
4219   IVERB2=2
       GO TO 4250
4220   IF(IFELD4(IFIELD)-4200)4250,4221,4222
4221   IVERB3=1
       GO TO 4250
4222   IVERB3=2
       GO TO 4250
C      IS THE VERBAL BEHAVIOR THE SAME AS THE CONCEPT
4250   IF(IVERB1-ICONC1)4260,4251,4260
4251   IF(IVERB2-ICONC2)4260,4252,4260
4252   IF(IVERB3-ICONC3)4260,4253,4260
C      YES IF WE REACH 4253--SO HOW MANY CONCEPTS REMAIN ON THE TABLE
4253   IF(ILEFT2-1)4300,4256,4258
4256   ISKIP=2
       GO TO (9525,9526),IPRINT
9525   PRINT 4257
4257   FORMAT(30HTHE SUBJECT SOLVED THE PROBLEM)
       GO TO 4001
9526   ISOLVE=ISOLVE+1
       IUNNEC=IINEFF+IREDUN+IREPET
       IINAPP=ICONBI+IINCON+IVM+IVPR
       IEXBEH=IP1+IP2+IP3+IM1+IM2+IM3+IE1+IE2+IPW
       PUNCH 9555,A,            (J(I),I=1,5),(J(I),I=7,11),IEFFIC,
      1   IINEFF,IREDUN,IREPET,ICONBI,IINCON,IVM,IVPR,IP1,IP2,IP3,IM1,
      2           IM2,IM3,IE1,IE2,IPW,IRC,IITERM,ISOLVE,IUNNEC,IINAPP,
      3           IEXBEH
       GO TO 4001
```

ERIC

```
4258  GO TO (9527,9528),IPRINT
9527  PRINT 4259
4259  FORMAT(46HSUBJECT GIVES CONSISTENT-BUT-INVALID STATEMENT)
9528  ICONBI=ICONBI+1
      GO TO 4000
4260  GO TO (9529,9530),IPRINT
9529  PRINT 4261
4261  FORMAT(36HSUBJECT GIVES INCONSISTENT STATEMENT)
9530  IINCON=IINCON+1
      GO TO 4000
C                     TRANSFER TO 130 INDICATES THIS WAS AN ELICIT
130   IF(IFELD2(IFIELD)-7100)3518,6046,6048
6046  GO TO (9531,9532),IPRINT
9531  PRINT 6047
6047  FORMAT(30HEXPERIMENTER ELICITS STATEMENT)
9532  IE1=IE1+1
      GO TO 4000
6048  GO TO (9533, 9534),IPRINT
9533  PRINT 6049
6049  FORMAT(30HEXPERIMENTER ELICITS STATEMENT)
9534  IE2=IE2+1
      GO TO 4000
C                     TRANSFER TO 131 INDICATES THIS WAS A MOTIVATE
131   IF(IFELD2(IFIELD)-7200)6050,6052,6054
6050  GO TO (9535, 9536),IPRINT
9535  PRINT 6051
6051  FORMAT(31HEXPERIMENTER MOTIVATES SLIGHTLY)
9536  IM1=IM1+1
      GO TO 4000
6052  GO TO (9537,9538),IPRINT
9537  PRINT 6053
6053  FORMAT(33HEXPERIMENTER MOTIVATES MODERATELY)
9538  IM2=IM2+1
      GO TO 4000
6054  GO TO (9539,9540),IPRINT
9539  PRINT 6055
6055  FORMAT(31HEXPERIMENTER MOTIVATES STRONGLY)
9540  IM3=IM3+1
      GO TO 4000
C     TRANSFER TO 132 INDICATES THIS WAS A PROMPT OR A PUT IN WORDS--PW
132   IF(IFELD2(IFIELD)-7200)6056,6058,6060
6056  IF(IFELD2(IFIELD)-6600)4102,9130,9132
9130  GO TO (9541, 9542),IPRINT
9541  PRINT 9131
9131  FORMAT(12HPUT-IN-WORDS)
9542  IPW=IPW+1
      GO TO 4000
9132  GO TO (9543,9544),IPRINT
9543  PRINT 6057
6057  FORMAT(32HEXPERIMENTER GIVES SLIGHT PROMPT)
9544  IP1=IP1+1
      GO TO 4000
6058  GO TO (9545,9546),IPRINT
9545  PRINT 6059
6059  FORMAT(34HEXPERIMENTER GIVES MODERATE PROMPT)
9546  IP2=IP2+1
