72-23,093

DANNEFFEL, George J., 1910-
A COMPARISON OF THE NON-LANGUAGE MULTIPLE-CHOICE PROBLEM SOLVING ABILITY OF VERY BRIGHT, AVERAGE, AND HIGH-GRADE MENTALLY DEFECTIVE EIGHT-YEAR-OLD CHILDREN.

The University of Oklahoma, Ph.D., 1972
Education, special

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A COMPARISON OF THE NON-LANGUAGE MULTIPLE-CHOICE PROBLEM SOLVING ABILITY OF VERY BRIGHT, AVERAGE, AND HIGH-GRADE MENTALLY DEFECTIVE EIGHT-YEAR-OLD CHILDREN

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

GEORGE J. DANNEFFEL

Norman, Oklahoma

1972

A COMPARISON OF THE NON-LANGUAGE MULTIPLE-CHOICE PROBLEM

SOLVING ABILITY OF VERY BRIGHT, AVERAGE, AND

HIGH-GRADE MENTALLY DEFECTIVE

EIGHT-YEAR-OLD CHILDREN

APPROVED BY

B. T. Tesha

P. J. Pupiper

Mary Clare Petty

Lloyd P. Williams

DISSERTATION COMMITTEE

ACKNOWLEDGMENTS

The writer wishes to express his sincere appreciation to Dr. P. T. Teska, Chairman of his committee, who not only directed this study but whose confidence and guidance inspired the writer to complete his graduate program. Gratitude is also extended to Dr. Omer J. Rupiper whose suggestions in the statistical treatment were most valuable; to Dr. Mary Clare Petty for her cooperation; and to Dr. Lloyd P. Williams whose keen observations provided the incentive for the writer to continue his program.

The writer expresses thanks to Principal Bill Martin, Jeanne Putman, and to others of the Edmond schools. Thanks is extended to the many administrators, teachers, and pupils from the schools in Norman and vicinity who cooperated in this study.

Special appreciation is extended to Linda Maffucci who administered the intelligence tests. Sincere appreciation is also extended to Dr. Bonnie K. Roberts for her valuable encouragement; other members of the special education staff at the University of Oklahoma; and to Ethel Warner who assisted in locating subjects.

Finally, the author expresses sincere appreciation to his daughters: Mary Beth, Betty Ann, and Sue Ellen and to his son, George John, for their continued understanding and inspiration during the past several years. Appreciation is extended to his sister, Eleanor C. Stanhiser, for her encouragement and understanding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

A COMPARISON OF THE NON-LANGUAGE MULTIPLE-CHOICE PROBLEM SOLVING ABILITY OF VERY BRIGHT, AVERAGE, AND HIGH-GRADE MENTALLY DEFECTIVE EIGHT-YEAR-OLD CHILDREN

CHAPTER I

INTRODUCTION

The problem solving performance of children has been analyzed and compared by many investigators. Several different methods and devices have been used for the purpose of analyzing this complex process. Tests of problem solving ability in which reading and enriched vocabulary play a large part are likely to be prejudiced in favor or subjects with higher intelligence and against those from deprived areas. The Problem Box was designed by Teska (1942) in an effort to hold these factors to a minimum while measuring the problem solving ability of children or adults from various cultures and environments.

Comparative studies of problem solving ability have been made comparing such factors as mental age, chronological age, intelligence, and socio-economic level. Researchers generally agree that problem solving ability is least affected by chronological age. Few investigators have compared groups of the same chronological age having different levels of intelligence. The interest in research in problem solving declined around the early part of the twentieth century. The interest of investigators in problem solving has gradually increased during the last few years, especially since 1960. Problem solving is a significant function of human behavior appropriate for further investigation.

Steps in Problem Solving

The five classical steps in problem solving were initially proposed by Dewey (1910): first, a felt difficulty or uneasiness; second, its location and definition; third, suggestions for a possible solution; fourth, development of the bearing of the suggestions by reasoning; and fifth, further observation and experimentation leading to verification or rejection of the proposed solution. Dewey (1933) restated and elaborated upon these five phases, observing that the sequence is not fixed. In practice one of the steps may be passed over, one phase may be expanded to include subphases, or two phases may telescope.

Symonds (1936) defined the process in four steps: first, isolating and defining the values that operate; second, proposing a variety of solutions; third, estimating the consequences of the various alternatives; and fourth, making decisions based on proposed solutions whose outcomes have most bearing on or relationship to the values at stake, and based on the probability that certain outcomes would result.

Vinacke (1952) stated that only three stages may be logically distinguished: first, confrontation by a problem; second, working toward a solution; and third, solution. His analysis of stage two includes steps two, three, and four as defined by Dewey.

D. M. Johnson (1955) said that in a very general way, problem solving consists of three steps: first, survey of the problem and analysis of the goal; second, production of probable solution attempts; and third, evaluation and judgment of the attempts. He suggests that if the only solution produced is obviously correct, no separate act of evaluation is necessary.

Merrifield et al. (1962) developed a modified theory of problem solving involving five phases as a result of study designed for this purpose: first, preparation; second, analysis; third, production; fourth, verification; and fifth reapplication. The writers stated that much backtracking can take place at almost any phase whenever progress is unsatisfactory. Phase five is only necessary when a tentative solution has been rejected. When the solution is found to succeed, the problem no longer exists.

Guilford (1967) proposed a general problem solving operational model, based upon the above study in which he was a co-author, that resembles a flow-chart. The five principal operations are: first, filtering (attention aroused and directed); second, cognition (problem sensed and structured); third, production (answers generated); fourth, cognition (new information obtained); and fifth, production (new answers generated).

He suggests an exit after each of the five operations to indicate cessation of the process. The first exit may be a rejection of the problem; the second exit may be a postponement or a problem impossible to solve, and the other exits may mean a satisfactory solution to the problem has occurred. A liberal allowance is made for looping phenomena that provides for feedback information and permits some flexibility in the order of events.

In a survey of the literature, D'Zurilla et al. (1971) found a remarkable degree of agreement among several theorists and investigators in the general kinds of operations involved in effective problem solving. Disregarding some minor variations in wording and categorization, the consensus viewpoint is closely represented by the following five general stages: first, general orientation (set and attitudinal factors); second, problem definition and formulation; third, generation of alternatives; fourth, decision making (evaluation and selection); and fifth, verification. The writers suggest that problem solving seldom proceeds in these neatly ordered stages but more typically the stages overlap and interact with each other.

In view of the various descriptions of the problem solving process cited above, it is apparent that the first step is one of recognition. A problem has to be recognized as such and there has to be a need or desire for a solution on the part of the individual. The word "problem" implies choice. If there are no alternatives or choices available and understood by the individual, a problem does not exist for him.

Thus, a problem may be presented by an investigator but the subject may not be aware that a problem exists. Dewey (1910) feels that at some point problem solving has to be a conscious process in order for one to reflect upon the factors involved. There is perhaps a synthesis in problem solving of the conscious and the unconscious. The unconscious lends to the process spontaneity and freshness; consciousness gives to the process a measure of conviction and control. If a solution is perceived along with the recognition of the problem, either a problem did not exist or the other steps in the process were omitted.

The second step is one of induction. It involves the location, isolation, definition and organization of the various elements or clues in search for generalizations. Isolation means to identify and separate the various clues. Definition is recognizing the values of the clues as either relevant to some degree or irrelevant. The relevant clues or elements are organized into some meaningful wholes or generalizations. Organizing the data into related wholes leads to a better understanding of the data so that the process of hypothesizing is less difficult and the suggested hypotheses are more likely to be relevant to the solution of the problem. Dewey (1910) says, "The meaning suggested supplies a mental platform, an intellectual point of view, from which to note and define the data more carefully, to seek for additional observations, and to institute, experimentally, changed conditions" [p. 79].

The third step is to discover or to formulate the hypotheses. About this Symonds (1936) says, "The matter of proposing hypotheses is in one sense the very heart of problem solving, particularly of the more constructive or creative sort" [p. 126]. The data must be logical in nature and organized to the point where meaningful constellations are generated or the proposed hypotheses may not be appropriate to the solution of the problem.

The fourth step is syllogistic in form, thus deductive in nature. It is the reasoning process of testing the validity of the generated hypotheses or generalizations against the organized data in search of a possible solution. This not only involves a preliminary check of the correctness of the proposed hypotheses but also leads to a better understanding of the data. If the testing proves the proposed hypothesis to be wrong, other previously proposed hypotheses are tested or the data is reorganized in order to generate new and better hypotheses which is a return to the inductive reasoning step with a better understanding of the data.

The fifth step is one of further observation. This leads to a final acceptance or rejection of the solution.

As a result of this classical description of problem solving, an adequate test of the process can be better formulated so that all of the steps will be required as discretely as possible in order to evaluate the total process. However, problem solving may take place without proceeding in this sequence or through these formal steps. Some steps may be omitted, some steps may be telescoped, or the order may be inverted or transposed. In some instances, such as in trial and error, inductive reasoning is not used since analyzing and organizing the clues are not involved. If the nature of the data is such that it does not lend itself to an adequate test of all relevant hypotheses, then deductive reasoning cannot be utilized. Inductive and deductive reasoning are fundamental in any problem solving process.

Review of the Literature

Most of the early studies using multiple-choice were done with animals. These included studies by Burtt (1916), Coburn and Yerkes (1915), Hamilton (1911), Yerkes (1914, 1917), Yerkes and Coburn (1915) and Yerkes and Rossy (1917).

Yerkes (1921) investigated the ideational behavior of individuals using a modification of his earlier apparatus. The earlier device consisted of an arrangement of twelve wooden keys that could be raised or lowered in any combination by the experimenter. The subject made a choice by depressing one of the lowered keys. When the correct key was depressed a bell would ring. The modifications consisted of replacing the bell with a buzzer and arranging the twelve keys in such a way that they could be extended or withdrawn in any combination. Only the keys projecting out of the keyboard were accessible to be depressed by the subject.

Regardless of the combination of keys extended, the correct solutions of the four problems presented in sequence were; the first key on the left, the first key at the left alternately with the first key on the right, the third key from the left, and the middle key. Two solutions were accepted; pressing the correct key in ten successive trials without formulation of the generalization involved, or pressing the correct key plus verbalization of the principle involved.

Yerkes (1921) compared groups of superior, average, defective, and pathological subjects. The superior group required fewer trials to solve each problem. The average group was next but here was an overlapping between the average group and the defective group. It was pointed out that some subjects were capable of selecting the correct key each time but quite incapable of verbalizing the principle involved in the solution. He suggests that such a solution is in "motor terms," common to average, mentally defective and pathological subjects. In the discussion he stated that regardless of the statistical significance of the data, the responses of each subject are illuminating and considered important in determining the "ideational characteristics." Some subjects eliminated keys systematically, revealing confidence in their ability to solve the problem. Others used a random approach without displaying any confidence in their ability to solve the problem.

The multiple-choice technique of Heidbreder (1928) involved the use of boxes in four problems: the one on the right, the flowered versus plain, the farther with plain figure, and the nearer with red dotted figures. The subjects ranged in chronological age from three to adults. It was found that the number of trials decreased with age, that verbal generalizations were given by all subjects above six, and that the verbalizations of adults were more clear and objective.

Roberts (1932) used a multiple-choice apparatus consisting of a toy house with three different colored doors, a visible toy inside could be obtained by opening the door matching the color of the toy. The subjects were between the ages of two and five. It was found that all subjects solved the problem but none of the subjects below the age of three were able to verbalize their solutions.

Aarons (1933) compared serial learning and generalizing abstraction using Yerkes (1921) multiple-choice technique versus learning the order of cards in a deck. He found a low positive correlation between multiple-choice problem solving and serial learning as measured by learning the order of cards.

Long and Welch (1941) compared the ability of 135 children between the ages of 30 and 83 months to match and discriminate numbers. Three problems were used: selecting the box containing the largest number of marbles in order to get a reward, matching the number of marbles in a single group, and matching the number of marbles in four groups. A steady improvement by age in the ability to discriminate and match numbers was found, but a low positive correlation with intelligence was reported.

House and Zeaman (1958) compared two groups of institutionalized mentally defective children with two groups of public school normal children of professional parents using a multiple-choice technique. The apparatus, modeled after the Wisconsin General Test Apparatus, had a low worktable toward the subject and a square box 72 inches high attached to the back of the worktable. During testing, this box concealed the examiner and the sliding stimulus tray when it was withdrawn from the subject between trials. In the side of the box facing the subject was mounted a one-way mirror. The sliding tray 30 inches wide contained two round food wells 2 inches in diameter and 12 inches apart. These food wells were covered by a pair of stimulus bases made of 4 inch by 4 inch Masonite. While the tray was concealed from the subject the examiner placed candy in the well under the correct response and then pushed the tray in front of the subject for each trial. The correct stimulus card was varied irregularly from left to right for each trial. There were 25 pairs of stimuli consisting of five forms (circle, square, triangle, cross, and T) in five different colors (red, green, yellow, white, and black). Twenty-five trials with about 15 second intervals between trials were given each day. The problem was considered as being solved when the subject made 20 correct choices out of 25 trials during a session. Each subject was placed in one of four groups according to mental age; defective with a M.A. of 4 years, normal with a M.A. of 4 years, defective with a M.A. of 5 years, and normal with a M.A. of 5 years. It was found that as mental age increased, errors decreased for both normals and defectives.

