ORAL VERSUS ORAL-PICTORIAL PRESENTATION OF A SIMILARITIES TASK

by Sheila D. Roberts

Thesis presented to the School of Psychology and Education of the University of Ottawa as partial fulfillment of the requirements for the degree of Master of Arts in Psychology.

Toronto, Canada, 1962

ACKNOWLEDGMENTS

This thesis was prepared under the direction of Professor Maurice Chagnon, Ph.D., of the School of Psychology and Education of the University of Ottawa.

The writer is indebted to the North York Board of Education and to its Coordinator of Psychological Services, J. G. McMurray, Ph.D., for assistance in obtaining the subjects used in this research, to the staffs of the schools in which testing was carried out, and to the professional colleagues who assisted with the testing and scoring.

CURRICULUM STUDIORUM

Sheila D. Roberts was born September 12, 1933, in Montreal, Quebec. She received the Bachelor of Arts degree from Marianopolis College of the University of Montreal in 1954.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

In similarities tasks such as the one included in the WISC, the subject is asked orally to state in what way two items are alike. Oral presentation refers to the standard presentation of the task in spoken words, and oral-pictorial presentation to the presentation of a line drawing of the items to be compared in addition to the spoken words.

In its present form, the similarities subtest of the WISC is too difficult for young children. Wechsler has attempted to solve this problem by adding a different task, the analogies, at the lower end, to be used with subjects under eight years of age, and others who obtain a score of less than three on the similarities items. Other subjects are automatically given the full four points credit for the analogies portion. However, as it has not been demonstrated that the analogies task measures the same ability as the similarities task, this method of adjusting the difficulty level of the test for young children is questionable.

The WISC is designed for use with subjects from ages five to fifteen years. Yet at age eight, the mean raw score on the similarities portion of the subtest is only three[1].

---

1 The WISC manual shows a mean of seven for the similarities task at age eight, but four of these seven points are for the analogies portion.

The possible scores on the negative side of the mean are only two, one, or zero. The range of intelligence among children of one age group is large. If a test is to discriminate adequately among the individuals, it must reflect this range of ability. With such a low mean score, the similarities task can hardly reflect the wide individual differences in intelligence found in subjects of the same chronological age.

The present study is an attempt to solve this problem that the similarities task does not adequately reflect the range of ability found among subjects at the lower end of the age and intelligence groups. The method used is the varying of the stimulus presented in such a way as to make the task as it stands easier, more comprehensible, and more vivid for the subjects, so that the failure rate will be lessened. Should this method succeed in raising the mean number of successes on the task, a concomitant result should be an increase in the variability of the scores because there would be fewer zero scores. This could take care of the objection that individual differences in ability are not adequately reflected by the test in its present form.

Considering the extensive use made of the WISC in all possible settings, these problems in regard to its use with young children are very much in need of study.

The first chapter consists of a review of the literature pertinent to the topic of this thesis. The use of the

WISC with young children is discussed. The question of varying the nature of the stimulus task is then treated. Differences in the type of responses given are also discussed. The chapter is terminated with a statement of the hypothesis.

The experimental design of the project is described in the second chapter. The sample, the formation of the experimental groups, the nature of the experimental conditions, and the testing procedure are outlined.

The results obtained are presented and discussed in the third chapter. The comparability of the males and females used in the study is demonstrated. The data for the two experimental conditions is then analyzed for differences between means and in variability. The type of responses given under the two conditions is examined by means of an item analysis. The implications for subsequent research are indicated. Photographs of the drawings used and a sample record sheet are found in the appendices.

CHAPTER I

Review of the Literature

This survey of the literature will consider the problems involved in the use of the WISC with young children. Studies concerned with varying the nature of the stimulus task will be considered in the second section, and finally, the question of the type of responses will be discussed.

1. Use of the WISC with Young Children

Both Delp[1] and Anderson[2] in reviewing the WISC, have referred to the problems involved in using this test with young children. Delp discusses the fact that the test is too difficult with lower age brackets. Anderson points out the surprising fact that complete failure on a task, that is, zero raw score, is assigned as much as five units of scaled score. This is hardly a defensible practice. Thus a child who obtained a raw score of one would likely obtain a scaled score of six. The difference in ability

---

1 Harold A. Delp, "Review of the WISC", in O.K. Buros, Ed., Fourth Mental Measurement Yearbook, Highland Park, N.J., The Gryphon Press, 1953, p. 477-479.

2 James M. Anderson, "Review of the WISC", in O.K. Buros, Ed., Fourth Mental Measurement Yearbook, Highland Park, N.J., The Gryphon Press, 1953, p. 477-479.

between him and one child who obtained zero raw score might be small, while between him and another child who obtained zero, the difference in ability might be great. Yet the difference in scaled score would only be one in both cases.

In an item analysis of the WISC using possible mental defectives between the ages of seven years six months and fifteen years eleven months, Carleton and Stacey[3] found that a large number of items contribute little to the total test variance. They felt that the subtests would not be particularly sensitive to revealing individual differences among subjects chosen from a comparable population.

Littell[4], in his review of the research done during the first decade of the WISC's existence points out the fact that a young child may be given subtests with as few as four or five items. He states that the reliability of such short scales is open to considerable question. He recommends adding more items to the lower ends of the subtests.

A possible outcome of the use of drawings of the items in the second experimental condition in this study

---

3 Frederick O. Carleton and Chalmers L. Stacey, "An Item Analysis of the WISC" in the Journal of Clinical Psychology, Vol. 11, 1955, p. 149-154.

4 William M. Littell, "The Wechsler Intelligence Scale for Children: Review of a Decade of Research", in the Psychological Bulletin, Vol. 57, No. 2, issue of March 1960, p. 132-156.

is the lowering of the difficulty level of the test. Thus the use of these drawings might solve the problem in the use of the WISC with young children which these authors have pointed out.

No studies dealing with the Similarities subtest of the WISC specifically were found in the literature. In his article, Littell reviews all the research done on the WISC since its publication in 1949. There is no mention of studies treating of the Similarities subtest independently.

Nor were there any reports of research into the effect of varying the nature of the stimulus task which were directly comparable to the present study. However, the question of using a visual stimulus as opposed to an oral one has been discussed by several authors.

2. Varying the Nature of the Stimulus Task

In his book, The Clinical Interaction[5], Sarason says:

> If an individual performs poorly on a verbal similarities task (e.g. in what way are an apple and a peach alike?), it does not follow that he would necessarily perform as poorly on a similarities task in which the objects are visually presented.

In speaking of a Similarities task in which the objects are visually presented, Sarason is referring to a

---

5 Seymour B. Sarason, The Clinical Interaction, New York, Harper and Brothers, 1954, p. 48.

performance type of similarities task. The subject is presented a number of objects which he is told to group according to any criterion he chooses. When he has finished grouping them, he is asked to give an explanation of his groupings. Thus Sarason suggests that there might be a difference between performance on a verbal similarities task and performance on a similarities task involving real objects. He did not carry out research into the question. In the present study both experimental conditions involved verbal similarities tasks, but a graphic representation of the real objects was presented in addition to the verbal stimulus in the oral-pictorial condition.

Welch, who has done much research into children's thinking, discusses this question of varying the nature of the stimulus task[6]:

> In any learning situation there may be at least three types of media: three-dimensional material, pictorial material, and linguistic material. The number and the vividness of the characteristics of an object are never as great in a picture of that object as in the object itself when we perceive it directly. A description of the object is still more unsatisfactory.... Many problems have been failed simply because the subject did not understand the instructions or the nature of the material which the problem involved..... Some mental patients who had no trouble with these (problems) even when they

---

6 Livingston Welch, "The Transition from Simple to Complex Forms of Learning", in the Journal of Genetic Psychology, Vol. 71, 1947, p. 238-239.

> involved abstract concepts (e.g. the classes of animals, fruit and boats), were unable to solve the same type of problem when words at the concrete level were substituted in place of three dimensional material.

As the thinking of a child is necessarily more dependent on the presence of real objects, he would be likely to experience difficulty more than any other subject when words rather than a visual stimulus is presented.

In the Similarities test, the problem is presented to the subject linguistically. It can be asked whether this form of presentation is vivid enough for children. The similarities task is a difficult one. The child must call to mind the two objects, and compare them. He is presented a word image, and he must call up a visual image from memory. It is possible that many of them fail at this level and do not reach the stage of making the comparison, which is the task on wnich they are being judged.

Speaking of images, Russell[7] says:

> An image is a centrally aroused experience which reproduces in part some previous perceptual experience in the absence of the original sensations. A child's image accordingly, is a product of his own thinking or memory, rather than of direct sensory stimulation. As such it is often incomplete or even distorted in comparison to the original perception, but it is similar to a sensory experience.

---

7 David Harris Russell, Children's Thinking, New York, Ginn and Company, 1956, p. 68.

It is this possibility that the child's image of the objects will be incomplete or distorted which makes the use of pictures of the objects to be compared in the similarities test desirable. Children of the same age will necessarily differ in their ability to call to mind clear images of objects. The Similarities task, however, was not designed to differentiate among the subjects at this level of imagery, but rather at the level of conceptualization. Therefore increasing the stimulus cues by the addition of the pictures puts the subjects on a more equal starting point for their conceptual thinking. Thinking is affected by the individual's mental ability and by his stage of development. Using a single chronological age group controls one aspect of development, and using the pictures controls for developmental differences in ability to form clear images. Thus differences in performance will reflect differences in the area we wish to measure, namely in the area of conceptualization.