      GO TO 4000
6060  GO TO (9547,9548),IPRINT
```

```
9547  PRINT 6061
6061  FORMAT(32HEXPERIMENTER GIVES STRONG PROMPT)
9548  IP3=IP3+1
C     THE FOLLOWING FEW STATEMENTS TAKE CARE OF FIELD CHANGING
4000  GO TO (4100,4099),ICHANG
4099  IFIELD=0
      ICHANG=1
4100  CONTINUE
      GO TO 9999
3500  PRINT 6062
6062  FORMAT(38HTHE TASK (COLUMNS 10-12) IS MISPUNCHED)
      PRINT 1,A,J
      ISKIP=2
      GO TO 4001
3501  PRINT 6063
6063  FORMAT(57HTHE FIRST COLUMN OF THE CONCEPT (COLUMN 20) IS MISPUNCHE
     1D)
      PRINT 1,A,J
      ISKIP=2
      GO TO 4001
3502  PRINT 6064
6064  FORMAT(47HPROGRAM ERROR IN COUNTING DIMENSIONS IN CONCEPT)
      ISKIP=2
      GO TO 4001
3503  PRINT 6065
6065  FORMAT(39HTHE FOCUS (COLUMNS 26-28) IS MISPUNCHED)
      PRINT 1,A,J
      ISKIP=2
      GO TO 4001
3504  PRINT 6066,ITRIAL
6066  FORMAT(6HTRIAL ,I1,15H WAS MISPUNCHED)
      PRINT 1,A,J
      ISKIP=2
      GO TO 4001
3513  PRINT 6075
6075  FORMAT(21HERROR ONE BEFORE 4005)
      ISKIP=2
      GO TO 4001
3514  PRINT 6076
6076  FORMAT(12HERROR AT 115)
      ISKIP=2
      GO TO 4001
3515  PRINT 6077
6077  FORMAT(41HERROR THREE BEFORE 1282, OR 1283, OR 1286)
      ISKIP=2
      GO TO 4001
3516  PRINT 6078
6078  FORMAT(21HERROR AT 1305 OR 1306)
      ISKIP=2
      GO TO 4001
3517  PRINT 6079
6079  FORMAT(24HA TRIAL SIGN WAS MISSING)
      PRINT 1,A,J
      ISKIP=2
      GO TO 4001
3518  PRINT 6080
6080  FORMAT(24HAN ELICIT WAS MISPUNCHED)
```

```
      PRINT 1,A,J
      ISKIP=2
      GO TO 4001
3519  PRINT 6081
6081  FORMAT(21HERROR ONE BEFORE 6000)
      ISKIP=2
      GO TO 4001
3520  PRINT 3521
3521  FORMAT(13HERROR AT 1273)
      ISKIP=2
      GO TO 4001
4043  PRINT 4047
4047  FORMAT(21HERROR FOUR AFTER 4008)
      ISKIP=2
      GO TO 4001
4044  PRINT 4048
4048  FORMAT(23HERROR AROUND OR AT 4010)
      ISKIP=2
      GO TO 4001
4045  PRINT 4049
4049  FORMAT(13HERROR AT 4028)
      ISKIP=2
      GO TO 4001
4046  PRINT 4050
4050  FORMAT(13HERROR AT 4038)
      ISKIP=2
      GO TO 4001
4054  PRINT 4055
4055  FORMAT(20HERROR ONE AFTER 4031)
      ISKIP=2
      GO TO 4001
4102  PRINT 4103,IFIELD
4103  FORMAT(6HFIELD ,I1,15H WAS MISPUNCHED)
      PRINT 1,A,J
      ISKIP=2
      GO TO 4001
4300  PRINT 4301
4301  FORMAT(13HERROR AT 4253)
      ISKIP=2
      GO TO 4001
5058  PRINT 5059
5059  FORMAT(23HCOLUMN 32 IS MISPUNCHED)
      PRINT 1,A,J
      ISKIP=2
      GO TO 4001
4001  GO TO 4002
9999  CONTINUE
      END
```

Appendix II

Intercorrelations Among Aptitude Tests

	1	2	3	4	5	6
1	----					
2	4878	----				
3	4185	5232	----			
4	2816	2421	3634	----		
5	3501	2929	3246	1166	----	
6	2868	6343	5129	2253	3264	----