At the mental age level of four years the normals performed significantly better than the defectives, and at the mental age level of five years there was no significant difference. The investigators pointed out that a more difficult problem would be required to discriminate between normals and defectives at the five year mental age level. It was concluded that intelligence is related to learning ability with the mental age controlled.

Harter (1965) compared three groups of children using the same apparatus and procedures as House and Zeaman (1958) described above, except that an opaque screen was used as a shield instead of withdrawing the tray and different stimulus objects were used. These were 100 pairs of "junk" stereometric objects, one of each pair being arbitrarily designated as the "correct" object. A marble placed in the food-well was used as a reinforcer for a correct choice. Each subject continued daily sessions of ten problems (40 trials) until he attained the learning set criterion of five successive problems with no more than one error on trials two, three, and four of all five problems. Each subject was required to verbalize the solution to each problem solved. A total of 81 subjects were used, 42 boys and 39 girls, all in public school classes and never institutionalized. I.Q. scores were used as the basis for assigning subjects to one of three groups according to M.A.: 5 years, low group; 7 years, normal group; and 9 years, high group. The approximate mean I.Q. scores of the groups were: low, 70; normal, 100; and high, above 130.

The mean C.A. for each group was: 7-0, 10-1, and 13-2 for the low; 5-0, 6-9, and 9-1 for the normal; and 3-4, 4-9, and 7-1 for the high. The chronological ages of all subjects ranged from 3-4 to 13-2 years. An analysis of the data revealed that: (1) C.A. contributes very little to learning set performance. (2) M.A. and I.Q. both influenced the learning set as revealed by comparing the number of problems required to form the learning set. (3) Some subjects who reached the criterion were unable to verbalize the solution. (4) All subjects reached the criterion in seven days or less. (5) There were no significant differences between the performances of boys and girls. Harter (1965) pointed out that at each M.A. level, as the I.Q. level increased the learning set acquisition was accelerated. It appeared that some combination of I.Q. and M.A. level would be the best predictor of learning. M.A. level when taken as a developmental measure of cognitive level would be expected to predict speed of learning.

Harter (1967) used the same apparatus described previously [House and Zeaman (1958)] to investigate the role of M.A. and I.Q. by comparing three levels of I.Q.; 65, 100, and 130 and two levels of M.A. $5\frac{1}{2}$ and $8\frac{1}{2}$. The low group, I.Q. level of 65, was represented by both institutionalized and noninstitutionalized subjects. Two social conditions were used, one with the examiner behind the one-way mirror and the other a face-to-face situation with the one-way mirror removed plus friendly conversation before testing and verbal praise for correct solutions. Two equally qualified examiners were used in order to partially control for differences in effectiveness of adults as socially reinforcing agents. The same learning set criterion used by Harter (1965) was used in this study. The finding were: (1) There were no significant sex differences. (2) Learning set acquisition was significantly faster at both the higher I.Q. and M.A. levels. (3) The correlation between C.A. and learning set acquisition was not significant. (4) Learning set acquisition was faster in the noninstitutionalized than in the institutionalized retardates. (5) Learning set acquisition was faster for the lower I.Q. levels without social interaction. (6) Learning set acquisition was considerably faster for the higher I.Q. levels with social interaction, particularly at M.A. 5½.

Moffitt (1969) compared the problem solving performance of two groups of severely retarded children from the Cambridge State School using the same apparatus as House and Zeaman (1958) described previously. The stimuli, 4 inch by 4 inch cardboard bases covered with rough felt, consisted of an eight by eight matrix, eight forms and eight colors. On a day prior to the nine daily experimental sessions consisting of one 30 trial problem, each subject demonstrated that he could find the candy reinforcer. A reinforcement occurred once in each block of three problems. There were three controlled reinforcement conditions; 100%, 66%, and 33%. Subjects were placed in groups on the basis of test information on file; two levels of I.Q. (25-35, 36-46) and two levels of M.A. (4.0-6.9, 7.0-9.9).

The higher I.Q. groups performed significantly better than the lower I.Q. groups under all three percentages of reinforcement. There were no significant differences between the low M.A. groups and the high M.A. groups.

Gozali (1969) investigated the cognitive styles of retarded primary students using a circular version of Kagan's Matching Familiar Figures instrument consisting of 20 pictures of familiar objects. Each item consisted of a picture of an object to be matched in a circle in the center of a larger circle. Spaced around the smaller circle within the larger circle were six pictures of the same object, only one of which matched the details of the picture in the inner circle. Measures of latency of the first response and the total number of errors were recorded. The median response latency was calculated. Those subjects above the median were classified as "reflective" and those below as "impulsive." The impulsive subjects made more errors.

Lehtiö (1970) investigated the order of the importance of clues in the formulation of hypotheses. The stimuli were Ektachrome color slides of 72 combinations of numbers in two sizes (small and large) and two colors (red and blue) with one dot or two dots at the end of each target row. The slides were projected onto a rear-projection screen with an Agfa Diamator 150 projector. For each exposure, the examiner pushed the start button of an electronic timer, the solenoid-operated shutter opened and the timer started. The correct choice by the subject was to press the left key if there was one dot at the end of the target row and the right key if there were two dots. When the subject pushed either of the keys, the shutter closed and the timer recorded the time. Feedback for the subject was provided when he heard the examiner read aloud the recorded time to a tape recorder. Three problems were presented using six subjects for each problem. The target for the first problem was red, small, digit "3". The second problem separated the multiple attribute target into the single attributes of red, small, and digit "3". The third problem was the same as the second using six new subjects. No information about the subjects was given. The time and number of errors were recorded for each subject. The order of analysis of the clues presented was color, size, and then shape.

Mazes usually consist of several points at which two or three choices are available. Selection of the correct path at each point results in reaching the goal or the end of the maze. Early studies using children include the ball rolling maze by Mattson (1933), the finger-maze by Wieg (1932), the stylus-maze by McGinnis (1929) and the body-maze by Wenger (1933).

Various results were reported by other maze studies as follows: adults slightly better than children and children slightly better than rats by Hicks and Carr (1912), only a slight difference between children of age 11 and adults of age 21 in a stylus-maze by Gould and Perrin (1916), university students slightly better than rats by Husband (1929), and a low correlation between intelligence test scores and maze performance of psychology students in an electrical type maze by Lumley (1931). Mattson (1933) using maze learning found the learning curve for boys and girls almost identical. McGinnis (1929) found boys initially ahead of the girls, who soon equaled their skill. Thus the girls made greater absolute gains during the test than the boys.

Visual apprehension was investigated by De Santis (1931) using 9 normal children from 3 to 6 years of age, 9 feeble-minded from 7 to 12, 10 imbeciles, 1 idiot, and 1 moron. He reported that the feeble-minded were slower but even down to the idiot level, some spatial data was learned. The slowness was probably due to poor attentive capacity.

Gellerman (1931) compared 38 children ranging in age from 3 to 13 years and 25 college students using a large body-size temporal alley maze. Whenever an incorrect turn was made, a door blocked further progress so the subject had to go back and make the correct move. Three correct trials in succession were required. For children the solution was twice right, then twice left; for adults, it was the same sequence plus one repetition. The children required more than two and a half times as many trials as the adults. The average number of errors was 30 for children and 16 for adults. Two children ages three and four failed to solve the problem. Beyond the age of four there was a gradual decrease in the number of trials required. For adults there was a correlation of .58 between intelligence score and number of trials to learn and similarly high between accuracy and intelligence.

Maier (1936) used a swastika-shaped maze through which the children walked. There was a goal booth at the end of each of the four pathways, any one could be reached from the others, but only one contained a toy house. The subjects were 38 children in 5 successive age groups, the youngest age group averaged 4 years. All of the children in the four-year group failed and the average five-year-old group did not do much better. It was not until the age of six that accuracy significantly exceeded chance. Maier (1936) concluded that the average child below the age of six had not acquired the ability to combine experiences required by his test. Performance improved with both C.A. and M.A. but was not significant until a C.A. of six or a M.A. of six.

Mussen (1960) points out that most studies in concept formation have been to determine the concepts subjects have already acquired at different ages rather than testing to determine the original learning from concept teaching. He uses the terms "concept learning" and "concept discovery." He suggests that clear-cut studies of concept learning are rare in the literature, a systematic study has scarely begun.

Peterson (1918) used the association of a number with a letter in a study of rational learning. Psychology students who were the subjects had to guess at the number associated with each letter until they had learned the whole series. The test scores obtained in this investigation correlated highly with estimates of intelligence and with class standing in psychology.

One of the few systematic investigations of concept learning in adults was done by Hull (1920). He used Chinese characters, unfamiliar to his subjects, in complex stimulus patterns as a measure of the evolution of concepts. Improvement with age was found. He suggested that the selection of a common identical element may not necessarily involve the same function as the arriving at an abstraction from symbols where the common element is a broad term.

Vickers and Hoskins (1929) used an obstacle test, inverse form board, area sorting, substitution, and card sorting according to geometrical design with groups of children and adults. These tests, thought to measure intellectual ability in a practical situation, showed a steady increase in solutions with increase in age.

Hazlitt (1929) used two problems, the sorting of small and large black cards and assembling the Russian egg after instruction. The subjects were told to replace all of the smaller eggs within the larger egg except the green one. The subjects were from three to seven years of age. The results suggest that regardless of chronological age, mental age, or relative brightness, the reasoning process is essentially the same when the experience factor is eliminated and all of the data is within the understanding of the subject. This is another way of saying that if the problem is simple enough, all subjects, dull or bright and younger or older, will not be different in their solutions.

The formation of concepts through learning to associate nonsense syllables with symbols such as a circle and two dots was investigated by Smoke (1932). Twenty psychology students were the subjects. A positive correlation between the ability to learn the concepts rapidly and high intelligence was reported.

Ray (1936) used an apparatus that automatically presented a new pair of pictures after each choice was made by pressing a key under either of the two pictures. A green light would flash with a correct choice and a red light when incorrect. The 20 series of pictures consisted of complete versus incomplete objects, curvilinear versus non-curvilinear objects, and animate versus inanimate objects. The subjects were twelve year old children, six dull, six normal, and six bright. The brights solved more problems using fewer hypotheses, but the number of hypotheses tested, correlated negatively with intelligence. Ray (1936) points out that the perception of a great variety of clues to generate a larger number of hypotheses was not necessarily a sign of intelligence. The generation of fewer hypotheses relevant to the data may be a sign of intelligence as suggested by the negative correlation. The use of irrelevant clues and perseveration of hypotheses were common to the dull and average but not among the bright. Verbalization of the correct solution was more frequent among the bright than the dull subjects.

Blake and Williams (1963) used a paired-associates problem to compare groups of retarded, I.Q. 78; normal, I.Q. 95-99; and superior, I.Q. 132-136, students selected from the public schools. Each group was subdivided into two sets, one set was equated by M.A., the other set by C.A. The paired-associates test consisted of four object-level words and two-digit numerals as labels as follows: mammals 47, flowers 91, trees 28, and weather 75. Testing was preceded by a training phase. Four to 10 subjects were tested in groups composed of subjects from each I.Q. level. The findings were that the groups equated by M.A. were not significantly different in level of concept attainment. For the groups equated by C.A., the superior group performed significantly better than the normal and the normal group significantly better than the retarded.

Three studies comparing normal to subnormal boys using the same materials were performed by Stephens (1964, 1966, 1968). A series of 27 cards were constructed, 2 sample cards and 25 test cards. Each card was 8 inches by 18 inches with seven pictures on each. Four of the pictures were representative of a given test category (e.g. color, form, detail, sex, age, odd vs. even, number of dots, etc.) and three were not. The subject was required to place a marker on each of the four pictures selected as being alike and then to state in his own words the characteristic common to his choices.

For the first study, Stephens (1964) compared 30 normal with 30 subnormal children attending urban public schools in central Oklahoma. The subjects were placed into groups according to chronological age and intelligence test scores. Ten normals and 10 subnormals were in each age group; 7-6 to 8-5, 8-6 to 9-5, and 9-6 to 10-6. On an intelligence test the normals scored between 96 and 108 with a mean of 101 and the subnormals scored between 47 and 72 with a mean of 60. Each subject was given the two sample cards and helped to solve these examples. Then each of the 25 test cards were presented, the examiner naming the appropriate category for each card, and then the subject placed markers on the four items appropriate to the category. In fourteen of the 25 categories tested, the number of normal subjects attaining the correct solution was significantly greater than the number of subnormals. It was concluded that the normal group made significantly more correct responses than the subnormal group. Stephens (1964) pointed out that the subnormals exhibited a smaller repertory of categories indicating that they could be expected to perform at a lower level even when the cues were strong and relevant to the appropriate category.