Sarason[8] gives the example of a boy who showed a marked difference in adequacy of response between items involving a verbal response to an orally presented problem (verbal items on the Stanford-Binet) and items involving a response to a visually presented stimulus (performance

---

8 Seymour B. Sarason, Op. Cit., p. 49.

items on the Stanford-Binet). On the former, his performance was hesitant and variable in efficiency: "It would seem that when he can formulate a problem in his own way, aided by the constant presence of a visual stimulus, Harold is likely to perform efficiently."

The visual stimulus keeps the task at hand constantly before the mind of the subject. Children as a group tend to be distractible. The pictures of the items in the Oral-Pictorial condition have the advantage of serving to keep the child thinking about the task at hand.

Underwood[9] presents a point of view concerning thinking for the purpose of giving direction to research on certain variables which appear to influence efficiency in thinking. One of the variables he discusses is the mode of presentation of the stimulus. He refers to presenting an object or picture of an object as perceptual presentation of a stimulus, and presenting the object name as symbolic presentation. On this question of mode of presentation he writes[10]:

> Properties of objects may not be remembered when the stimuli are symbolic (words), and if these properties are germane to the solution of

---

9 Benton J. Underwood, "Orientation for Research on Thinking", in Psychological Review, Vol. 59, No. 3, issue of May 1952, p. 209-220.

10 Benton J. Underwood, Op. Cit., p. 213.

> a given problem, solution will not be achieved... To illustrate: If we ask a child what relationship exists between a table and a sheep, it is likely that the common four-leg property will be discovered more readily if the sheep and the table are presented perceptually (real or by pictures), than if the child must depend upon memory as he searches for common properties.
>
> On the other hand, if the forgotten properties of the symbolized objects are irrelevant and the remembered properties relevant, solution might be more rapid for symbolic presentation than for perceptual presentation. For, in the case of perceptually-presented objects, properties which were forgotten for the symbolized objects may actually serve as potent distracting stimuli.

In order to decide which mode of presentation is more fruitful in a particular situation, Underwood indicates that we need to know the dominant responses to a perceptual stimulus and to the corresponding symbolic one. Thus it is necessary to try out both and compare the types of responses elicited. In the present study, the method which resulted in the lower failure rate would be preferred. Another consideration would be the type of response elicited by the two methods. The responses to the similarities task can be of three types: abstract-conceptual, functional, and concrete. Would one mode of presentation tend to result in one of these types of responses in preference to the other two? This question of the type of responses elicited is taken up again later.

There are no studies where this suggested comparison of mode of presentation of a similarities task has been made.

However, there have been studies of reasoning and concept formation which are pertinent.

Two studies by Long and Welch compared pictorial with written word presentation. In a study in 1941[11], involving children from six to eight years, Long and Welch used a variety of problems to evaluate their reasoning ability. They found that twice as many children (ten out of fifteen as opposed to fifteen out of fifteen), were successful in solving picture problems than were successful in solving written-word problems. The same authors, in 1942[12], used more complex problems with children from eight to eleven years six months. In this study, they found no significant difference when pictures or written words were used.

An important difference from the present study is that written rather than spoken words were used. The written word is constantly before the subject to stimulate and facilitate his thinking about the problem. The standard mode of presentation of the similarities task, called the Oral Condition in this experiment, involves a spoken word stimulus only.

---

11 L. Long and L. Welch, "Reasoning Ability in Young Children", in the Journal of Psychology, Vol. 12, issue of July 1941, p. 21-44.

12 L. Long and L. Welch, "Influence of Level of Abstractness on Reasoning Ability", in the Journal of Psychology, Vol. 13, 1932, p. 41-59.

In a study by Maier[13], the process used in solving one problem is helpful in solving a subsequent one. When the model of the first was left intact, the second problem was solved more quickly and more frequently. According to Maier, the immediate presence of the structure was more effective than a memory of the structure.

The present study is based on the assumption that a picture of the objects to be compared will be more effective than the child's memory of the object. The task will be more vivid and more meaningful to the children. It is anticipated that this will lower the floor of the test, reducing the percentage of children who obtain a zero score.

A number of studies concerned with the mode of presentation of a task have been discussed. There remains the question of differences in the type of responses elicited by the various modes of presentation of a task.

---

13 Norman R. F. Maier, "Reasoning in Humans III. The Mechanisms of Equivalent Stimuli and of Reasoning", in the Journal of Experimental Psychology, Vol. 35, 1945, p. 349-360.

3. Type of Response

In discussing the adult Wechsler-Bellevue test, Rapaport[14], states that verbal concept formation is reflected in the similarities task. Concept formation informs human beings about the "belonging together" of the objects and events of his everyday world. According to Rapaport:

> The definitions offered are usually based on the verbal coherence established in everyday experience between the objects and their generic terms, (coat - dress - clothing). The subject has relatively simple task as long as this verbal coherence is operating. If, however, it has never been created or is weakened, the subject may follow either of two courses. First, he may take reference to images of the object, which generally results in concretistic or at best functional definitions. Or he may attempt generalizations.

Thus, Rapaport distinguishes three levels of concept formation: concrete (both round), functional (you eat both), and abstract-conceptual (both fruit). Rapaport indicates that the subject faces a different task if his conceptual thinking has not developed to that degree at which the verbal coherence between the objects and their generic terms is established. This must be kept in mind when considering a similarities task for children. A child's conceptual thinking is just in the process of developing. It

---

14 David Rapaport, Diagnostic Psychological Testing, The Theory, Statistical Evaluation, ane Diagnostic Application of a Battery of Tests, Vol. I, The Year Book Publishers Inc., 1946, p. 394.

seems likely that his conceptual thinking would not be developed to the extent that the "verbal coherence" mentioned would be established. Thus while an adult is likely to produce a correct response spontaneously, a child would have to refer to a mental image of the object. Rapaport feels this will result in either a concrete or a functional response. Can we not expect abstract-conceptual responses from children? This question of the type of responses that we can expect from children is an important one. The results of the present study will be examined in the light of this question.

Also in regard to type of responses, a study with the Goldstein-Weigl Color-Form and Sorting Test by Reichard, Schneider, and Rapaport[15], approached the problem from the developmental standpoint using children as subjects. They wanted a set of developmental norms for the evaluation of clinical anomalies.

This sorting test consists of two parts. In the first part, the subjects form their own groupings of the objects and give an explanation of their groupings. Children of all ages were more successful on this first part than on the second part of the test. Results on the first

---

15 S. Reichard, M. Schneider, and D. Rapaport, "The Development of Concept Formation in Children", in the American Journal of Orthopsychiatry, Vol. 14, issue of Jan. 1944, p. 156-161.

part showed a steady increase with age in ability to group together objects which belong together and in ability to give abstract-conceptual explanations of the groupings. The results appear to indicate three levels of conceptual development: concrete responses were characteristic of children up to age eight, functional responses reach a peak at eight or nine, and abstract-conceptual appear to reach relative maturity around age eleven.

In the second part of the test, the subjects had to define conceptually groups formed by the examiner. As mentioned, the failure rate was higher here. This method of presentation is more comparable to the present study in that the subjects do not form their own groupings, but rather a grouping is presented to them and they must state the relationship that exists among its members. Likewise, in the similarities task, two items are presented to the subjects and they must state their relationship.

In the first part of the test, the percent of conceptual responses does not begin to exceed the percent of functional until the age of nine. In the second part, the percent of conceptual responses outnumbers the percent of functional at all age levels. The results of the present study will be compared to these contrasting findings.

In addition to this study using a performance test, there have been studies of the type of responses to the

vocabulary subtest of the WISC and Wechsler-Bellevue Intelligence Scale. In two such studies, Stacey and Portnoy[16, 17], investigated the relationship between the number of concrete, functional, and abstract-conceptual responses given and both age and I.Q. The population in their first study was of borderline intelligence and lower, and in the second of dull normal intelligence and lower.

In regard to intelligence, they found a greater number of concrete than functional responses with higher I.Q. In the first study, involving younger subjects, this result is significant at the one percent level. They also found a trend toward a greater proportion of abstract-conceptual responses with higher I.Q.

In regard to age, the older subjects gave a greater proportion of abstract-conceptual responses than the younger: the older subjects gave a greater proportion of concrete responses than the younger: and the younger subjects gave a greater proportion of functional responses than the older subjects.

---

16 Chalmers L. Stacey, and Bernard Portnoy, "A Study of the Differential Responses on the Vocabulary Sub-test of the WISC", in the Journal of Clinical Psychology, Vol. 6, No. 4, issue of Oct. 1950, p. 401-403.

17 Chalmers L. Stacey, and Bernard Portnoy, "A Study of the Differential Responses on the Vocabulary Sub-test of the Wechsler-Bellevue Intelligence Scale", in the Journal of Clinical Psychology, Vol. 7, No. 2, issue of April 1951, p. 144-148.

Stacey and Portnoy's results in regard to the incidence of concrete and functional responses differ from the results of Reichard, Schneider, and Rapaport which were just discussed[18]. However, the task used differed in the two cases, and also Stacey and Portnoy used subjects of below normal intelligence.