7	2419	3518	2847	3038	0924	5559
8	3116	3883	5845	3174	1692	5250
9	2772	3277	5414	2488	3172	3761
10	3214	3004	3415	4098	1620	2780
11	1877	3259	4367	2603	1481	4003
12	3529	5696	5803	3124	3040	6142
13	3501	4365	5322	3026	2860	4284
14	2843	5637	4958	4043	2794	6195
15	1527	3125	3237	2135	2075	3756
16	0422	1919	1949	1984	1140	1945
17	2638	2862	4021	2431	2723	3491
18	0330	0738	1163	2205	0819	1445
19	1696	4633	3609	1957	2056	4949
20	1349	3294	4319	1979	1658	3717
21	1522	1365	2637	-0110	1906	2214
22	1643	3044	2886	1250	2060	3957
23	2000	1887	2094	1208	1737	1784
24	0790	1283	2648	1323	1116	1123

Intercorrelations Among Aptitude Tests (continued)

	7	8	9	10	11	12
1						
2						
3						
4						
5						
6						
7	----					
8	3968	----				
9	3271	4882	----			
10	3799	3494	3631	----		
11	2577	4751	2483	3124	----	
12	3912	5706	4338	3300	4726	----
13	3324	4686	4048	3520	4293	4542
14	3591	5325	3738	3115	4199	5477
15	1937	2410	3844	0746	1319	3313
16	1173	2319	1784	1263	2511	2365
17	1968	2811	3312	3005	2251	2832
18	1621	1992	0601	2027	1394	0546
19	3543	3584	2912	2222	2469	4581
20	2248	3903	2990	1733	2447	4433
21	1619	2070	2417	2090	2480	2184
22	2048	3374	2602	2090	1783	3126
23	1030	2293	2115	1439	0985	1812
24	1912	2456	2568	1710	1621	1809

Intercorrelations Among Aptitude Tests (continued)

	13	14	15	16	17	18
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13	----					
14	4968	----				
15	3317	4478	----			
16	1054	1869	2738	----		
17	3660	3637	5075	2625	----	
18	0610	1774	1380	1859	1941	----
19	3429	4455	2820	1159	2936	0437
20	3695	3407	4339	2824	3077	2239
21	2481	3722	1994	0265	1517	-0280
22	3127	3818	2325	1638	2678	1492
23	2383	2998	3913	1505	4780	2215
24	1992	2293	2178	1554	2235	2913

Intecorrelations Among Aptitude Tests (continued)

	19	20	21	22	23	24
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19	----					
20	1895	----				
21	2277	0526	----			
22	3436	2667	1151	----		
23	1970	2065	1045	1621	----	
24	0392	2882	-0168	3368	2628	----

NOTE:--Decimal points omitted.

Appendix III

Publications, programs, technical reports and theses prepared in whole or in part under Contract OE 5-10-299

Anderson, Richard C. Individual differences and problem solving. In Robert M. Gagne´, Editor, Learning and Individual Differences. Columbus, Ohio: Merrill Books, 1967. pp. 66-89. (Also Chapter 3).

Anderson, Richard C. Manual for presenting problems. Mimeo. 1965. 7 pp.

Anderson, Richard C. Note on an experiment to teach first graders a problem solving skill. Mimeo. 1965. 14 pp.

Anderson, Richard C. Part vs. whole task procedures for teaching children a problem solving skill. Journal of Educational Psychology. In press. (Also Chapter 6).