For the second study, Stephens (1966) used the same boys in the same groups and the same technique that he used in his former study in 1964. The results of two problems were considered. For the first problem, each subject was required to place the markers on his four choices and then to state in what way he believed the items he had chosen to be alike. The subject was not told whether his choice was correct or incorrect. The cards were presented again to the same subjects for the second problem, the examiner stating the name of each category as the card was shown, requiring the subject to place the markers again. The results of this problem were reported previously by Stephens (1964). It was found that the number of correct responses attained by the normal group was significantly greater than for the subnormal group and the number of correct category verbalizations by the normal group was significantly greater than for the subnormal group. Stephens (1966) pointed out that even though the subnormal subjects were able to use the proper category for grouping the test items, these subjects had great difficulty in specifying the name of that category. This restricted use of functional categories may suggest that the subnormal is less able to comprehend and interpret a wide range of stimuli found in everyday experiences.

The third comparative study was concerned with the types of errors exhibited by the normal and subnormal subjects in their efforts to verbalize the correct conceptual categories which they had used correctly in a problem solving situation. The results of the unstructured problem were reported by Stephens (1966) and of the structured problem by Stephens (1964). The recorded verbalization errors were organized into seven categories: over-generalization, under-generalization, incorrect generalization, enumeration of names, enumeration of functions, similarity statement, and no-knowledge. The no-knowledge errors were considered to be at the lowest level of conceptualization. The generalization errors of the normal subjects comprised 92.36% of the total errors and only 50.44% for the subnormals. Consequently, the remaining six categories were then combined, resulting in only two classifications, errors of generaliztion and errors of non-generalization. Stephens (1968) pointed out that the findings indicated that the subnormals were equally likely to commit both types of errors, generalization and non-generalization, while the normals were more likely to attempt to generalize and thus to make more errors. The retarded subjects produced fewer verbal responses and said "I don't know" more often which was interpreted as expressing no knowledge of the categories even though they had been able to use the concept successfully. Their verbal responses, even though fewer, tended to be low level descriptions of appearance or function rather than a relevant category description.

Various kinds of mechanical puzzles have been used in experiments on problem solving in bright and dull children. The child must make effective use of sticks or other potential implements in order to reach a lure.

Ruger (1910) used puzzles that required the separation of such things as rings from each other. When the subject failed to perceive the logical solution, trial and error behavior was evident.

Eagleson (1940) used various match stick problems and mystery type puzzles in a study using 33 negro subjects. He found a higher correlation between the time required to solve the puzzles and the number of overt manipulations than between the time required and intelligence. No clear-cut differences between the bright and dull were noted.

An instrumentation problem similar to the kind

Kohler (1925) used with chimpanzees was used by Alpert (1928) with 44 nursery school children. For one problem, an object was suspended from the ceiling that could be obtained by placing a nearby block underneath and standing upon it. In another problem, an attractive toy was placed too far from bars of the pen to be reached with the hand. A stick was lying within the pen that could be used to poke through the bars to pull in the toy. In another situation, the toy was placed farther from the bars and two halves of a fishing rod were lying inside the pen. The two halves could be fitted together to be of sufficient length to reach the toy and pull it in. Three types of insight were reported; solution with immediate insight, solution with partial insight, and solution with sudden insight.

Matheson (1931) performed an extensive investigation of instrumentation using 28 children between the ages of 2 and 4½ years. Some of the problems were: to get a lure by unhooking a ring to lower a basket, stack boxes in order to reach a high lure, and one of several crossed strings attached to a lure. Much of the activity was characterized as trial and error. A low positive correlation was found between chronological age and solution of problems and between intelligence and solution of problems.

Syllogistic reasoning, involving only the deductive process, was investigated by Winch (1921, 1922), Wilkins (1928), Moore (1929), Ewert and Lambert (1932), Pyle (1935), and Sells (1936). The findings reported were: steady improvement with age, low positive to .71 positive correlations with I.Q. scores, highly correlated with school success, and high correlation between generalizing and intelligence.

Broady (1940) set up comparable tests of verbal and nonverbal reasoning of concrete classification and abstract classification. All four types developed as a function of increasing chronological age except for nonverbal abstract in which no further development was found after the age of 174 months. The development sequence was: verbal concrete, nonverbal concrete, nonverbal abstract, and verbal abstract.

Various other types of problem solving techniques were used in making comparisons of performance with intelligence. Harter (1930) reported a low positive correlation, Bedell (1934) found that those subjects in the lower quartile in intelligence scored little better than chance on an inference test, Billings (1934) reported moderate correlations with various academic fields. Roslow (1936) found correlations from .32 to .81, and Graham (1938) found success, tracing geometric figures without lifting the pencil, eight times as great for those in the upper decile as compared with those in the lower decile.

Sargent (1940) attempted to investigate the thinking process, both quantitatively and qualitatively by using anagrams. He found that the correlation between levels of I.Q. decreased as the level of difficulty increased. The solving of easy anagrams was facilitated most by speed of reorganization and reaction to clues. The more difficult ones were facilitated by richness of hypotheses, overcoming of interference, and an analysis based on experience. An individual not being proficient in both aspects would account for the decrease in correlation as the difficulty of the anagrams increased.

Stevenson (1968) investigated the performance of 258 boys and 271 girls on an anagram task. Instructions to make as many words as possible in eight minutes from the letters contained in the word "generation" were presented by a professional television announcer on television film. The subjects were public school students in grades 3-9 present in 19 classrooms in middle-class sections of Minneapolis. I.Q. and C.A. were obtained for each grade and each sex. Additional available data on 32 retarded boys and 32 retarded girls of the same C.A. as the normal subjects in grades 7,8, and 9 whose average M.A. was comparable to the M.A. of the normal subjects in grade 4 was recorded. It was found that: girls performed significantly better than boys in all grades, consistent increases occurred at each higher grade level and higher intellectual level within each grade, and anagrams are a productive and efficient means of investigating developmental change in the verbal processes of children. The retarded subjects constructed approximately the same number of words as their M.A. peers in the fourth grade but 27% of the words were not appropriate, nonsense words being more prevalent.

Lipton and Overton (1971) investigated the performance of 80 grammar school children in grades 2, 4, 6, and 8 on an anagram task consisting of scrambled words printed in two inch high block letters. The subjects in each grade were subdivided into high and low reading achievement groups. Each subject was required to print the word after saying it aloud. The time between the presentation of the card and the verbal reply was recorded. There was a significant improvement in performance by grade level and by reading ability level within each grade. Word length was found to have a significant effect on both the number of solutions and solution time. The greatest improvement in performance occurred between the second and fourth grade groups.

After a review of the methods and devices used, Teska (1942) concluded, "None of these tests provided an opportunity for the exercise of all the elements of the problem-solving process. The mazes and puzzles are weak as tests of problem solving because the data do not yield to inductive reasoning. The tests of syllogistic reasoning emphasize the deductive element to the exclusion of the inductive element. The tests of concept formation, while well balanced as problem solving tests within the confines of a given age level, are limited because it is difficult to develop a list of concepts common to both younger and older subjects not too difficult or complex for the younger or too simple for older subjects. The multiple-choice technique is adapted to testing over a wide age range. The generalizations necessary for the solution arise logically from the data--that is, inductively--and can be checked systematically against the data--that is, deductively--and at the same time can be stated with clarity in the language of the very young subject. The data is such that the older or brighter subjects can extract more meaning from it. The type of clues provided by the Yerkes' multiple choice is limited to clues of position and relationship. This limitation is a definite weakness of the test. A test providing a wider variety of clues, thus giving greater range to inductive processes but still retaining the systematic means of checking hypotheses, should be a more adequate test of the problem-solving process" [pp. 26-27].

After a review of the current literature to include the techniques used, the writer agrees with the conclusions of Teska (1942). The Problem Box incoporates all of the better ideas found in the literature. Lehtiö (1970) projected color slides of large and small numbers in red or blue on a rear-projection screen and incoporated a time recording device. The examiner pushed a start button to open the shutter for each trial and the subject pressed either a right hand or left hand key that closed the shutter. The feedback to the subject was delayed and verbal. House and Zeaman (1958), Harter (1965, 1967), and Moffitt (1969) used an apparatus that was not portable. Reinforcement was provided by candy or a marble as a reward for a correct response which is a form of external motivation as well. Gerjuoy (1953) provided feedback by turning out one of three lights when the correct button was pressed. Mussen (1960) pointed out that pushing a button is a clear-cut discrete response which makes observation and recording easy. Push buttons were used by Gerjuoy (1953), Jeffrey (1958), and Terrell and Kennedy (1957). Multiple-choice was concluded to be quite easy for preschool children when used by Kuenne (1946) and Stevenson and Zigler (1958). Gerjuoy (1953), Cantor (1955), and Norcross (1958) demonstrated conclusively that the ability to learn names for the stimuli plays an important role in solving discrimination problems. Mussen (1960) pointed out that the difficulty of a problem can be changed by varying the number of stimuli with the number of choices held constant, by varying the number of response choices with the number of stimuli held constant, or by varying the number of stimuli and response choices. These variations can be provided for using the Problem Box by re-arranging the order of the slides in the carousel of the projector. The Problem Box was used successfully by Pepper (1966) to investigate verbalization in problem solving and by Teska (1942), Hensley (1957), and Heath (1970) in comparative studies of the problem solving performances of children.

Teska (1942) designed and built the Problem Box for the purpose of providing a means of investigating independently or as a group, all of the five elements of the problem solving process discussed previously. The Problem Box was then used by Teska (1942) in order to verify the adequacy of it in providing a means of determining the step in the problem solving process in which success or failure had occurred and in order to use the data provided by the Problem Box to compare the performances of 34 dull and bright subjects of different chronological ages. Ten problems were used with a maximum of 100 trials permitted per problem. After an analysis of the data obtained as a result of the testing procedures used with the Problem Box, Teska (1942) stated, "The test provided several means of determining to what elements in the problem-solving process success or failure could be traced" [p. 45]. It was further concluded, "The test was successful in revealing the differences in the performance of dull and bright children. It was possible to trace success or failure to particular steps in the problem-solving process. Acquired concomitants of intelligence such as good reading ability and rich vocabulary played a small role in success or failure. The test is adapted to use over a wide age range" [p. 52]. After a comparison of the performances of the 34 subjects, it was found that the percentage of solutions increased at each successive age level for both the bright and dull. The youngest of the bright, C.A. 8, was superior to the oldest of the dull, C.A. 14. The bright were superior to the dull in the number of problems solved and in the number of trials required for solution. In comparing the bright with the dull having the same M.A., the bright were superior to the dull in number of problems solved and the number of trials required for solution. The predictive value for the degree of success in the problem solving situation used was greater than that of either M.A. or C.A. It was pointed out that the dull frequently solved problems without verbal generalizations.

Hensley (1957) used the Problem Box to compare the problem solving ability of bright and dull children. Ten problems were used with a maximum of 100 trials permitted per problem. Thirty-seven subjects were placed in groups so that a comparison could be made between bright and dull having the same I.Q. with varying C.A. and M.A., the same C.A. with varying M.A., and the same M.A. with varying C.A. and I.Q. He concluded that at each age level the bright were superior to the dull in the number of problems solved and the number of trials. There was no appreciable difference between the younger bright and the older dull. The ability to solve problems was probably more a function of M.A. than of C.A. or I.Q. The bright were superior to the dull in frequency of verbal generalization at both age levels, apparently the chief point of differentiation.

Pepper (1966) used the Problem Box to investigate verbalization of problem solving behavior when the examiner did or did not reinforce correct responses and when the problem was or was not stated by the examiner. One simple problem was used with a maximum of 60 trials for this problem. The subjects were 66 volunteer undergraduate students at the University of Oklahoma. It was concluded that learning occurred more efficiently when the problem was stated and the reinforcement was clearly identified.

Heath (1970) remodeled the Problem Box because some of the electrical parts were no longer functional and were irreplaceable as they were obsolete. The method of projecting the stimuli objects onto the translucent window was replaced with a Kodak Carousel 650 Automatic Focus slide projector. The four push buttons used by the subject to make his choice were replaced with toggle switches. The electrical system was redesigned without the use of a response recording system. The basic ideas were retained including flexibility and portability. Heath (1970) used the Problem Box to compare the problem solving ability of average I.Q. children representative of a lower socio-economic white neighborhood, a lower socio-economic black neighborhood, and a higher socio-economic white neighborhood. He used 10 problems with a maximum of 100 trials per problem. It was found that there were no significant differences in the problem solving ability between black and white subjects from similar socio-economic backgrounds nor any significant differences between socio-economic levels as measured by the Problem Box. Problem solving ability increased as C.A. and M.A. increased. There were no significant differences between the mean number of trials used for solution and achieving verbal generalization except for three problems solved by Group III, high socio-economic white. These three problems tended to be solved with fewer trials when correct verbal generalizations were not given.