In the present study, the question of the type of the responses resulting under the two conditions will necessarily have to be considered. The reason is that the use of the pictures of the items might cause the subjects to focus on the concrete properties of the objects and thus result in concrete responses more so than would be the case using the standard form of presentation of the similarities task. It is conceivable that the oral-pictorial mode of presentation could raise the subject's over-all score by increasing the number of items successfully handled, while at the same time lowering the qualitative level of the responses by discouraging the production of abstract-conceptual responses. For this reason, an item analysis of the type of responses will be carried out.

4. Summary and Hypothesis

In summary, it has been indicated by a number of authors that the difficulty level of the WISC is too high.

---

18 See page 12.

The use of drawings in the oral-pictorial condition may serve to lower the difficulty level of the similarities task.

A number of studies which treat of the effect of varying the nature of the stimulus task have been discussed. The present study involves a comparison of two methods of presenting the same similarities task.

Finally, the question of the type of responses produced has been considered. This will also enter into the present study in comparing the types of responses elicited by the two methods of presentation of the similarities task used, namely, oral and oral-pictorial presentation.

The hypothesis may thus be stated in null form as follows: there is no significant difference in performance on a similarities task between a group presented an oral stimulus alone and an equivalent group presented an oral-pictorial stimulus.

CHAPTER II

EXPERIMENTAL DESIGN

This chapter will describe the sample, the formation of the experimental groups, the experimental conditions, and the procedure followed in an experiment designed to test the hypothesis proposed at the end of the preceding chapter.

1. The Sample

The final sample for this study consisted of 140 eight-year-olds, seventy of them male and seventy female, drawn from two public schools in the Township of North York, which is part of Metropolitan Toronto. It was decided to use eight-year-olds because the success of children at lower ages on the similarities task is so limited that there would be little scope for comparison between the two groups. The two schools were chosen because they were situated in areas whose populations were considered to be heterogeneous. All the eight-year-olds in these two schools whose entire schooling so far had been in the North York system were included. This restriction was decided upon because North York students are given a Primary Mental Abilities test in grade one, and this information in regard to intelligence was considered desirable in order to describe the population. It resulted in the elimination of nineteen eight-year-olds who had been

transferred from other school systems. This left 146 subjects, seventy-two boys and seventy-four girls. It was decided to assign all of these to the experimental conditions being mindful that children from among the boys or the girls might be absent during testing. It was felt that there should be an equal number of each sex in the final group.

There were two boys absent and one girl transferred from school at the time of testing, so three girls were eliminated randomly such that there was an equal number of boys and girls in both experimental conditions.

According to the Primary Mental Abilities test given in grade one, the range of I.Q.'s was 82 to 136, with a mean of 105.5[1]. For this experiment eight-year-olds were considered to be those between seven years, seven months and eight years, six months.

2. Experimental Groups

An equal number of boys and girls was assigned to the two groups by means of random numbers using the tables

---

1 The high mean I.Q. results from the fact that those with I.Q.'s below 80 are removed from the school and placed in centrally located special classes.

of Arkin and Colton[2]. The seventy-two boys and seventy-four girls were separately listed alphabetically, and assigned numbers from one to seventy-two and seventy-four respectively. The table of random numbers was entered randomly, and all numbers between one and seventy-four were assigned in alternating pattern to the two experimental conditions. As just mentioned some subjects were not present during testing, so others were eliminated such that there was an equal number (thirty-five) of boys and girls in both experimental conditions. The total sample was 140 subjects.

3. Experimental Conditions

The first experimental condition involved the presentation of the twelve items of the similarities subtest of the WISC, using the directions in the WISC manual[3]. According to these directions the examiner said: "In what way are a (plum) and a (peach) alike?" This experimental condition is referred to as oral presentation because a spoken stimulus was presented to the subjects.

---

2 Herbert Arkin and Raymond R. Colton, Tables For Statisticians, New York, Barnes and Noble, 1950, p. 142-145.

3 David Wechsler, Wechsler Intelligence Scale For Children, New York, The Psychological Corporation, 1949, p. 66.

The second experimental condition consisted of the presentation of a line drawing of the two items to be compared in addition to the verbal stimulus given in the first experimental condition. These line drawings were done with india ink on cards of three by five inches[4]. Drawings of the items were submitted by three persons. Those considered to be the best were selected. A pilot study revealed faults in two items which were subsequently corrected.

In this second condition, the subjects were presented two cards with drawings of the items to be compared. At the same time, the examiner said: "In what way are a (cat) and a (mouse) alike?" This is referred to as oral-pictorial presentation because the stimulus consisted of both spoken words and a drawing.

4. Experimental Procedure

The testing was carried out by two qualified psychologists each of whom did half the testing under each experimental condition in order to eliminate the effects of examiner bias.

To avoid the possibility that the subjects might be alarmed when summoned out of the classroom, the two examiners introduced themselves to each class as a whole. They asked

---

4 Photographs of the drawings used are found in Appendix 1.

several simple questions and made comments intended to establish a relaxed atmosphere. They then informed the subjects that the examiners would like to see them individually for a few moments. The examiners carried on sufficient conversation with each individual to establish rapport before beginning testing.

Two rooms suitable for testing were available at both schools. The testing was done over a period of three days. The time needed per child was approximately twelve to fifteen minutes.

The responses were recorded on a specially devised record sheet[5]. The scoring was done by a third qualified psychologist who was not involved in the testing. In order that the scorer would not be influenced in his scoring by the information about the subjects which is contained on the record sheet, the tops of the record sheets were folded back to hide the information from view. The responses were given a score of two, one, or zero based on the criteria set forth in the WISC manual[6]. A total score was then obtained for each subject.

---

5 A sample record sheet is found in Appendix 2.

6 David Wechsler, Op. Cit., p. 95-97.

CHAPTER III

PRESENTATION AND DISCUSSION OF RESULTS

This chapter is concerned with the compilation and analysis of the results of the experiment in the light of the null hypothesis and of the previous studies mentioned in the review of the literature. The first section considers the comparability of the males and females used in the sample. The second section involves the question of differences between means and differences in variability for the two experimental conditions. The type of responses given is then investigated by means of an item analysis. A final section is devoted to a summary of these findings.

1. Comparability of Males and Females

The sample for this experiment consisted of an equal number of males and females. Before they can be treated as a uniform population, their comparability must be demonstrated. Therefore the results of the males are compared to those of the females under the two conditions. The means, the standard deviations, and the standard errors of the means of the males and females are presented in Table I. The mean for the males under the oral condition is 3.5, and for the females, 3.1. This difference of .4 is not significant, the value of P being .37. In the oral-pictorial condition, the mean of

Table I.-

Means, Standard Deviations, and Standard Errors of the Means of the Males and Females for the Oral and Oral-Pictorial Groups

Condition	Sex	N	Mean	σ	σ_M
Oral	Males	35	3.5	2.05	.351
	Females	35	3.1	1.69	.289
Oral-Pictorial	Males	35	4.2	2.05	.351
	Females	35	3.9	1.70	.291

the males is 4.2, and of the females, 3.9. This difference of .3 is also not significant, the value of P being .50. These results are found in Table II.

There is a difference in standard deviations between the males and females of .36 and .35 under the oral and oral-pictorial conditions respectively. Because of the size of the sample, the F-test of differences between the two variances was used[1]. In the case of the comparison of males to females under the oral condition, the F is 1.47, and under the oral-pictorial condition, the F is 1.46. Neither of these is significant. These results are found in Table III.

Thus, there is not a significant difference in variability between the males and females under either condition. It has just been seen that the difference between the means of the males and females was also not significant under either condition. The conclusion can therefore be drawn that the males and females used were comparable and can be considered as a uniform population.

It was then possible to proceed to a consideration of the differences between means and differences in variability under the two conditions for the total population

---

1 Lawrence T. Dayhaw, Manuel de Statistique, Ottawa, Editions de l'Universite d'Ottawa, 1958, p. 369.

Table II.-

Comparison of the Mean of the Males with the Mean of the Females in the Oral Group and in the Oral-Pictorial Group

Condition	D_M	σ_{dM}	CR	P
Oral	.4	.446	.90	.37
Oral-Pictorial	.3	.457	.66	.50

Table III.-

Comparison of the Variance of the Males with the Variance of the Females in the Oral Group and in the Oral-Pictorial Group

Condition	D_{σ}	F	Significance
Oral	.36	1.47	N.S.
Oral-Pictorial	.35	1.46	N.S.

of 140 subjects.

2. Differences between Means and in Variability

The means, the standard deviations, and the standard errors of the means were calculated for the two experimental groups. These are presented in Table IV. The mean of the oral group was 3.3, and the mean of the oral-pictorial group was 4.0. The standard deviation in both cases was 1.89, and the standard error of the mean was .227 for both means. In order to determine if the difference between means of .7 was significant, the standard error of the difference between means and the significance of the difference were calculated. The standard error of the difference between means was .321, and the critical ratio, 2.18. Thus the difference between means is significant at the three percent level of probability.

A one percent level of significance is usually demanded in psychological research in order to indicate a difference greater than chance. However, a difference significant at the three percent level would strongly suggest the possibility of a true difference.

In the present study, a significant difference in variability would be required in addition to a significant difference in means in order to reject the null hypothesis.