Anderson, Richard C., Boone, Julianne, Daniel, Jessica. Common elements program for the card array task. University of Illinois Training Research Laboratory Monograph, 1965, 1, 1-200.

Anderson, Richard C., Yavin, Rivka, and Daniel, Jessica. Common elements program for the pencil task. University of Illinois Training Research Laboratory Monograph, 1965, 3, 1-200.

Anderson, Richard C., and Yavin, Rivka. Common elements program for the cowboy task. University of Illinois Training Research Laboratory Monograph, 1965, 4, 1-200.

Anderson, Richard C., and Yavin, Rivka. Whole-training program for the cowboy task. University of Illinois Training Research Laboratory Monograph, 1965, 5, 1-112.

Anderson, Richard C., and Yavin, Rivka. Whole-training program for the pencil task. University of Illinois Training Research Laboratory Monograph, 1965, 6, 1-112.

Anderson, Richard C., and Yavin, Rivka. Whole-training program for the card array task. University of Illinois Training Research Laboratory Monograph, 1965, 7, 1-112.

Guthrie, John T. The function of rules in learning, retention, and transfer. Master's thesis. University of Illinois, 1966. (Also in Journal of Educational Psychology, 1967, 58, 45-49.)

Rubovits, James J. Expository versus discovery methods in acquisition and transfer of principles. Master's thesis. University of Illinois, 1968.

Willis, Richard C., and Anderson, Richard C. Dimensions of ability among Mattoon, Illinois fourth grade students. Mimeo, 1965. 9 pp.

OE 6000 (REV. 9-66)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

OFFICE OF EDUCATION

ERIC REPORT RESUME

(TOP)

ERIC ACCESSION NO.

CLEARINGHOUSE ACCESSION NUMBER

RESUME DATE	P.A.	T.A.
01-31-68		

IS DOCUMENT COPYRIGHTED? Chapter 3 only YES [X] NO [ ]

ERIC REPRODUCTION RELEASE? YES [X] NO [ ]

001

TITLE

100 An Analysis of a Class of Problem Solving Behavior.

101 Final Report

PERSONAL AUTHOR(S)

200 Richard C. Anderson

INSTITUTION (SOURCE) — SOURCE CODE

300 University of Illinois, Urbana, Illinois, Training Research Laboratory

310 REPORT/SERIES NO. N.A.

OTHER SOURCE — SOURCE CODE

320 N.A.

330 OTHER REPORT NO. N.A.

OTHER SOURCE — SOURCE CODE

340 N.A.

350 OTHER REPORT NO. N.A.

400 PUB'L. DATE Jan. 68 — CONTRACT/GRANT NUMBER OE 5-10-299

PAGINATION, ETC.

500 181 pp. OE Project 2942

RETRIEVAL TERMS

IDENTIFIERS

ABSTRACT

Completed was a theoretical and experimental analysis of a strategy for solving concept attainment problems, the strategy of varying each factor in succession while holding all other factors constant. Standardized procedures for presenting and scoring problems were developed. A computer program was prepared for the content analysis of problem solving protocols. Three experiments were completed with first, second, third, and fourth graders. An experiment with fourth graders investigated the effects on problem solving performance of order of presentation, type of materials, number of stimulus dimensions defining the concept to be attained, and the specific characteristics of the concept to be attained. Also investigated were the relationships between seven aptitude factors and problem solving performance. Modest but significant correlations were obtained for Spatial Orientation and Figural Adaptive Flexibility. The data from a second experiment seemed to indicate that concrete stimulus objects are important in the problem solving of children because they serve as memory aids rather than because they help the child to select logically appropriate instances. A final experiment demonstrated that first graders can be taught to solve concept attainment problems at a high level of proficiency. While some improvement in performance occurred with simple practice at problem solving, best results were obtained with a small-step, programed procedure.

ERIC

E-BR

must

FROM:

ERIC FACILITY

SUITE 601

1735 EYE STREET, N. W.

WASHINGTON, D. C. 20006

ERIC