Review of the Cognitive Growth of Children

In designing a study to compare the problem solving ability of children, the age of the subjects was one of the variables to be considered. Problem solving is a complex process that evolves in relation to the developmental stages of intelligence as the child grows older.

Piaget (1962) defined intelligence in terms of operations, coordination of operations. He said, "I shall distinguish four great stages, or four great periods, in the development of intelligence: first, the sensori-motor period before the appearance of language; second, the period from about two to seven years of age, the preoperational period which precedes real operations; third, the period from seven to twelve years of age, a period of concrete operations (which refers to concrete objects); and finally after twelve years of age, the period of formal operations, or positional operations" [p. 121]. About the period from seven to twelve years of age he observed, "At this stage the child is still reasoning on the basis of what he sees because there is no conservation. He is able to master this problem only when the operations are formed and these operations, which we have already sensed at the sensori-motor level, are not formed until around seven to eight years of age. At that age the elementary problems of conservation are solved, because the child reasons on the basis of the transformation per se, and this requires a manipulation of the operation. The ability to pass from one stage to the other and be able to come back to the point of departure, to manipulate the reversible operations, which appears around seven to eight years of age, is limited when compared with the operations of the superior level only in the sense that they are concrete. That is to say, the child can manipulate the operations only when he manipulates the object concretely" [p. 125]. "It is only around seven to eight years of age that a child is capable of solving a problem of inclusion. Another system of operation that appears around seven to eight years of age is the operation of serializing; that is, to arrange objects according to their size, or their progressive weight. It is also a structure of the whole, like the classification which rests on concrete operations, since it consists of manipulating concrete objects" [p. 126].

Piaget (1964) wrote, "...the child of seven or eight thinks before acting and thus begins to conquer the difficult process of reflection" [p. 40]. "Intuition is the highest form of equilibrium attained by the thinking of young children and, in this sense, corresponds to the concrete operation thinking of middle childhood" [p. 48]. "At around seven to eight years, on the average (needless to say, these average ages are dependent on the social and educational environment), the child--after interesting transition phases whose details we shall not discuss here--arrives at the constitution of a logic and of operational structures which we shall call "concrete." This "concrete" stage, as opposed to formal characteristics of true logic, is particularly instructive for the psychology of logical operations in general. It signifies that at this level, the level of the beginnings of logic proper, the operations are not as yet concerned with propositions or verbal declarations but with objects themselves and are confined to classifying, seriating, putting in correspondence with one another,..." [p. 124].

Piaget (1928) observed, "...it is no exaggeration in this sense to say that there is no logical reasoning before the age of 7-8. . . . For by the term 'reasoning' should be understood the work of controlling and proving hypotheses, which work alone creates conscious implications among judgements" [p. 234].

Inhelder and Piaget (1958) stated, "In contrast, stage III subjects view the experiment from the start both in terms of the total number of possibilities and in terms of necessary relations, since they possess operations which both are combinatorial and contain the potential assurance of deductive necessity. In their first correspondence operations they do not merely take note of the empirical relationships but immediately proceed to search for an explanation--i.e., they consider the correspondences as implications" [p. 17]. "Thus, generally, verification at stage III consists of two procedures: (1) separating out variables according to combinations not given by direct observation, and (2) the composition of these relationships according to operations of conjunction and implication such as those of proposition. . . ." [p. 45].

Inhelder and Piaget (1964) observed, "The third stage (starting at 7-8 years) is marked by a new type of anticipation, one that refers to transformations as opposed to static collections, as well as showing an understanding of inclusion. The schema of inclusion may begin as a mere form in search of content. This is what happens when the very diversity of possible criteria makes it difficult for the subject to decide on his first classification without trial-and-error. Alongside this development in anticipation, we find a correspondingly greater flexibility in changing criteria: sometimes an alternative criterion is adopted immediately, while at others the new classification may be preceded by trial-and-error, but even the trial-and-error shows that the subject is aware of the different possibilities. . . ." [p. 221].

Guilford (1952) after reviewing tests composed of syllogisms found that the results usually show that there is a marked spurt in the development of logical powers at about the ages of seven to eight. Even though many examples of spontaneous reasoning in children appear below the age of six, they rarely do well in standard tests of the syllogistic type.

Cattell (1963) isolated "fluid" and "crystallized" components of intelligence. The "crystallized" factor consists largely of "process" functions, presumably not much influenced by learning or educational experience, and reaches maturity at a relatively early age.

Ausubel (1968) found evidence in various studies that bright, dull, and average children grow intellectually at different rates. The terminal age of intellectual growth was the same for all levels, but the dull attain a much larger percentage of their ultimate intellectual status during the early years and tend to grow in spurts and pauses. Normal children grow at a more constant rate and the bright grow at an accelerated rate in later childhood that slows down in middle and late adolescence.

Maier (1936) in a problem where subjects had to put "two and two together" found that not until the average age of six was accuracy far better than chance. Harter (1930) in a pencil-maze problem found fewer inappropriate moves as age increased from three to six, no errorless solutions occurred below the age of five. Gellerman (1931) in a previously discussed study, found that beyond the age of four years there was a gradual decrease in the number of trials required for learning at successive age levels. Braine (1959) in a study repeating two of Piaget's logical operations studies found logical operations concepts evident in children at the age of 5, two years younger than Piaget had found. Bloom (1964) after several extensive studies of development concluded that in terms of intelligence measured at age 17, at least 20% is developed by age 1, 50% by about age 4, 80% by about age 8, and 92% at age 13.

Thurstone and Ackerson (1929) reported the inflection point of the mental growth curve to have a positive acceleration up to the age of about 10 with an inflection point somewhere near the age of 11. They suggest that mental growth curves have initial positive acceleration, an inflection point, and a later negative acceleration for all children, the inflection point coming earlier for the bright than for the dull.

In reviewing the literature discussed above, it was reported that children could solve non-verbal problems as young as 4 or 5, at that age children below I.Q. 70 would have such a low M.A. that they would not be able to perform adequately in a problem solving situation. Piaget (1962) described the period up to 12 years of age as the concrete operations stage and Thurstone and Ackerson (1929) found the mental growth curve to have positive acceleration up to about the age of 10 with an inflection point at about the age of 11. The median age between 5 and 11 or 12 would be more appropriate and since children at the age of 8 were reported by Bloom (1964) to have acquired about 80% of their intelligence, the writer concluded that children at the chronological age of 8 would be most appropriate for a study of non-verbal problem solving.

Statement of the Problem

There were two purposes of this study. The first was to compare the problem solving ability as measured by the Problem Box of three groups of eight-year-old children; very bright, average, and high-grade mentally defective, as measured by the Stanford-Binet Intelligence Scale (1960). The second purpose was to find out whether a relationship existed between the total number of trials used in the attempt to solve the problems and the total number of problems solved by all subjects as measured by the Problem Box.

Definitions

The following operational definitions were used for this study:

Problem Solving Ability: measured by recording for each subject; the number of trials used in the attempt to solve the 10 problems, the number of problems solved, and the number of problems solved with a correct verbalization of the solution.

Solution of the Problem: recognized two criteria as a solution; one was 10 consecutive correct responses indicated by 10 consecutive red lights, and the other was a correct verbalization of the generalization regardless of number of correct trials.

Problem Box: used in this study means the Problem Box designed by Teska (1942), remodeled by Heath (1970), and improved by the writer.

Group I: the bright group, means the white eight-year-old children who scored 130 or above on the short form of the Stanford-Binet Intelligence Scale (1960).

Group II: the average group, means the white eight-year-old children who scored between 90 and 110 on the short form of the Stanford-Binet Intelligence Scale (1960).

Group III: the high grade mentally defective group, means the white eight-year-old children who scored between 55 and 70 on the short form of the Stanford-Binet Intelligence Scale (1960).

Limitations

This study was limited to white eight-year-old (± three months) children attending the public schools in the vicinity of Norman, Oklahoma, who scored between 55 and 70, 90 and 110, and 130 or above on the short form of the Stanford-Binet Intelligence Scale (1960). Parental consent was obtained.

Hypotheses

The following null hypotheses were tested for statistical significance at the .05 level by the Mann-Whitney U Test (Siegel, 1956) using Table K and the Spearman Rank Correlation Coefficient (Siegel, 1956).

1) There is no statistically significant difference in the problem solving ability between; Group I and Group II, Group I and Group III, and Group II and Group III as measured by the total number of trials required in the attempt to solve the ten problems by use of the Problem Box.

2) There is no statistically significant difference in the problem solving ability between; Group I and Group II, Group I and Group III, and Group II and Group III as measured by the number of problems solved by use of the Problem Box.

3) There is no statistically significant difference in the number of correct verbalizations of the correct solution of the problems solved between; Group I and Group II, Group I and Group III, and Group II and Group III.

4) There is no statistically significant correlation between the number of trials used by all subjects in attempting to solve the ten problems and the number of problems solved by all subjects as measured by the Problem Box.

CHAPTER II

METHOD AND DESIGN

Selection of Subjects

White children having a chronological age of eight years (± three months) attending the public schools in the vicinity of Norman, Oklahoma, were selected. Test scores on file were used as a pre-screening method to identify those children who would be most likely to score within the range of the prescribed groups. The short form of the Stanford-Binet Intelligence Scale (1960) was individually administered to all of the subjects by the same examiner who was eligible for certification by the State of Oklahoma as a psychometrist. Those children who scored 130 or above, between 90 and 110, and between 55 and 70 were placed in the appropriate group and tested individually with the Problem Box. The writer administered all of the individual tests on the Problem Box. Table 1 is a description of the subjects and results of the testing.

Description of the Problem Box

The Problem Box designed by Teska (1942) and remodeled by Heath (1970) was improved and altered by the writer as follows:

1) The toggle switches were replaced with push-

button double-throw switches. The push buttons are a little over one-half inch in diameter, molded out of pale red plastic.

2) The viewing screen made of frosted glass was replaced with a Kodak rear-view screen about 4 inches by 4 inches in size that produced more distinct images with more uniform color.

3) The color slides were replaced with Kodacolor II color pictures taken with a 35 mm. single reflex camera of a square and a triangle made of white, dark green, or bright red photographic paper placed in the various positions on a black background. A jig for this purpose was made by a person skilled in mechanical drawing. The new slides produced more uniform positioning of the square and the triangle on the rear-view screen, more uniform color of better quality, and a more uniform distribution of light.

4) The angle of the face of the Problem Box was changed so that the screen was held parallel to the lens of the projector providing a more uniform distribution of light.

5) The delay between the time the subject pushed a button and when the slide changed was shortened from five seconds to three seconds. It seemed that the shorter delay might eliminate some probable boredom possibly felt by the subject.

6) The maximum number of trials permitted each subject in attempting to solve a problem was changed from 100 to 64. It seemed that 100 trials was excessive because many of the children who solved a small number of problems would experience too much failure. After examining the data of

Hensley (1957) and Heath (1970), the writer decided that a maximum of 64 trials permitted per problem would be more appropriate and would not appreciably change the total number of problems solved.

The Problem Box, made of 1/4 inch plywood, is 10 inches high by 20 inches wide by 26 inches long and is open at the end opposite the screen. The box contains a 35 mm. Kodak 850 Carousel Projector placed at the open end of the box opposite the screen. A Kodak rear-view screen approximately four inches square is centrally positioned in the face of the box with a red light centered above it. There are four push buttons, one placed adjacent to each quadrant of the screen, any one of which when pushed will activate a 20-position variable (1-30 sec.) delayed-action relay switch. This delayed-action switch is pre-set at three seconds. After three seconds, it will activate the projector to automatically project the next slide onto the screen and to turn off the red light if it was on. Only the correct button will turn the red light on. The purpose of the three second delay is to provide the subject with an opportunity to review his choice. The box also contains various electrical panels. Three views and a schematic drawing of the electrical system of the Problem Box are presented in Figures 1, 2, 3, and 4 in the Appendix.

The slides used in the Problem Box consisted of a series of pictures of a square and a triangle in eight different positions and in varying combinations of colors; red, white, and green. Twelve different positions of the square and triangle were possible but the four in which both figures appear together on the right or the left were excluded. Each carousel contained 80 slides placed in the desired sequence, the first 10 slides having the figures in varying positions and colors, the other 70 slides a continuous duplication of the first 10. Four carousels were used. One carousel contained slides of both figures in white used for problem one, two, and three. The second carousel contained slides of one white figure and the other figure green used for problem four. The third carousel contained slides of both figures red or both figures green used for problems five, eight, nine, and ten. The fourth carousel contained slides of one figure in red and the other in green used for problems six and seven.