Table IV.-

Comparison of the Mean of the Oral Group with the Mean of the Oral-Pictorial Group

Condition	N	Mean	σ	σ_M	σ_{dM}	CR
Oral	70	3.3	1.89	.227		
					.321	2.18
Oral-Pictorial	70	4.0	1.89	.227		

The reason for this is that the present study is an attempt to solve the problem that the similarities task does not adequately reflect the range of ability found among subjects of the same age group. In order to reflect this variability in intelligence, a task must result in a wider variability in scores than results from the standard administration of the similarities task.

As can be seen in Table IV[2], however, there is no difference in variability when the standard deviation of the oral condition is compared to that of the oral-pictorial condition for the total group. The null hypothesis can therefore not be rejected.

A final consideration in this presentation and discussion of results was the question of differences in the types of responses given under the two conditions. This was investigated by means of an item analysis.

As mentioned in the review of the literature, vocabulary tests and other tasks have been studied in regard to the types of responses they elicited. It was felt that it would likewise be valuable to examine the type of responses given to the similarities task by the age group used in this study.

---

2 See page 28.

3. Item Analysis: Type of Response

It was noted in the review of the literature that the responses to a similarities task can be of three types conceptually: abstract-conceptual, functional, or concrete. The distribution of the responses given under the two experimental conditions according to these three types is found in Table V. The results obtained in this experiment were analyzed for possible differences in the type of responses resulting under the two experimental conditions. It could be that the presentation of a drawing of the items to be compared might favour the production of concrete responses.

In order to determine if this were so, an item analysis of the type of the responses was made. The results of the chi-squares are presented in Table VI. Only item one reached an acceptable level of significance. The level of probability for item one is .01. The number of abstract-conceptual responses to item one is five under both the oral and oral-pictorial conditions. There are zero functional responses under the oral, and two under the oral-pictorial condition (see Table V). Thus the significance must lie in the number of concrete as opposed to the number of unacceptable (scored zero) responses. Under the oral condition, there are thirty-six concrete and twenty-nine unacceptable responses, and under the oral-pictorial there are fifty-three concrete and ten unacceptable responses. The chi-square

Table V.-

Distribution of Responses to the Twelve Items According to the Type of Response Under the Oral and Oral-Pictorial Condition

Item	Abs.-Conc.		Functional		Concrete		Unaccept.	
	O	O-P	O	O-P	O	O-P	O	O-P
1	5	5	0	2	36	53	29	10
2	6	6	0	0	40	43	24	21
3	4	6	21	33	0	0	45	31
4	2	6	35	40	13	15	19	9
5	0	0	16	15	4	3	50	52
6	3	8	4	5	0	0	63	57
7	6	3	0	0	3	0	61	67
8	0	1	0	0	0	1	70	68
9	0	0	1	0	0	0	69	70
10	0	0	0	0	0	0	70	70
11	0	0	0	0	0	0	70	70
12	0	0	0	0	0	0	70	70
Totals	27	35	77	95	96	115	640	595
Percent	3.2	4.2	9.2	11.3	11.4	13.7	76.2	70.8

Table VI.-

Comparison of the Oral and Oral-Pictorial Groups According to the Types of Responses Given to the Twelve Items

Item	Chi^2	P
1	14.48	.01
2	.32	.90
3	5.64	.20
4	5.04	.20
5	.24	.90
6	2.72	.30
7	4.28	.20
8-12[a]		

a Only three out of 140 subjects obtained any score on these items.

of this difference is significant at the .001 level of confidence.

A survey of the actual responses given reveals that fourteen of these thirty-six concrete responses to item one under the oral condition as opposed to forty-one of the fifty-three concrete responses to item one under the oral-pictorial condition is the response "both round". Thus there is a significant difference here between the two conditions because the pictorial stimulus makes it easier for the subjects to give a response based on the fact that the items are similarly shaped. The fact that they are both round is immediately evident in a drawing. The drawings therefore, did not affect the quality of the responses by encouraging concrete responses at the expense of abstract-conceptual as was thought possible. The use of the drawings did reduce the number of unacceptable responses to item one.

The fact that the difference between means of the oral condition and the oral-pictorial condition is significant at the three percent level is also explainable by this discrepancy in the responses to item one.

There remains to be discussed the relative frequency of the different types of responses (abstract-conceptual, function, and concrete). The number of responses of each type are recorded by item in Table V[3]. The total percent of

---

3 See page 31.

each column is found in the last row. Inspection of these percents reveals that concrete answers lead numerically, followed by functional, and then by abstract-conceptual. This seems to support the generally held opinion that functional responses are superior to and appear later than concrete responses. It agrees with the results of Reichard, Schneider, and Rapaport, but not with the studies of Stacey and Portnoy.

It must be borne in mind that the first-mentioned study involved a sorting task, the second, the vocabulary subtest of the WISC, and the present study, a similarities task. They have in common the fact that they all involve concept formation, and inferences regarding a child's level of concept formation are drawn from them.

As mentioned previously[4], Reichard, Schneider, and Rapaport found concrete responses characteristic of children up to age eight, functional responses reach a peak at eight or nine, and abstract-conceptual appear to reach relative maturity around age eleven.

Stacey and Portnoy[5], on the other hand, found that the number of concrete rather than the number of functional responses were higher with both age and I.Q.

---

4 See pages 12 and 13.

5 See pages 14 and 15.

In regard to the present study, reference to Table VII, where the percent of each type of response on the individual items for the oral condition[6] is presented, will cast doubt on the apparent hierarchy of abstract-conceptual, functional, then concrete responses based on the overall percents of each as seen in Table V[7]. For items one, two, and seven, the number of concrete responses far exceeds the number of functional, of which there are none. On the other hand, the functional responses are more numerous on items three, four, five, six and nine, although the difference is not as great as the reverse relationship is on the first two items. It would seem, therefore, that the likelihood of obtaining concrete responses to the similarities task in preference to functional, or vice-versa, varies depending upon the particular item.

The number of abstract-conceptual responses is small on all items compared to the number of concrete or functional, except on items six and seven. On item seven, there are twice as many abstract-conceptual as concrete responses (six versus three). Again, this is the result of the nature of the particular item. It is "in what way are scissors and a

---

6 The results are given for the oral condition only in order that these may serve as a standard against which an individual eight-year-old's responses to the WISC similarities task under its usual mode of presentation can be compared.

7 See page 31.

Table VII.-

Percentage Distribution of the Responses to the Twelve Items According to Type of Response Under the Oral Condition Alone

Item	Abs.-Conc.	Functional	Concrete	Total Accept.
1	7.2	0	51.4	58.6
2	8.6	0	57.1	65.7
3	5.7	30.0	0	35.7
4	4.3	50.0	18.6	72.9
5	0	22.9	5.7	28.6
6	4.3	5.7	0	10.0
7	8.6	0	4.3	12.9
8	0	0	0	0
9	0	1.4	0	1.4
10	0	0	0	0
11	0	0	0	0
12	0	0	0	0

copper pan alike?" The abstract-conceptual response "both are metals" would appear to be easier than a concrete answer such as "both shiny".

These results in regard to the type of response given do not support the view that functional responses per se are of a higher level conceptually than concrete. Further study of this question of the type of responses given to a similarities task is warranted using children of different age levels. If a trend were found, it could serve as a norm against which the performance of an individual child could be evaluated in relation to his age.

It is of interest to note what the results of this study indicate in regard to the order of presentation of the items of the similarities task in relation to their difficulty level. Reference to the last column of Table VII[8], where the total percent of acceptable responses for each item is given, suggests that the items are not arranged in ascending order of difficulty. Item four has the highest percent of correct responses: 72.9 percent, as compared to 58.6 percent for item one, 65.7 percent for item two, and only 35.7 percent for item three. Thus the first four items seem to be out of order as far as difficulty level is concerned. This question of the order of presentation of the items

---

8 See page 36.

could be important in the administration of the test because testing is discontinued after a certain number of items is failed. Further research into this question of the order of difficulty of the items would be of value.

4. Summary

The aim in this research project was to investigate whether altering the nature of the stimulus presented in the similarities task by adding a graphic representation of the object would make the task easier for the subjects, raise the group mean, and thus allow more scope for variability in scores. It was felt that if this were so, the results would better reflect the range of ability in the population. Although the difference between means was significant at the three percent level, further analysis revealed that this was the result of a large difference on one item. The difference in the number of successful responses to this one item was in turn due to the fact that the pictures of the item made one particular response easier to evoke. Thus the addition of the pictorial representation failed to affect the success of the subjects on the twelve items in general.

No significant difference in variability was found. The oral-pictorial condition failed to differentiate better among the subjects. The null hypothesis could therefore not be rejected.

In regard to the type of responses, few abstract-conceptual responses were given on any items. The expected difference between the number of concrete and functional responses in favour of concrete was not found as a general trend. Rather, some items tended to result in concrete responses predominantly, while others favoured functional responses.

While no quantitive difference was found between the two modes of presentation, both examiners felt that the children in the oral-pictorial group showed more interest and enthusiasm for the task, and seemed less threatened by it. In the WISC, the similarities subtest is the fourth of five verbal subtests. With young children, a definite improvement in their motivation is noted when the verbal portion of the test is completed and the performance part is reached. The verbal subtests seem too much like classroom material to them. While the use of the drawings may not improve performance quantitatively, their use might have a favourable effect on the subject's interest and motivation level.