The order of presentation of the ten problems, easiest to most difficult, was determined by Teska (1942) in his pilot study, verified by his final test, and used by Heath (1970). These problems in the same order were used in this study as follows:

1. Always the square (demonstration).
2. Always the triangle.
3. Always the figure on the left.
4. Always the green figure.
5. Triangle when both figures are red, square when both figures are green.
6. Always the figure on the right, but in each trial one figure is red and the other is green. Color is a false clue.
7. Red-green alternation, regardless of figure.
8. Square when both figures are red, triangle when both figures are green.
9. Single alternation of the figures, figures appearing both red or both green. Color is a false clue.
10. Square when the figures are in opposite corners, triangle when the figures are in adjacent corners, the figures appearing both red or both green. Color is a false clue.

Test Procedure

The Problem Box was placed on a table in a suitable position for the subject. The appropriate carousel was placed in the projector, the Problem Box was turned on and the first slide of the problem was projected onto the screen. The subject was seated in front of the Problem Box in a comfortable position and the examiner was seated in back of the subject where he could observe which button the subject pushed and whether or not the red light flashed on when the subject pushed a button. The examiner recorded each response, whether correct or incorrect, on a scoring sheet illustrated in Figure 5 of the Appendix.

All subjects were given these instructions on the first problem only: "One of these four buttons (pointing out the four buttons) will cause this light (pointing to the light) to flash red." The examiner then pressed the button beside the square and the red light came on. "Remember only one of these four buttons will cause the light to flash." The examiner again pressed the button beside the square and the red light came on. "The idea is to make the light flash red every time." The examiner again pressed the button beside the square and the light came on. Usually the subject began to push the buttons but occasionally it was necessary to say to the subject, "You do it now." No further demonstration trials were given until the subject had used 64 trials and had not solved the first problem. After getting the red light in a few consecutive trials, some of the subjects would generalize the solution by such statements as; "It's the square," "This is easy," or "I think I know." If the subject did not verbalize the solution voluntarily after ten consecutive red lights, considered a solution to the problem, he was asked, "You have been getting the light to flash every time. How do you do it?" If no response, he was questioned further, "How do you tell which button to push next?" On the first problem only, if the subject did not solve it and verbalize the solution, a demonstration was given until he understood the solution. He was then encouraged to push the buttons until he flashed the red light every time and was able to verbalize the solution. This gave the subject a "success experience" and encouraged him to continue with the other problems. A maximum of 64 trials was permitted for each problem. A problem was considered solved when the subject voluntarily verbalized the correct solution or after ten consecutive red lights.

After the subject had attempted to solve all ten of the problems, he was asked what he had "tried" in order to solve the problems. Most of the responses were "I forgot." Others were; "The triangle and the square," "The red ones and the green ones," or "This button, then this button, then this button, and this button (Pointing)." These responses and the generalizations verbalized during testing were recorded on the scoring sheet. The number of trials used per problem, the total trials for all problems, the total number of problems solved, and the total number of problems solved and the solution verbalized are tabulated for each subject in Table 1.

Statistical Analysis of the Data

The Mann-Whitney U Test was chosen to test hypotheses one, two, and three because Siegel (1956) stated that it is one of the most powerful of the nonparametric tests for independent samples and unequal groups. It tests whether two independent groups have been drawn from the same population. The groups were unequal; Group I had 18 subjects, Group II had 17, and Group III had 15. A parametric test could not be used because the assumptions of normality, homogeneity of variance, and continuous data with equal intervals could not be met. The subjects were not selected at random from the population and the measurements were ordinal. The scores of the two groups to be compared were ranked together. Tied observations were assigned the average of the tied ranks. Correction for ties was not made since Siegel (1956) stated the test is more conservative when a correction for ties is not made. The formula method for determining the value of U was used as the groups were fairly large. Since the sise of the largest sample was larger than 9 and less than 20, Table K. in Siegel (1956) was used to find the critical value of U. If the observed value was equal to or less than the critical value at the .05 level of significance, two-tailed test, the null hypothesis was rejected.

The Spearman Rank Correlation Coefficient was used to test hypothesis four. Siegel (1956) stated, "It is perhaps the best test known today" [p. 202]. Each of the three groups was tested for a relationship between the total number of trials used in attempting to solve the ten problems and the number of problems solved. There was a large proportion of tied observations requiring the use of the formula provided for that purpose by Siegel (1956) to find the observed value of rho. The critical value of rho at the .05 level, one-tailed test, was obtained from Table P in Siegel (1956). If the observed value was equal to or greater than the critical value, the null hypothesis was rejected.

CHAPTER III

RESULTS

Fifty white public school children eight years of age ($\pm$ three months) were placed in appropriate groups in order to compare their non-language problem solving ability as measured by the Problem Box. Group I was composed of 18 very bright children who scored 130 or higher on the short form of the Stanford-Binet Intelligence Scale (1960); Group II, 17 average children who scored between 90 and 110; and Group III, 15 mentally defective children who scored between 56 and 70. A qualified psychometrist in the graduate school at the doctoral level at the University of Oklahoma administered all of the intelligence tests. The writer tested all of the children with the Problem Box. Two children were tested but excluded from the study because of apparent emotional problems. The description of the subjects and the results of testing are presented in tabular form in Table 1.

The Mann-Whitney U Test was used to test at the .05 level (two-tailed) of significance, hypotheses numbered one, two, and three. All of these null hypotheses were rejected at the .002 level of significance, see Table 2.

Hypothesis 1, a comparison of the total number of trials required in the attempt to solve the ten problems, was tested. The results, presented in Table 2, were as follows:

1) Between Groups I and II, the observed value of U was 20.5. The critical value at the .05 level is 93 and at the .002 level (two-tailed test), it is 61. The observed value of U was less than the critical value at both levels of significance so the null hypothesis was rejected at the .002 level. The bright group used significantly fewer trials than the average group.

2) Between Groups I and III, the observed value of U was 0. The critical value at the .05 level is 80 and at the .002 level (two-tailed), it is 51. The observed value of U was less than the critical value at both levels so the null hypothesis was rejected at the .002 level. The bright group used significantly fewer trials than the mentally defective group.

3) Between Grpups II and III, the observed value of U was 4. The critical value at the .05 level is 75 and at the .002 level it is 47. The observed value of U was less than the critical value at both levels so the null hypothesis was rejected at the .002 level. The average group used significantly fewer trials than the mentally defective group.

Hypothesis 2, a comparison of the total number of problems solved, was tested. The results, presented in Table 2, were as follows:

1) Between Groups I and II, the observed value of U was 35.5. The critical value at the .05 level is 93 and at

the .002 level, it is 61. The observed value of U was less than the critical value at both levels so the null hypothesis was rejected at the .002 level. The bright group solved significantly more problems than the average group.

2) Between Groups I and III, the observed value of U was 0. The critical value at the .05 level is 80 and at the .002 level, it is 51. The observed value of U was less than the critical value at both levels so the null hypothesis was rejected at the .002 level. The bright group solved significantly more problems than the mentally defective group.

3) Between Groups II and III, the observed value of U was 2. The critical value at the .05 level is 75 and at the .002 level, it is 47. The observed value of U was less than the critical value at both levels so the null hypothesis was rejected at the .002 level. The average group solved significantly more problems than the mentally defective group.

Hypothesis 3, a comparison of the total number of problems solved and verbalized, was tested. The results, presented in Table 2, were as follows:

1) Between Groups I and II, the observed value of U was 8. The critical value at the .05 level is 93 and at the .002 level, it is 61. The observed value of U was less than the critical value at both levels so the null hypothesis was rejected at the .002 level. The bright group solved and verbalized significantly more problems than the average group.

2) Between Groups I and III, the observed value of U was 0. The critical value at the .05 level is 80 and at

the .002 level it is 51. The observed value of U was less than the critical value at both levels so the null hypothesis was rejected at the .002 level. The bright group solved and verbalized significantly more problems than the average group.

3) Between Groups II and III, the observed value of U was 5. The critical value at the .05 level is 75 and at the .002 level, it is 47. The observed value of U was less than the critical value at both levels so the null hypothesis was rejected at the .002 level. The average group solved and verbalized significantly more problems than the mentally defective group.

To test hypothesis four, the significance of the correlation between the total number of trials used in attempting to solve the ten problems and the number of problems solved, the Spearman Rank Correlation Coefficient was used. A significant positive correlation was found for all three groups at the .01 level. This indicates that there is a positive relationship between the number of trials used and the number of problems solved. The results, presented in Table 2, were as follows:

1) For Group I, the obtained rho was .76 for the 18 subjects. At the .05 level of significance, the critical value is .399 and .564 at the .01 level. The obtained rho was greater than the critical rho; therefore, the null hypothesis was rejected at the .01 level. A highly significant positive correlation between the number of trials used and the number of problems solved was found for the bright group. This

indicates that there is a positive relationship between the number of trials used and the number of problems solved.

2) For Group II, the obtained rho was .71 for the 17 subjects. The critical value at the .05 level is .402 and .583 at the .01 level (interpolated). The obtained rho was greater than the critical rho; therefore, the null hypothesis was rejected at the .01 level. A highly significant positive correlation between the number of trials used and the number of problems solved was found for the average group. This indicates that there is a positive relationship between the number of trials used and the number of problems solved.

3) For Group III, the obtained rho was .85 for the 15 subjects. The critical value of rho is .440 at the .05 level and .623 at the .01 level (interpolated). The obtained rho was greater than the critical rho; therefore, the null hypothesis was rejected at the .01 level. A highly significant positive correlation between the number of trials used and the number of problems solved was found for the mentally defective group. This indicates that there is a positive relationship between the number of trials used and the number of problems solved.

TABLE 1

SUBJECT DATA AND RESULTS OF TESTING, GROUP I

Subject No.	C.A.	I.Q.	M.A.	Problem Numbers and Trials 1	2	3	4	5	6	7	8	9	10	Total Trials	Total Problems Solved	Total Solved and Verbalized
1	7-9	138	10-9	31	3	50	4	16	11	61	5	28*	64	273	9	8
2	8-0	130	10-6	10	7	53	7	6	24	64	8	39	64	282	8	8
3	8-2	130	10-9	64	4	64	4	13	43	29	12	12	64	309	7	7
4	8-1	144	11-9	64	10	64	14	6	64	15	10	20	50*	317	7	6
5	7-11	141	11-3	2	10	64	4	5	64	23	64	19	64	319	6	6
6	7-11	151	12-0	10	10	55	10	7	30	64	6	64	64	320	7	7
7	8-2	145	12-0	22	6	64	10	5	64	7	47	43	64	332	7	7
8	7-9	135	10-6	59	4	18	5	6	27	33	64	64	64	344	7	7
9	7-10	133	10-6	33	3	21	5	63	34	61	11	64	64	359	8	8
10	8-0	146	11-9	64	7	64	10	16	64	12	12	49	64	362	6	6
11	8-0	143	11-6	10	10	64	10	23	64	64	34	20	64	363	6	6
12	7-10	146	11-6	58	10	64	12	6	53	64	23	24	64	378	7	7
13	8-3	135	11-3	42	5	64	24	64	64	32	8	16	64	383	6	6
14	8-0	136	11-0	6	10	36	8	64	22	64	64	45	64	383	6	6
15	7-9	141	11-0	64	10	55	6	29	28	64	18	64	64	402	6	6
16	8-2	130	10-9	64	10	64	14	21	35	58	13	64	64	407	6	6
17	8-0	139	11-3	64	3	55	2	64	12	64	25	64	64	417	5	5
18	8-0	130	10-6	64	3	37	10	55	8	64	59	64	64	428	6	6

* solution not verbalized

TABLE 1--Continued

SUBJECT DATA AND RESULTS OF TESTING, GROUP II

Subject No.	C.A.	I.Q.	M.A.	Problem Numbers and Trials 1	2	3	4	5	6	7	8	9	10	Total Trials	Total Problems Solved	Total Solved and Verbalized
19	7-10	110	8-9	30	10	38*	4	5	64	64	64	48	58*	378	7	5
20	7-9	98	7-9	18	10	60*	10	22	64	64	32	64	36	380	7	6
21	8-3	104	8-9	64	10	64	10	22	64	64	13	50	32*	393	6	5
22	8-2	96	8-0	64	5	64	4	13	64	64	6	64	64	412	4	4
23	8-1	110	9-0	64	10	64	11	20	64	35	11	62	64	425	6	6
24	7-11	106	8-6	64	10	9	10	64	64	64	13	64+	64	426	4	4
25	8-1	97	8-0	13	10	64	10	52	64	64	24	64	64	429	5	5
26	8-3	104	8-9	7	16	64	10	64	64	16	64	64	64	433	4	4
27	8-0	98	8-0	64	10	43	6	63	64	64	10	64	64	452	5	5
28	7-9	92	7-3	19	10	30	10	64	64	64	64	64	64	453	4	4
29	8-0	95	7-9	64	11	64	7	64	64	62	64	11	64	475	4	4
30	8-1	91	7-6	64	10	37*	10	35	64	64	64	64	64	476	4	3
31	7-11	109	8-9	64	5	63*	15	12	64	64	64+	64	64	479	5	4
32	8-0	92	7-6	64	10	49	12	64	29	64	64	64	64	484	4	4
33	7-10	104	8-3	64	4	39*	7	64	50*	64	64	64	64	484	4	2
34	8-0	101	8-3	25	10	64	13	64	64	64	64	64	64	496	3	3
35	7-11	93	7-6	64	18	64	10	28	64	64	64	57	64	497	4	4

* solution not verbalized
\+ solved

TABLE 1--<u>Continued</u>

SUBJECT DATA AND RESULTS OF TESTING, GROUP III

Subject No.	C.A.	I.Q.	M.A.	Problem Numbers and Trials: 1	2	3	4	5	6	7	8	9	10	Total Trials	Total Problems Solved	Total Solved and Verbalized
36	8-1	68	5-8	64	10	64	10	64+	64	64	18	64	64	486	4	4
37	8-2	68	5-9	64	10	64	13	64	18*	64	64	64	64	489	3	2
38	7-11	68	5-7	64	10	64	10	46	64	64	64	64	64	514	3	3
39	7-9	60	4-10	64	10	64	10	64	64	64	64	64	64	532	2	2
40	8-3	70	6-0	64	12	64	11	64	64	64	64	64	64	535	2	2
41	8-0	69	5-9	64	10	64	18	64	64	64	64	64	64	540	2	2
42	7-11	58	4-10	64	24*	64	11	64	64	64	64	64	64	547	2	1
43	8-3	65	5-7	64	17	64	24*	64	64	64	64	64	64	553	2	1
44	7-9	66	5-4	64	12*	33*	63	64	64	64	64	64	64	556	3	1
45	7-9	57	4-8	64	31	64	38	64	64	64	64	64	64	581	2	2
46	8-3	56	4-10	64	10	64	64	64	64	64	64	64	64	586	1	1
47	8-0	67	5-7	64	10*	64	64	64	64	64	64	64	64	586	1	0
48	7-11	67	5-6	64	64	64	14	64	64	64	64	64	64	590	1	1
49	7-11	58	4-10	64	20	64	64	64	64	64	64	64	64	596	1	1
50	8-3	59	5-1	64	27	64	64	64	64	64	64	64	64	603	1	1

* solution not verbalized

\+ solved

TABLE 2

OBSERVED VALUES AND CRITICAL VALUES OF STATISTICAL TESTS

MANN-WHITNEY U TEST

Groups	Observed Values of U			Critical Values of U		Significant difference*
	Number of trials	Number of problems solved	Number of problems solved and verbalized	.05 level two-tailed	.002 level two-tailed	
I vs. II	20.5	35.5	8	93	61	.002 level
I vs. III	0	0	0	80	51	.002 level
II vs. III	4	2	5	75	47	.002 level

*To be significant, the observed values must be equal to or less than the critical value.

SPEARMAN RANK CORRELATION

Number of Trials vs. Number of Problems Solved

Groups	Number in group	Observed rho	Critical Values of rho		Significant correlation*
			.05 level one-tailed	.01 level one-tailed	
I	18	.76	.399	.564	.01 level
II	17	.71	.402	.583	.01 level
III	15	.85	.440	.623	.01 level

*To be significant, the observed value must be equal to or greater than the critical value.

CHAPTER IV

SUMMARY AND CONCLUSIONS

A comparison of the problem solving ability as measured by the non-language multiple-choice Problem Box was made between very bright, average, and mentally defective white eight-year-old children attending the public schools in the vicinity of Norman, Oklahoma. The mental ages varied between M.A. 10-6 and 12-0 for the 18 very bright children, between M.A. 7-3 to 9-0 for the 17 average children, and between 4-8 to 6-0 for the 15 mentally defective children.

Intelligence was found to be a good predictor of the non-language problem solving ability of white eight-year-old children attending the public schools.

Either the total number of trials used or the total number of problems solved was found to measure about equally the non-language problem solving ability of white eight-year-old children attending the public schools.

Discussion

The conclusion reached in this study agrees with the findings of Teska (1942) who concluded that the predictive value of brightness for the degree of success in the problem solving situation was greater than either mental or chronological age. Hensley (1957) concluded that the ability to solve problems was probably more a function of mental age than either chronological age or intelligence. Both of these investigators used the original model of the Problem Box. Moffitt and Coates (1969) in a two-choice test of problem solving ability reported that performance was better for higher than lower I.Q. groups.

Thirteen children failed to verbalize the solution to sixteen solved problems. Two of the very bright failed to verbalize one problem each. Six average were unable to verbalize eight solved problems, five of them on problem 3. Five mentally deficient subjects were unable to verbalize six solved problems, three of them on problem 2. It was interesting to note that only one very bright child solved problem 10 but was unable to verbalize the solution while three average children solved problem 10 and one was able to verbalize the solution. A majority of the mentally defective children in this study said, "I don't know" when asked how they got the red light every time. Stephens (1966) reported that subnormal boys evidenced significantly less ability to provide names for categories they had used successfully in previous tasks than normal boys of the same age.

After the subjects in this study had attempted all ten of the problems, none of them solved all ten and only one very bright solved nine, each was asked what they had tried in order to solve the problems. Predominantly the responses were, "I don't remember." Other responses were not possible solutions to problems not already solved. This indicates that 64 trials per problem were sufficient. In describing the behavior of seven-and-eight-year old children Piaget (1928) observed, "In short, as soon as the problem contains ever so little complexity the child seems to conduct his reasoning process much as we do ours in the course of a purely empirical problem (such as a puzzle, etc.); i.e. he keeps no record of his successive attempts, and his mind proceeds by a series of fumbling movements, each of which is conscious, but that does not easily admit of retrospection as a whole. If after this is over we ask the child to describe his search, all he gives us is a device for finding the solution, and a device which presupposes the solution in question. Never does he succeed in describing his reasoning process as such" [p. 211].

The data obtained in this study indicates that the order of presentation of the problems, the least to the most difficult, should be changed. The order of difficulty indicated by this study was problems 1, 2, 4, 5, 8, 3, 6, 7, 9, and 10.

It is suggested that testing with the Problem Box might be discontinued after three consecutive problems are failed and the balance assumed unsolved and scored as the maximum number of trials permitted per problem. It did not appear that the number of trials permitted in an attempt to solve a problem should be increased from 64. It appeared to the examiner that many of the lower I.Q. children became restless and disturbed, possibly on account of continued failure, in spite of the occasional flash of the red light.

Suggestions for Additional Studies with the Problem Box

1) A further investigation of the order of difficulty of the problems should be made.

2) Additional problems with two different sizes of the square and triangle as another variable might be considered. Lehtiö (1970) with an apparatus similar to the Problem Box reported the order of difficulty as being color, size, and shape. A description of the 18 subjects was not reported.

3) An investigation using older subjects having the same chronological age with controlled I.Q. scores and varying mental ages should be made.

4) A timing and recording device might be added to the Problem Box so that a recorded tape of the time between choices and the button pushed might be automatically recorded. The subject might be less disturbed if the examiner did not have to score each choice, and the examiner would have more time to observe the behavior of the subject.

REFERENCES

Aarons, Leon. Serial learning and generalizing abstraction. American Journal of Psychology, 1933, 45, 417-432.

Alpert, Augusta. The solving of problems situation by preschool children. New York: Bureau of Publications, Teachers College, Columbia University, 1928.

Ausubel, David P. Educational psycholology a cognitive view. New York: Holt, Rinehart and Winston, Inc., 1968.

Bedell, R. C. The relationship between ability to recall and the ability to infer in specific learning situation. Bulletin of Northeastern Missouri State Teachers College, 1934, 34, No. 9, 55.

Billings, M. L. Problem solving in different fields of endeavor. American Journal of Psychology, 1934, 46, 259-272.

Blake, Kathryn A., and Williams, Charlotte L. Induction and deduction and retarded, normal, and superior subjects' concept attainment. American Journal of Mental Deficiency, 1963, 73, 226-231.

Bloom, Benjamin S. Stability and change in human characteristics. New York: John Wiley and Sons, Inc., 1964.

Braine, M. D. S. The ontogency of certain logical operations: Piaget's formulation examined by nonverbal methods. Psychological Monographs, 1959, 73, No. 5.

Broady, L. Comparable tests of verbal and non-verbal reasoning; their construction and application to developmental problems. Journal of Educational Psychology, 1940, 31, 180-194.

Burtt, Harold E. A study of the behavior of the white rat by the multiple choice method. Journal of Animal Behavior, 1916, 6, 222-246.

Cantor, G. N. Effects of three types of pretraining on discrimination learning in preschool children. Journal of Experimental Psychology, 1955, 49, 339-342.

Cattell, R. B. Theory of fluid and crystallized intelligence. Journal of Educational Psychology, 1963, 54, 1-22.

Coburn, C. A., and Yerkes, R. M. A study of the behavior of the Crow Corvus Americanus and by the multiple-choice method. Journal of Animal Behavior, 1915, 5, 75-114.

De Sanctis, Sante. Visual apprenhension in the maze behavior of normal and feeble-minded children. Journal of General Psychology, 1931, 39, 463-469.

Dewey, John. How we think. Boston: D. C. Heath and Co., 1910.

Dewey, John. How we think. (2nd ed.) Boston: D. C. Heath and Co., 1933.

D'Zurilla, Thomas J., and Goldfried, Marvin R. Problem solving and behavior modification. Journal of Abnormal Psychology, 1971, 78, 107-126.

Eagleson, O. W. A study of puzzle solving. Journal of Psychology, 1940, 9, 259-268.

Ewert, P. H., and Lambert, J. F. Part II, the effect of verbal instruction upon the formation of a concept. Journal of Genetic Psychology, 1932, 6, 400-413.

Gellermann, Louis W. The double alternation problem II. The behavior of children and human adults in a double alternation temporal maze. Journal of Genetic Psychology, 1931, 39, 197-226.

Gerjuoy, I. R. Discrimination learning as a function of the similarity of the stimulus name. Unpublished doctoral dissertation, Iowa State University, 1953.

Gould, M. C., and Perrin, F. A. C. A comparison of factors involved in the maze learning of human adults and children. Journal of Experimental Psychology, 1916, 1, 122-154.

Gozali, Joav. Impulsivity-reflectivity as problem solving styles among educable mentally retarded children. American Journal of Mental Deficiency, 1969, 73, 864867.

Graham, J. L. An experiment in generalizing: a universal problem. Journal of Experimental Psychology, 1938, 23, 96-100.

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

Guilford, J. P. General psychology. (2nd ed.) New York: D. Van Nostrand Co., Inc., 1952.

Hamilton, G. V. A study of trial and error reactions in mammals. Journal of Animal Behavior, 1911, 1, 33-36.

Harter, G. L. Overt trial and error in the problem solving of preschool children. Journal of Genetic Psychology, 1930, 45, 361-372.

Harter, Susan. Discrimination learning set in children as a function of IQ and MA. Journal of Experimental Child Psychology, 1965, 2, 31-41.

Harter, Susan. Mental age, IQ, and motivational factors in the discrimination learning set performance of normal and retarded children. Journal of Experimental Child Psychology, 1967, 5, 123-141.

Hazlitt, Victoria. Children's thinking. British Journal of Psychology, 1933, 45, 663-677.

Heath, Paul Allen. A comparative study in problem solving ability of a group of negro and white children of average intelligence by socio-economic level. Unpublished Ed.d. dissertation, University of Oklahoma, 1970.

Heidbreder, E. Problem solving in children and adults. Journal of Genetic Psychology, 1928, 35, 522-545.

Hensley, Horace Gene. A comparative study in problem solving of bright and dull children. Unpublished Ph.D. dissertation, University of Oklahoma, 1957.

Hicks, V. C., and Carr, H. A. Human reactions in a maze. Journal of Animal Behavior, 1912, 2, 980125.

House, Betty J., and Zeaman, David. A comparison of discrimination learning in normal and mentally defective children. Child Development, 1958, 29, 411-416.

Hull, C. L. Quantitative aspects of the evolution of concepts. Psychological Monographs, 1920, No. 28, 5-6.

Husband, R. W. A comparison of human adults and white rats in maze learning. Journal of Comparative Psychology, 1929, 9, 361-377.

Inhelder, Bärbel., and Piaget, Jean. The growth of logical thinking. New York: Basic Books, 1958.

Inhelder, Bärbel., and Piaget, Jean. The early growth of logic in the child. New York: Harper and Row, 1964.

Jeffrey, W. E. Simultaneous and successive presentation of stimuli in discrimination learning with children. American Psychology, 1958, 13, 333.

Johnson, Donald M. The psychology of thought and judgement. New York: Harper and Brothers, 1955.

Kohler, W. The mentality of apes. New York: Harcourt-Brace and Co., 1925.

Kuenne, M. R. Experimental investigation of the relation of language to transposition behavior in young children. Journal of Experimental Psychology, 1946, 36, 471-490.