SUMMARY AND CONCLUSIONS

In the comparison of two modes of presentation of a similarities task, it was found that the hypothesis of no significant difference in performance under the two conditions was upheld. The use of a pictorial representation of the items in addition to the oral stimulus did not lower the difficulty level of the task, or increase the variance among scores obtained by the subjects. It thus failed to discriminate more accurately among the subjects.

In addition to this principal problem, the question of the type of responses given under the two conditions was investigated. Very few responses were of the highest level of conceptualization, abstract-conceptual. It was found that the predominance of concrete or functional responses differed from item to item. It cannot be stated, therefore, that eight-year-olds tended in general to produce one or the other more commonly. Further study of this question of the quality of responses to a similarities task should be made, using children of different age levels. If a trend should be found, the performance of an individual child could be compared to the norm for his age.

The question of the order of difficulty of the items of the similarities task was considered. The results of this study suggest that the items may not be in order of

increasing difficulty. Further study of this problem was recommended.

BIBLIOGRAPHY

Carleton, Fredrick O. and Chalmers L. Stacey, "An Item Analysis of the WISC", in the Journal of Clinical Psychology, Vol. 11, 1955, p. 149-154.

This study points out the fact that the WISC subtests are not sensitive to revealing individual differences among certain groups of subjects. This is the problem which led to the undertaking of the present study.

Long, L. and L. Welch, "Reasoning Ability in Young Children", in the Journal of Psychology, Vol. 12, issue of July 1941, p. 21-44.

A number of problems was used to study reasoning in children of six to eight years. It compares their ability to solve problems of varying levels of abstractness.

--------, "Influence of Level of Abstractness on Reasoning Ability", in the Journal of Psychology, Vol. 13, 1942, p. 41-59.

Similar to the preceding study, using subjects of eight to eleven years, six months.

Rapaport, David, Diagnostic Psychological Testing, the Theory, Statistical Evaluation, and Diagnostic Application of a Battery of Tests, Vol. I, The Year Book Publishers Inc., 1946, xi-573 p.

Devotes a chapter to verbal concept formation as it is reflected in a similarities task. Discusses the different levels of concept formation, an important consideration in the present study.

Reichard, S., M. Schneider, and D. Rapaport, "The Development of Concept Formation in Children", in the American Journal of Orthopsychiatry, Vol. 14, issue of Jan. 1944, p. 156-161.

Studied the different levels of concept formation resulting on a sorting task, comparing the characteristic performance of different age groups. Serves as a source of comparison with the results of the present study.

Russell, David Harris, Children's Thinking, New York, Ginn and Co., 1956, xii-449 p.

A comprehensive review of research and findings in the area of children's thinking. A useful reference text.

Sarason, Seymour B., The Clinical Interaction, New York, Harper and Bros., 1954, x-425 p.
Discusses variables which affect the interaction of clinician and client, and also test performance. Treats of the effect of varying the nature of the stimulus task. Served as the impetus to the present study.

Stacey, Chalmers L. and Bernard Portnoy, "A Study of the Differential Responses on the Vocabulary Sub-test of the WISC", in the Journal of Clinical Psychology, Vol. 6, No. 4, issue of October 1950, p. 401-403.
Studies the different levels of concept formation resulting on a vocabulary task. Serves as a source of comparison with the results of the present study.

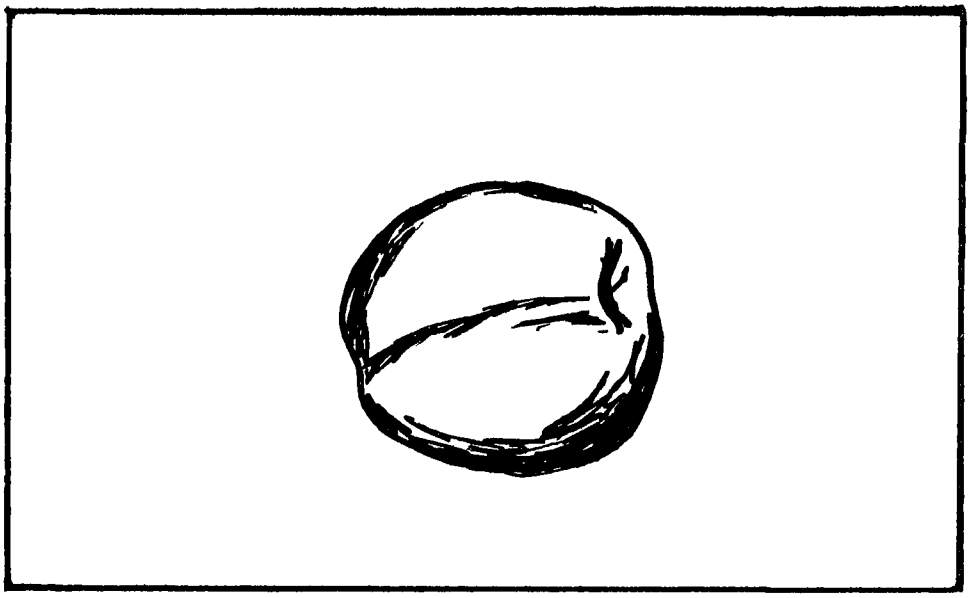
--------, "A Study of the Differential Responses on the Vocabulary Sub-test of the Wechsler-Bellevue Intelligence Scale", in the Journal of Clinical Psychology, Vol. 7, No. 2, issue of April 1951, p. 144-148.
Similar to the preceding study, using older subjects.

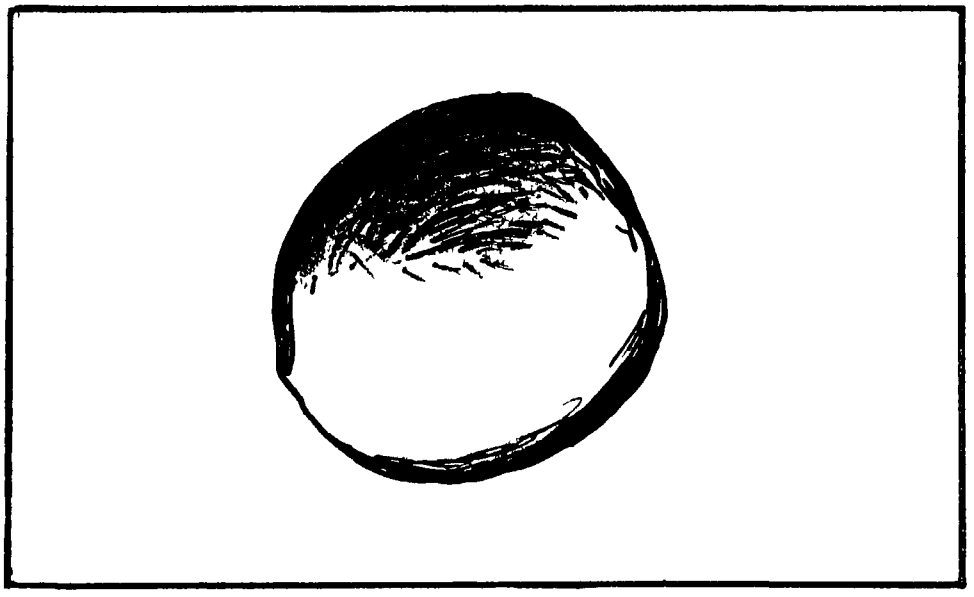
Underwood, Benton J., "Orientation to Research on Thinking", in the Psychological Review, Vol. 59, No. 3, issue of May 1952, p. 209-220.
Discusses variables to be considered in research on thinking, among them, varying the nature of the stimulus task. This is the experimental variable in this study.

Welch, Livingston, "The Transition from Simple to Complex Forms of Learning", in the Journal of Genetic Psychology, Vol. 71, 1947, p. 223-252.
Describes a number of tasks used in the study of problem solving which vary in degree of abstractness. An important study in the area of reasoning.