Lehtiö, Pekka K. The organization of component decisions in visual search. Acta Psychologica-European Journal of Psychology, 1970, 33, 93-105.

Lipton, Cheryl., and Overton, Willis F. Anticipatory imagery and modified anagram solutions: a developmental study. Child Development, 1971, 42, 615-623.

Long, L., and Welch, L. The development of the ability to discriminate and match numbers. Journal of Genetic Psychology, 1941, 59, 377-387.

Lumley, Frederick. An investigation of the responses made in learning a multiple choice maze. Psychological Monographs, 1931, 32, No. 189.

Maier, N. R. F. Reasoning in children. Journal Comprehensive Psychology, 1936, 21, 357-366.

Matheson, E. A study of problem solving behavior in pre-school children. Child Development, 1931, 2, 242-262.

Mattson, M. L. The relation between the complexity of the habit to be acquired and the form of the learning curve in young children. Genetic Psychological Monographs, 1933, 13, 299-398.

McGinnis, E. The acquisition and interference of motor-habits in young children. Genetic Psychological Monographs, 1929, 6, 209-311.

Merrifield, P. R., Guilford, J. P., Christensen, P. R., Frick, J. W. The role of intellectual factors in problem solving. Psychological Monographs, 1962, 76, 1-20.

Moffitt, Alan R., and Coastes, Brian. Problem-solving strategies and performance of severely retarded children. American Journal of Mental Deficiency, 1969, 73, 774-784.

Moore, T. V. The reasoning ability of children in the first years of school life. Studies in Psychology and Psychiatry, 1929, 29, No. 2.

Mussen, Paul H. (Ed.) Handbook of research methods in child development. New York: John Wiley and Sons, Inc., 1960.

Norcross, K. J. Effects on discrimination performance of similarity of previously acquired stimulus names. Journal Experimental Psychology, 1958, 56, 305-309.

Pepper, Sidney. Verbalization of problem-solving behavior. Unpublished Ph.D. dissertation, University of Oklahoma, 1966.

Peterson, J. Experiments in rational learning. Psychological Review, 1918, 25, 443-467.

Piaget, Jean. The stages of the intellectual development of the child. Bulletin of the Menninger Clinic, 1962, 26, 120-128.

Piaget, Jean. Judgement and reasoning in the child. London: Routledge and Kegan Paul Ltd., 1928.

Piaget, Jean. Six psychological studies. New York: Random House, 1964.

Pyle, W. H. An experimental study of the development of certain aspects of reasoning. Journal of Educational Psychology, 1935, 26, 539-546.

Ray, J. J. The generalizing ability of dull, bright, and superior children. Peabody Contribution to Education, 1936, No. 175.

Roberts, K. E. The ability of preschool children to solve problems in which a single principle of relationship is kept constant. Journal of Genetic Psychology, 1932, 40, 118-135.

Roslow, S. A statistical analysis of rational learning problems. Journal of Genetic Psychology, 1936, 48, 441-467.

Ruger, H. A. The psychology of efficiency. Archives of Psychology, 1940, No. 249, 58.

Sargent, S. S. The psychology of efficiency. Archives of Psychology, 1940, 249, 58.

Sells, S. B. The atmosphere effect: an experimental study of reasoning. Archives of Psychology, 1936, 29, No. 200.

Siegel, Sidney. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill Book Co., 1956.

Smoke, Kenneth L. An objective study of concept formation. Psychological Monographs, 1932, 42, No. 191, 1-40.

Stephens, Wyatt E. A comparison of the performance of normal and subnormal boys and structured categorization tasks. Exceptional Children, 1964, 30, 311-315.

Stephens Wyatt E. Category usage by normal and mentally retarded boys. Child Development, 1966, 37, 355-361.

Stephens, Wyatt E. Labeling errors of mentally subnormal children in a concept task. American Journal of Mental Deficiency, 1968, 73, 273-278.

Stevenson, Harold W., Klein, Robert E., Hale, Gordon A., and Miller, Leon K. Solution of anagrams: a developmental study. Child Development, 1968, 39, 905-912.

Stevenson, H. W., and Zigler, E. F. Probability learning in children. Journal of Experimental Psychology, 1958, 56, 185-192.

Symonds, Percival M., Education and the psychology of thinking. New York: McGraw-Hill Book Co., 1936.

Terman, Lewis m., and Merrill, Maud A. Stanford-Binet Intelligence Scale. Manual for the third revision form L-M. Boston: Houghton Mifflin Co., 1960.

Terrell, G., Jr., and Kennedy, W. A. Discrimination learning and tranposition in children as a function of the nature of the reward. Journal of Experimental Psychology,

Teska, Percy T. Performance of dull and bright children in a non-language multiple-choice problem situation. Unpublished Ph.D. dissertation, University of Wisconsin, 1942.

Thurstone, L. L., and Ackerson, L. The mental growth curve for the Binet tests. Journal Educational Psychology, 1929, 20, 569-583.

Vickers, W., and Hoskins, V. H. Results from some new tests of practical abilities. British Journal of Psychology, 1929, 20, 66-81.

Vinacke, W. Edgar. The psychology of thinking. New York: McGraw-Hill Book Co., 1952.

Wenger, M. A. Path-selection behavior of young children in body-mazes. Jounal of Experimental Education, 1933, 2, 197-236.

Wieg, E. L. Bi-lateral transfer in the motor learning of young children and adults. Child Development, 1932, 3, 247-268.

Wilkins, M. C. The effect of changed material on ability to do formal syllogistic reasoning. Archives of Psychology, 1928, 16, No. 102, 83.

Winch, W. H. Children's reasonings: experimental studies of reasoning in school children. Journal of Experimental Psychology, 1921-22, 6, 122-141.

Yerkes, R. M. The study of human behavior. Science, 1914, 39, 625-633.

Yerkes, R. M. Methods of exhibiting reactive tendencies characteristic of ontogenetic and phylogentic stages. Journal of Animal Behavior, 1917, 7, 11-18.

Yerkes, R. M. A new method of studying the ideational behavior of defective and deranged as compared with normal individuals. Journal of Comparative Psychology, 1921, 1, 369-394.

Yerkes, R. M., and Coburn, C. A. A study of the behavior of the Pig Sus Scrofa by the multiple-choice method. Journal of Animal Behavior, 1915, 5, 185-225.

Yerkes, R. M., and Rossy, Cecelio S. A point scale for measurement of intelligence in adolescent and adult individuals. Boston Medical and Surgical Journal, 1917, 176, 564-573.