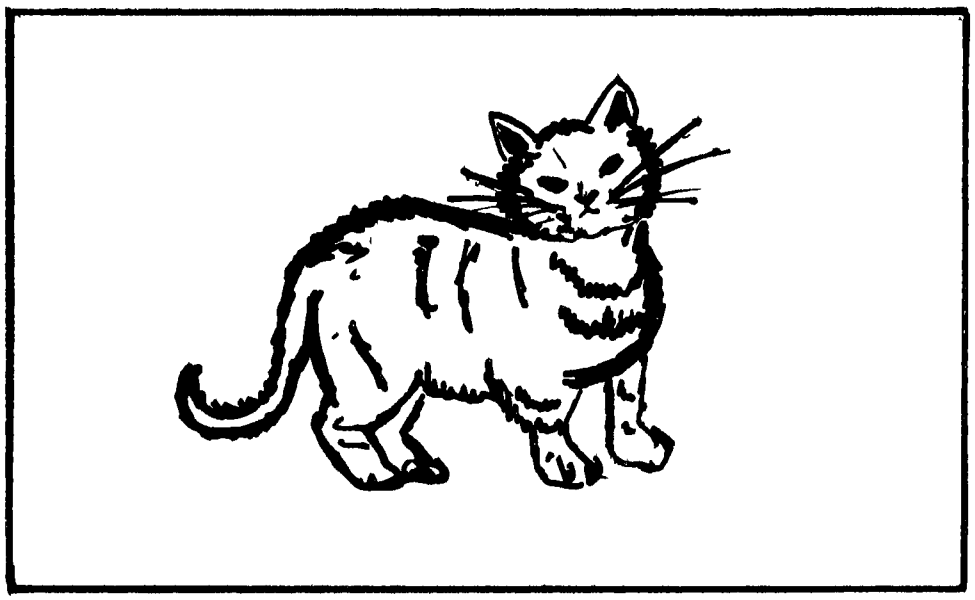
APPENDIX 1

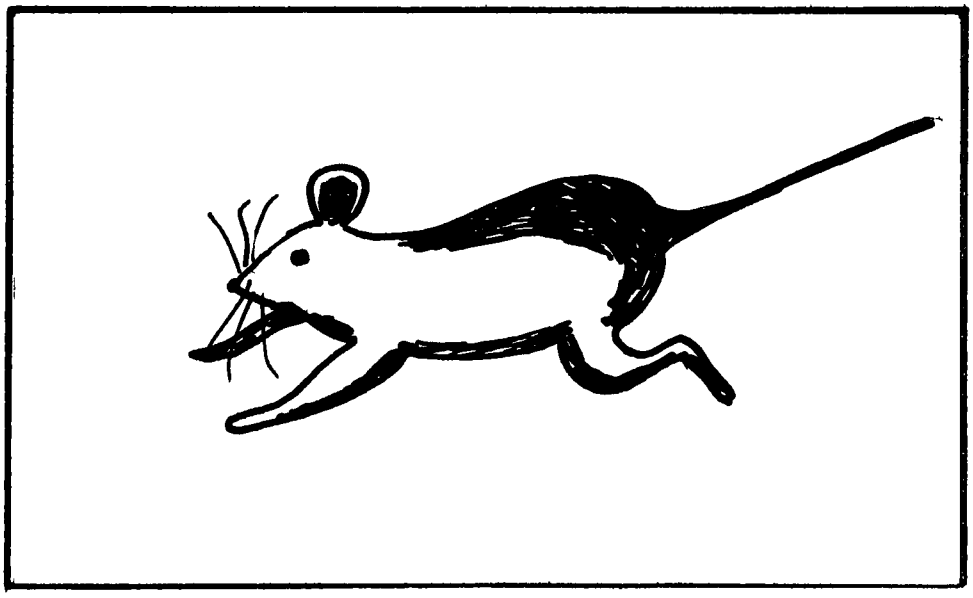
PHOTOGRAPHS OF THE DRAWINGS



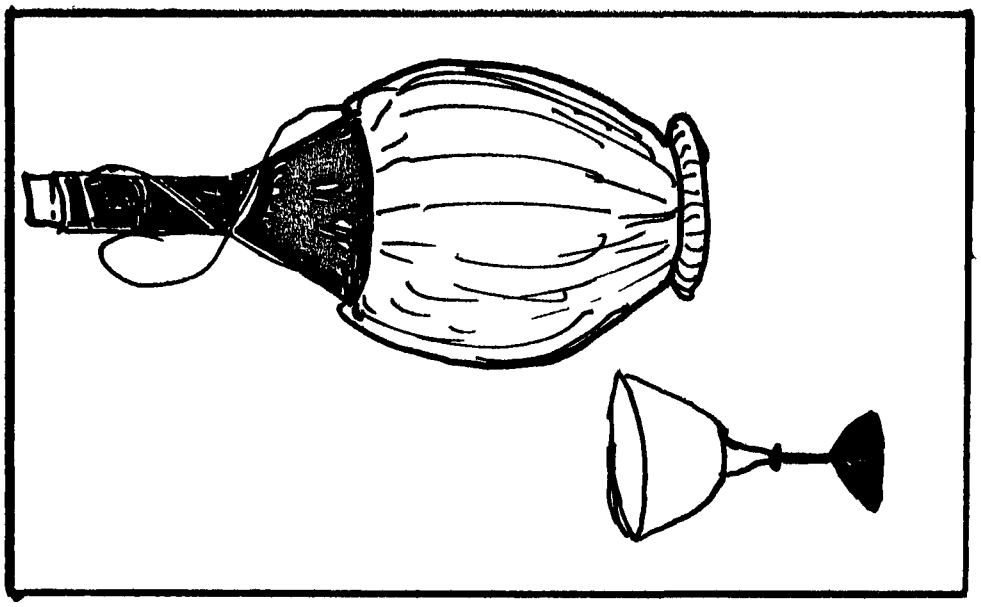


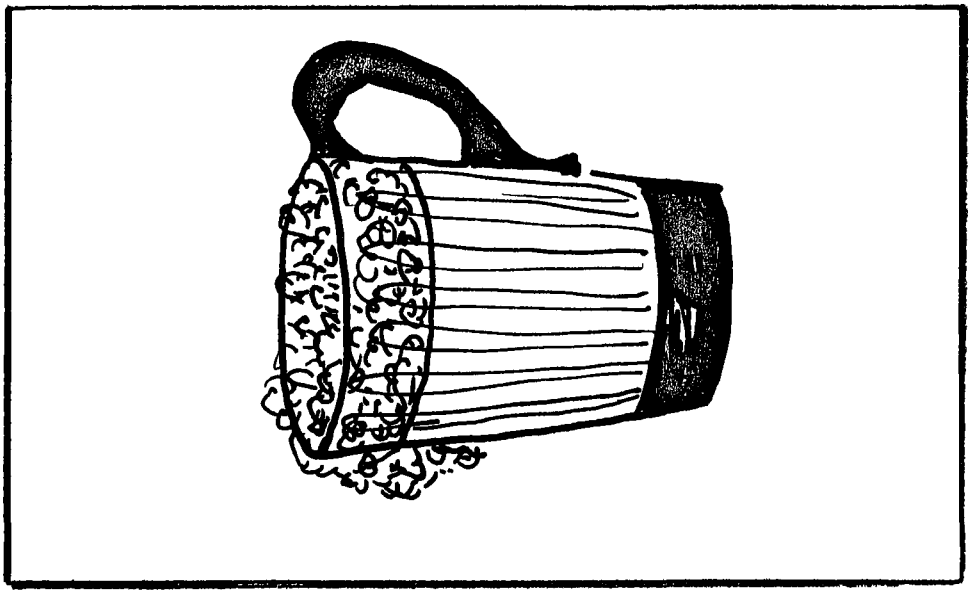
PLUM — PEACH



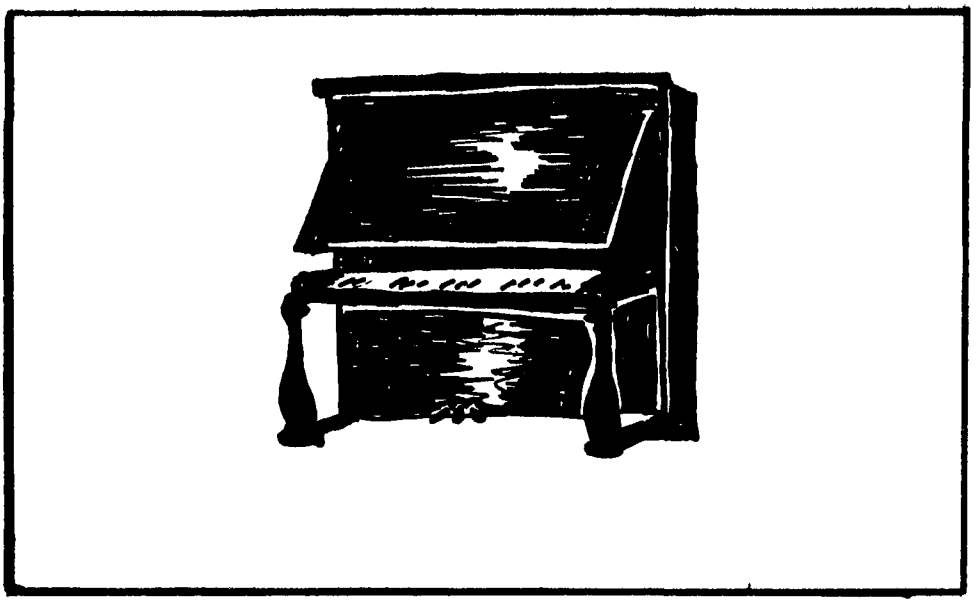


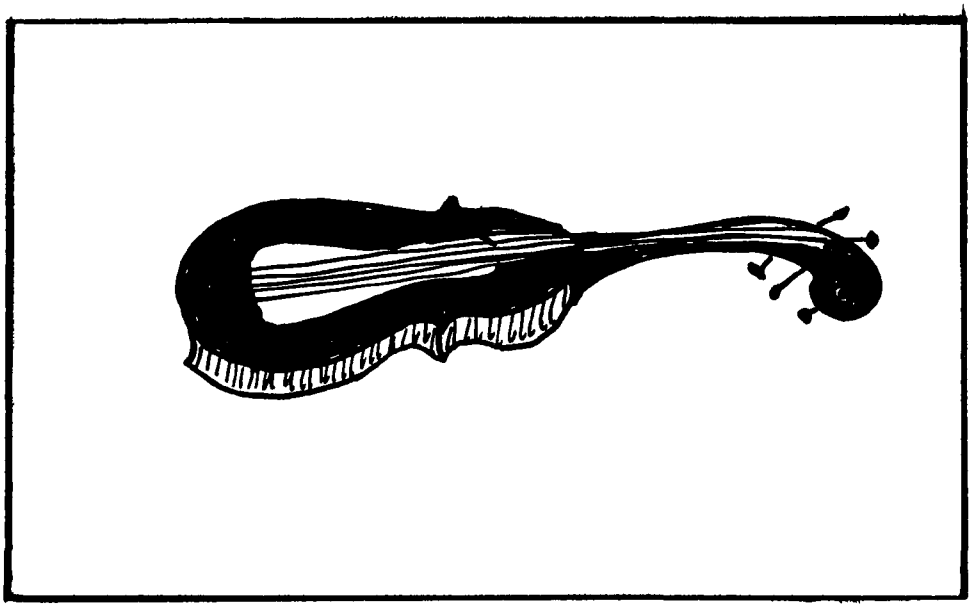
CAT – MOUSE



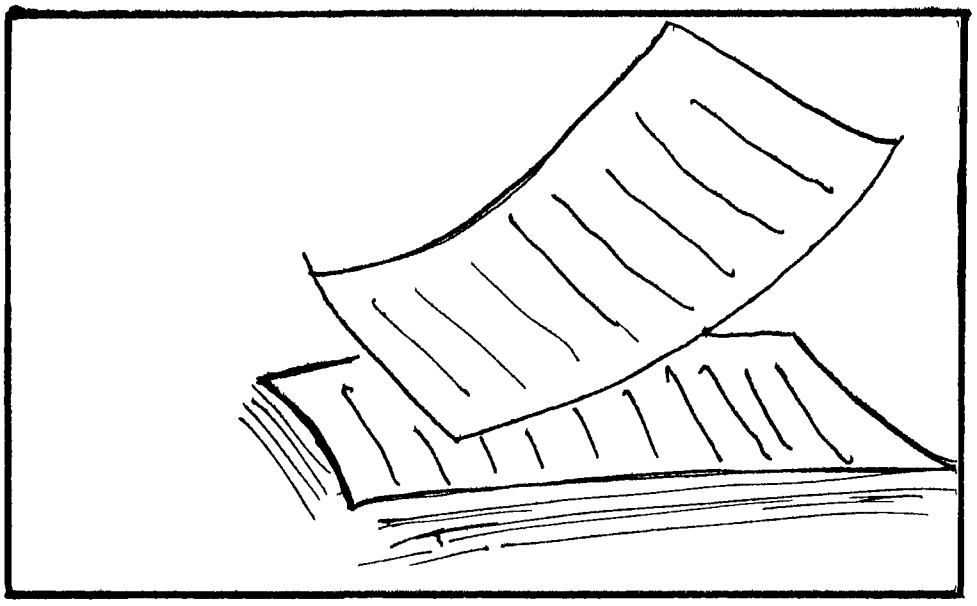


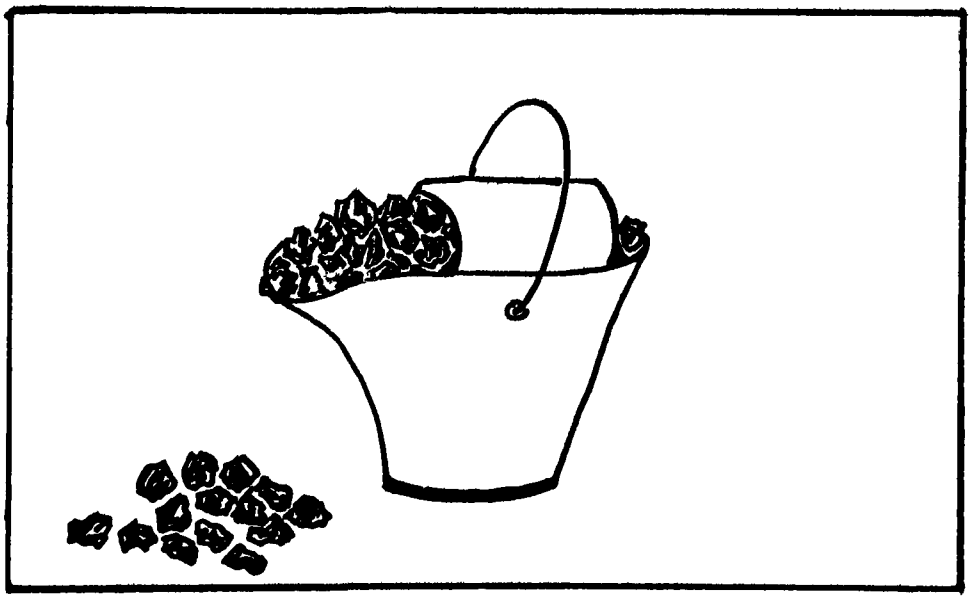
BEER — WINE



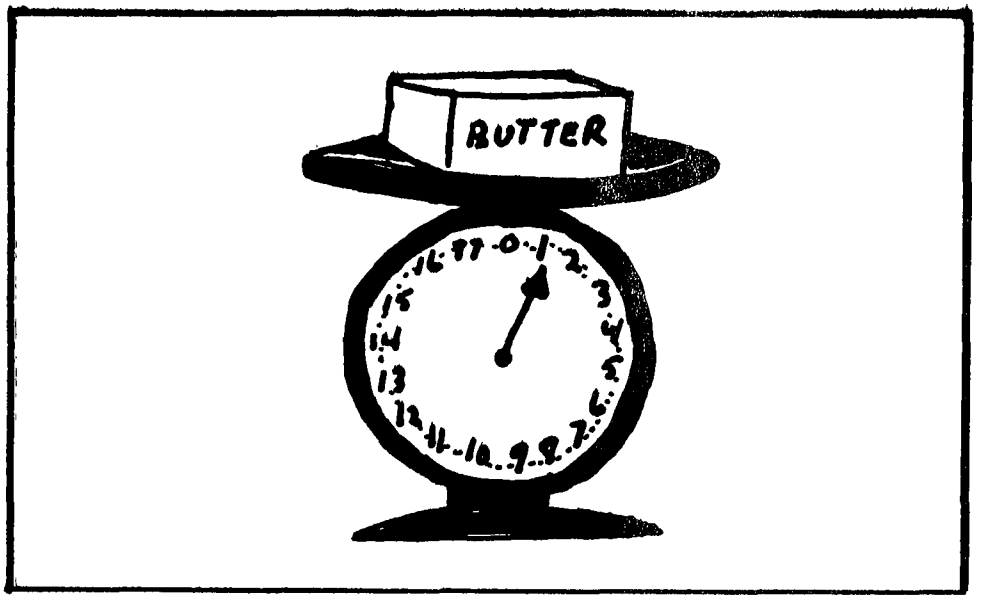


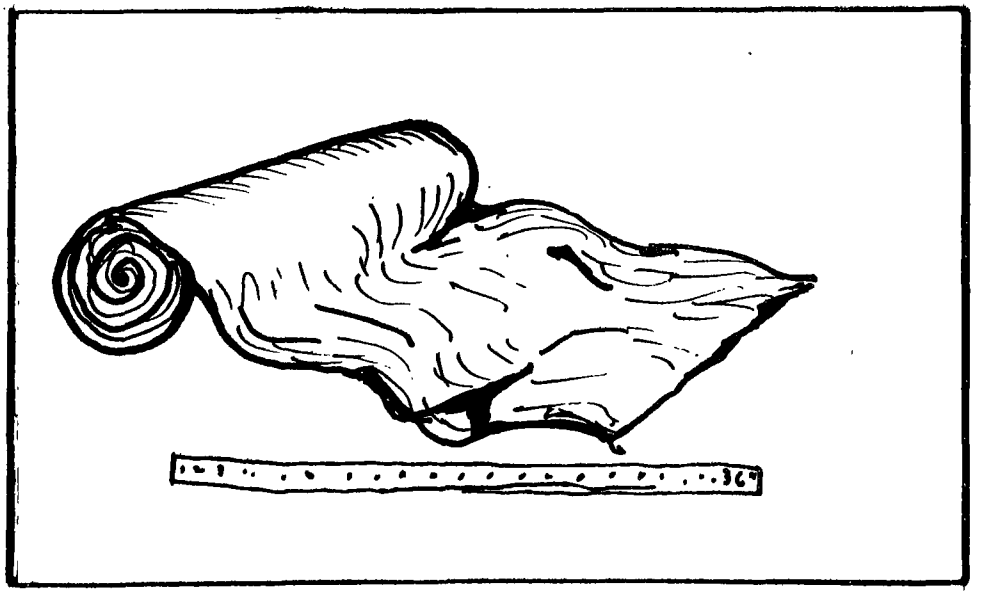