APPENDIX

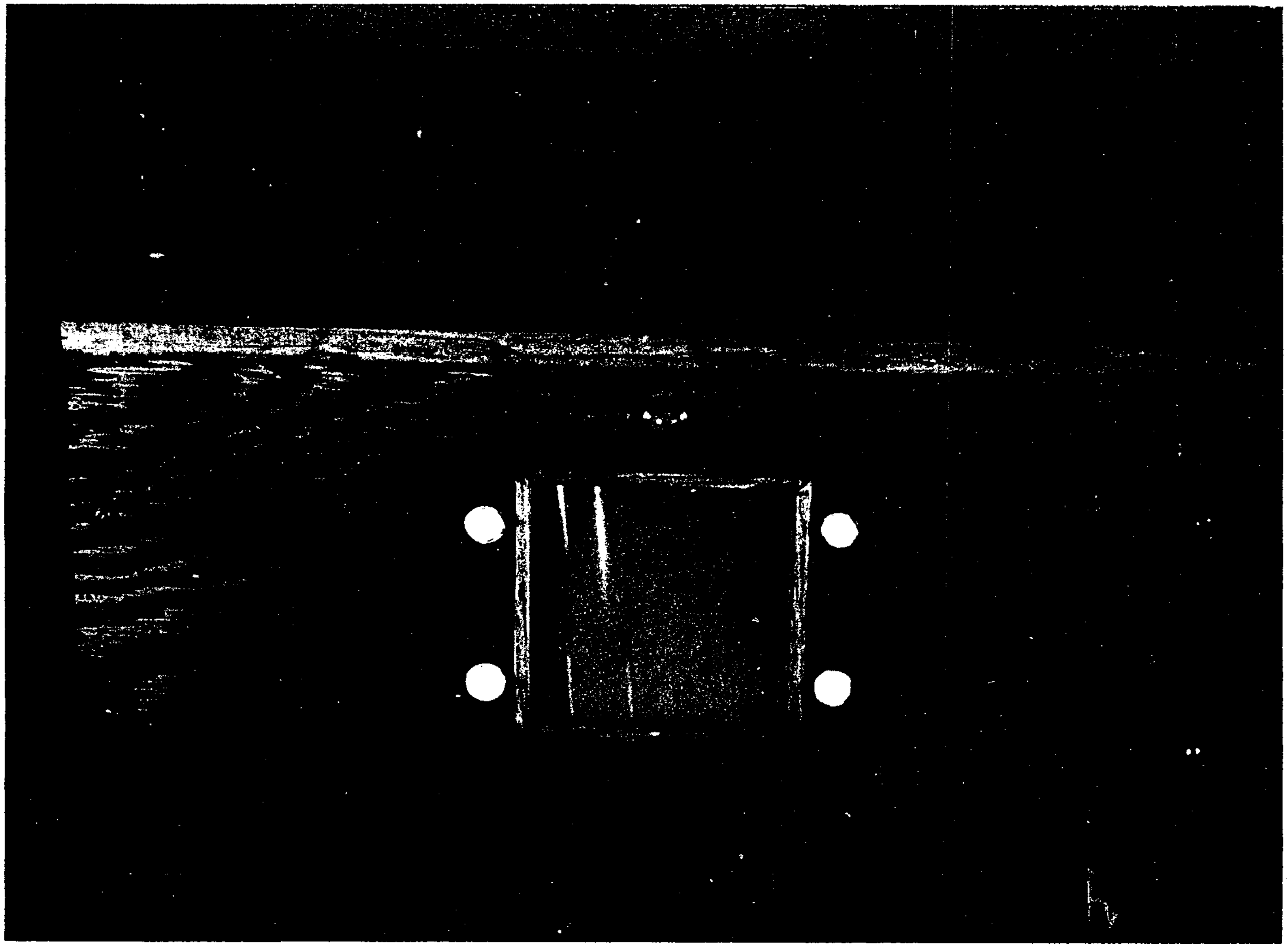

Figure 1. Front view of the Problem Box

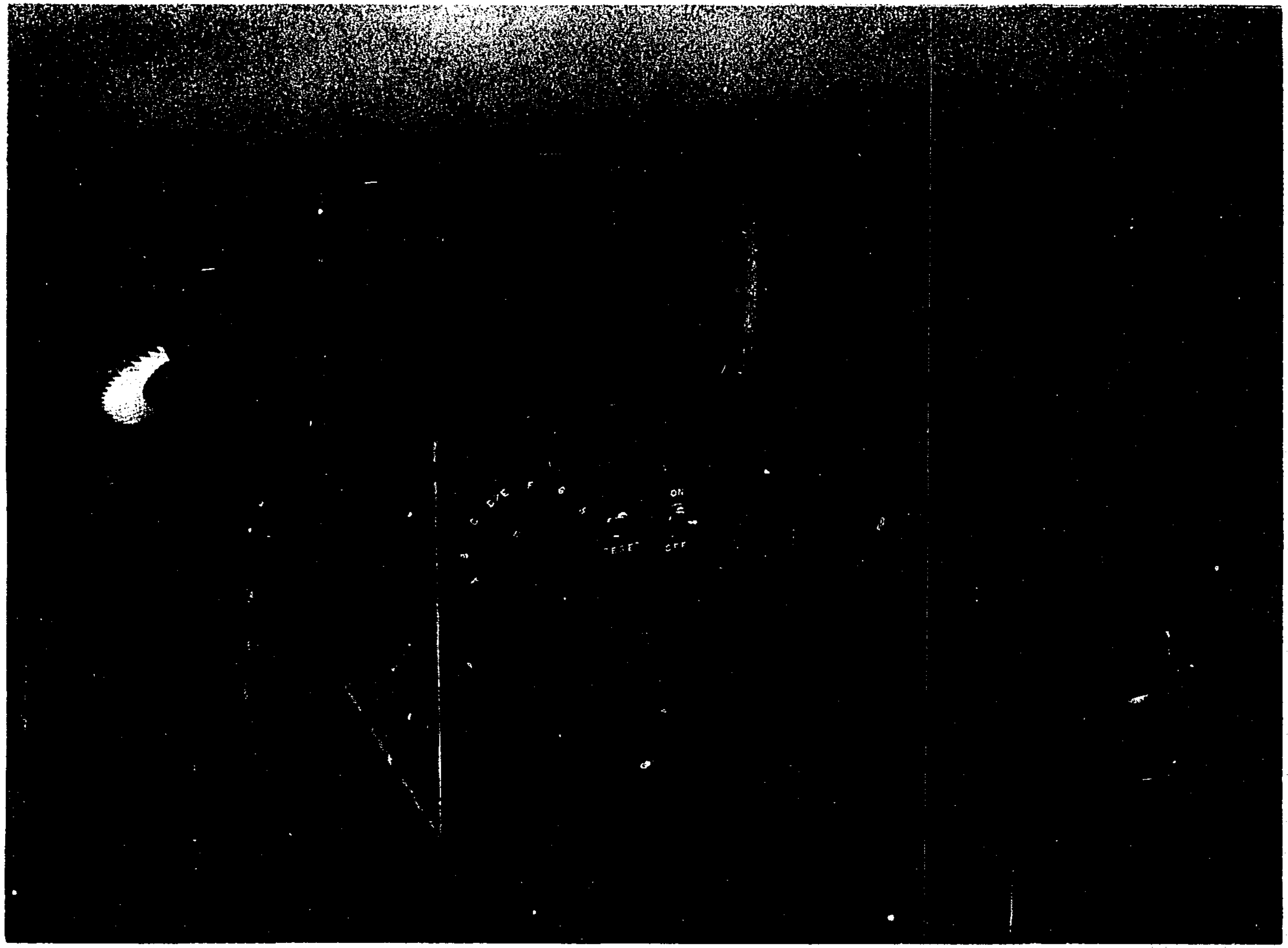

Figure 2. Side view of the Problem Box

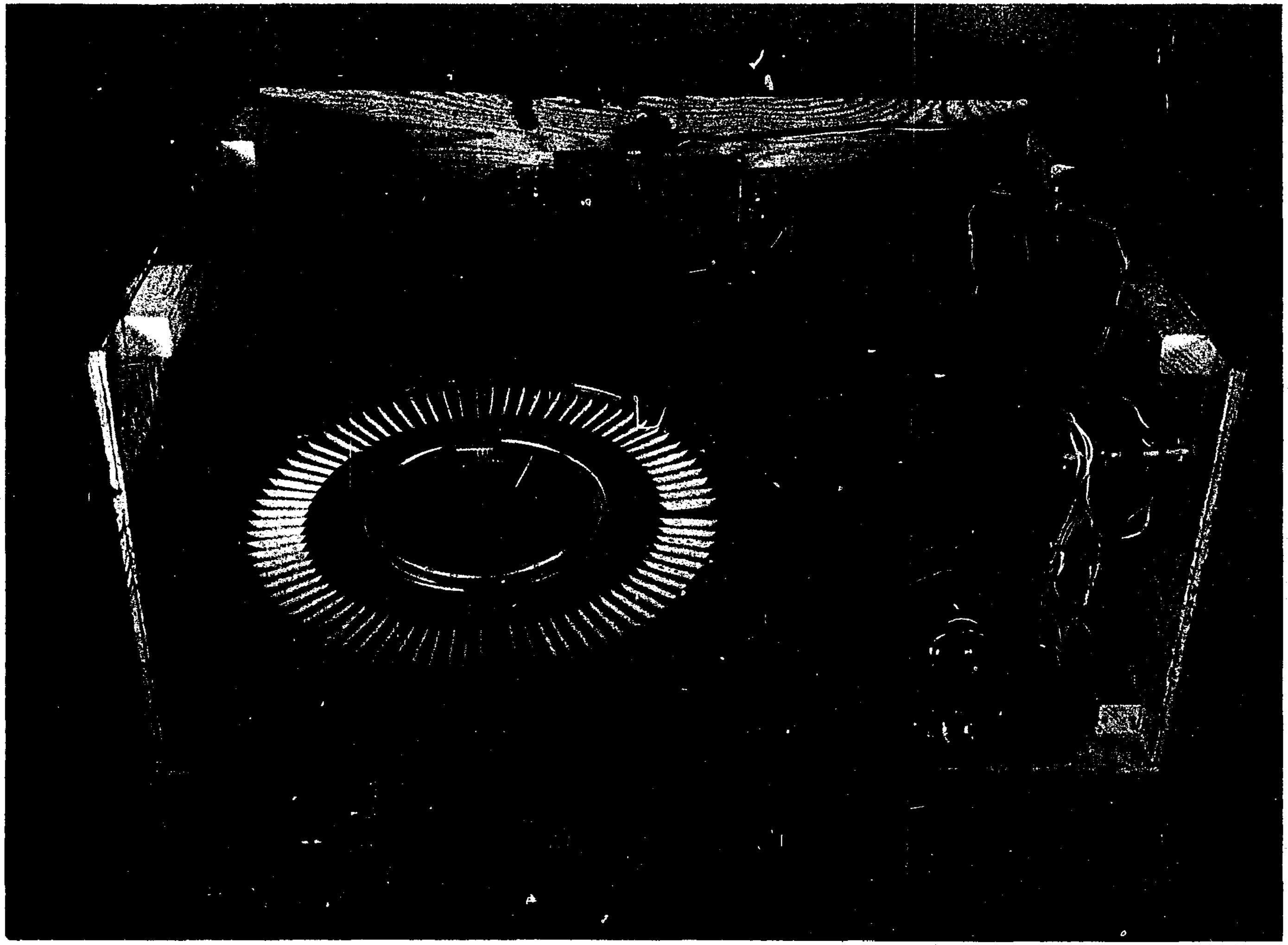

Figure 3. Inside mechanism of the Problem Box

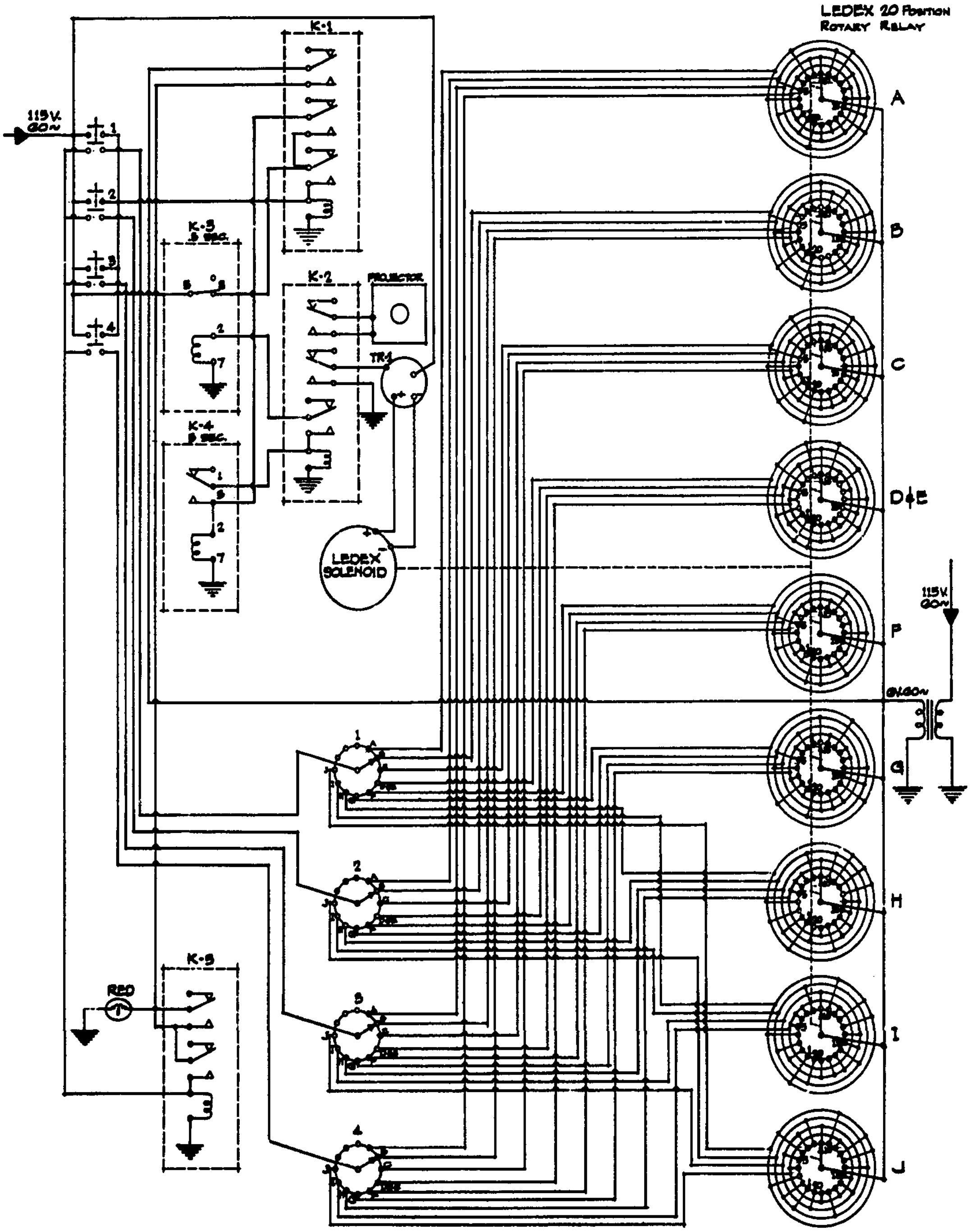

Figure 4. Electrical Schematic Diagram of the Problem Box

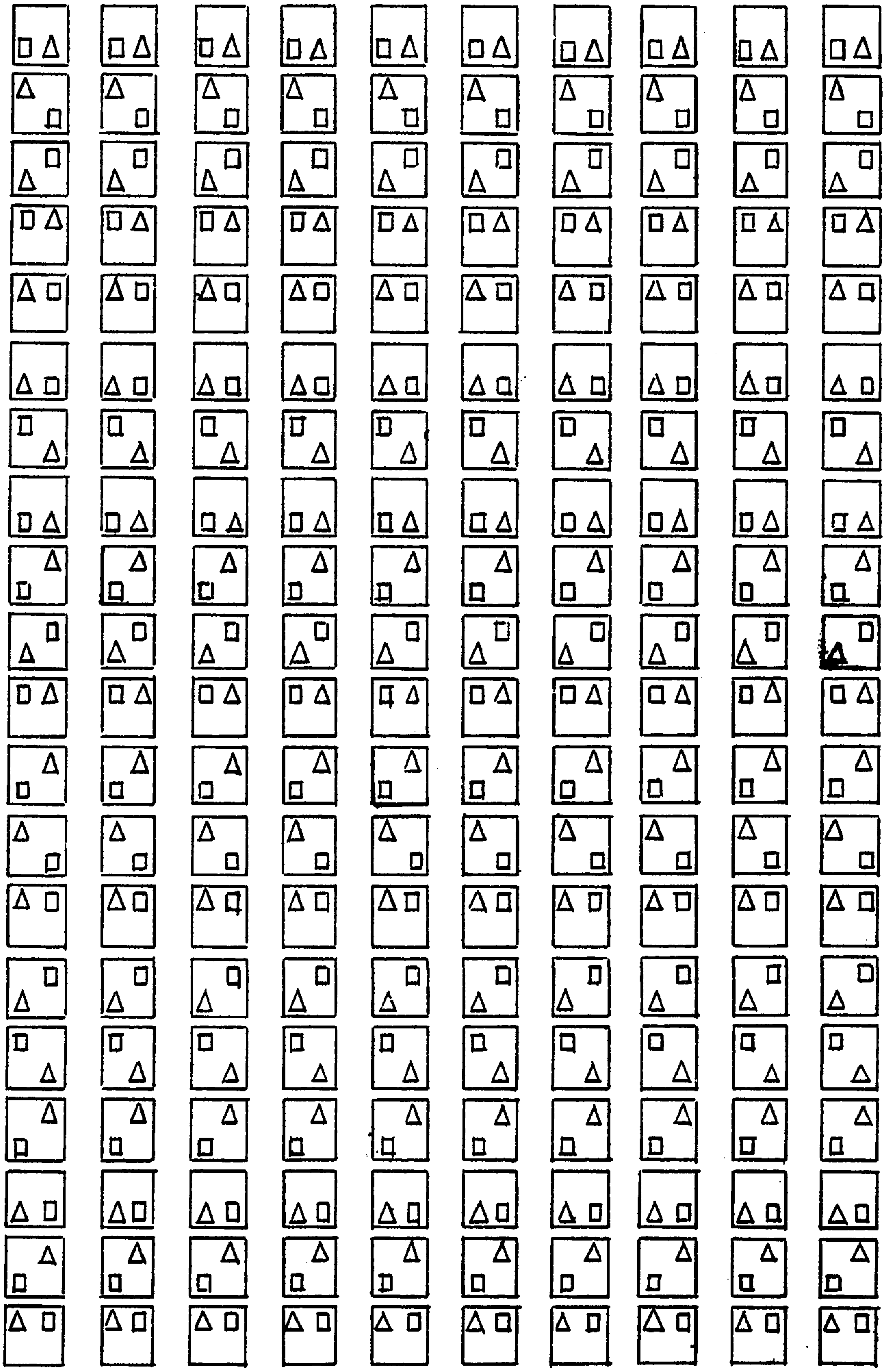

Figure 5. Scoring sheet