PIANO — VIOLIN



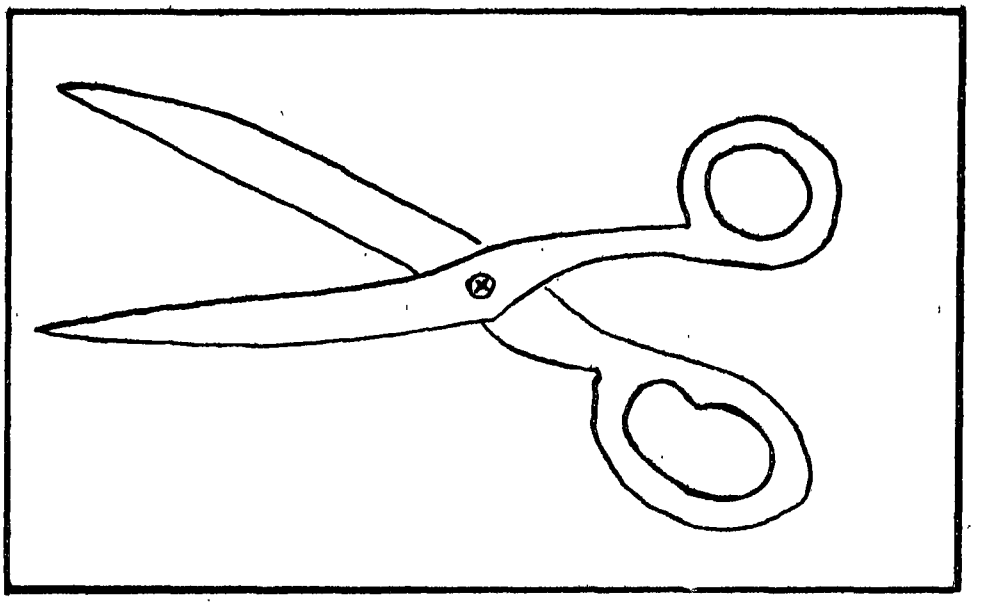


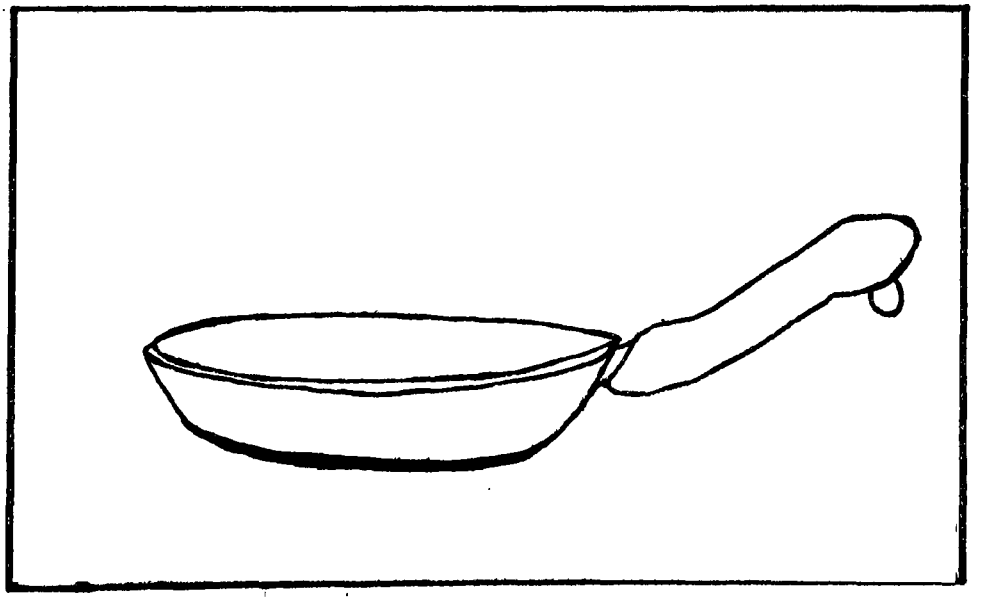
PAPER — COAL



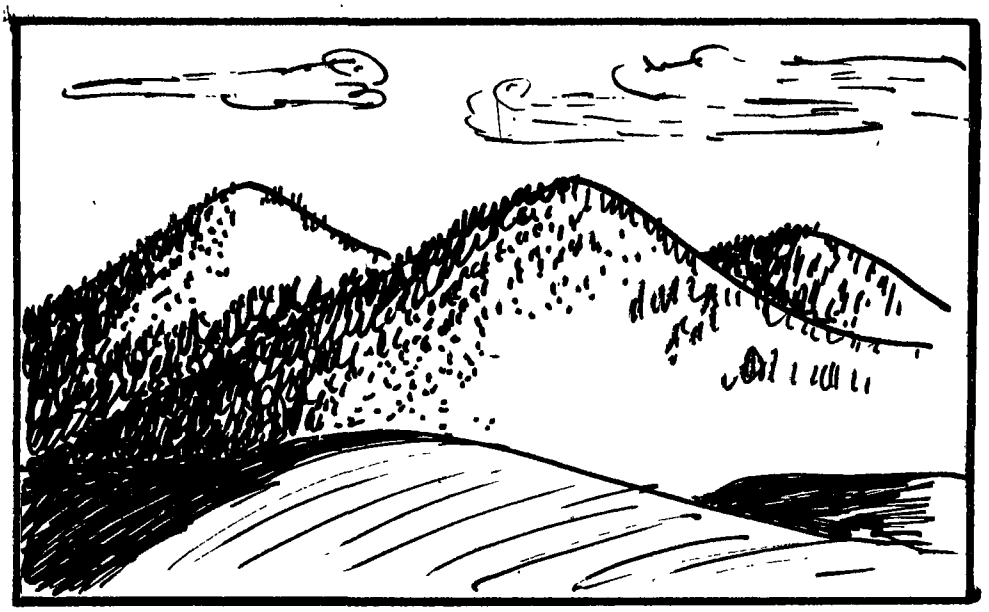


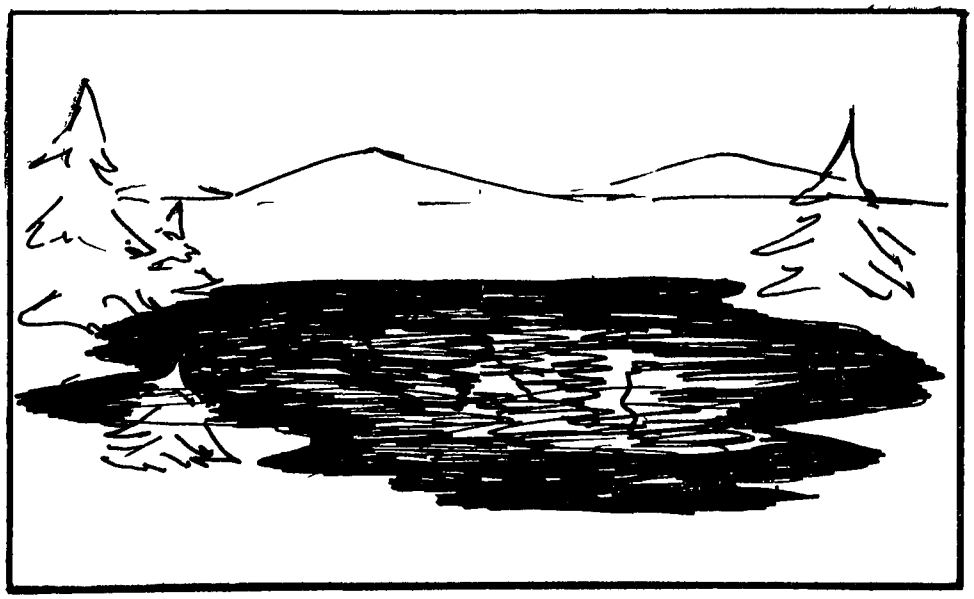
POUND — YARD



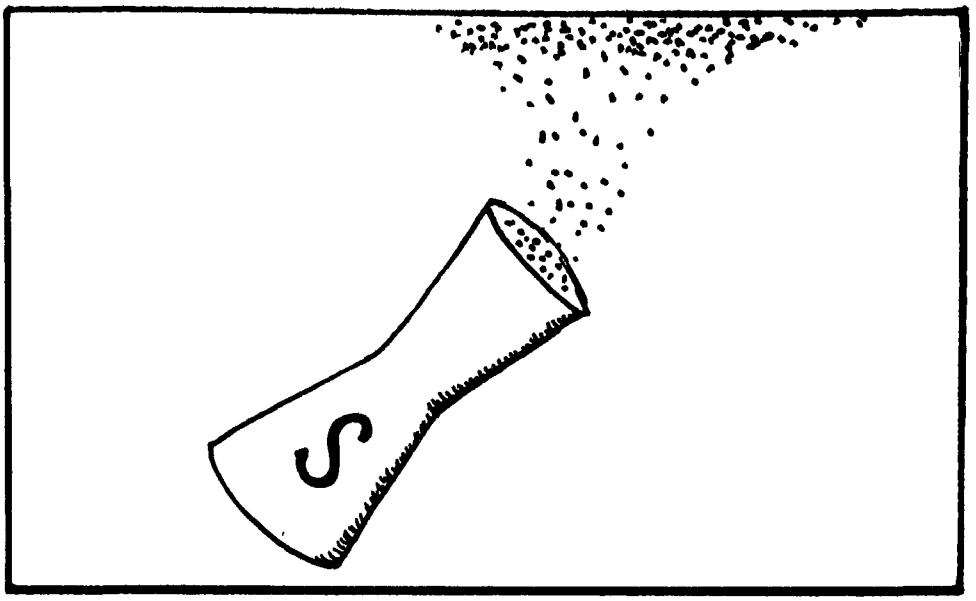


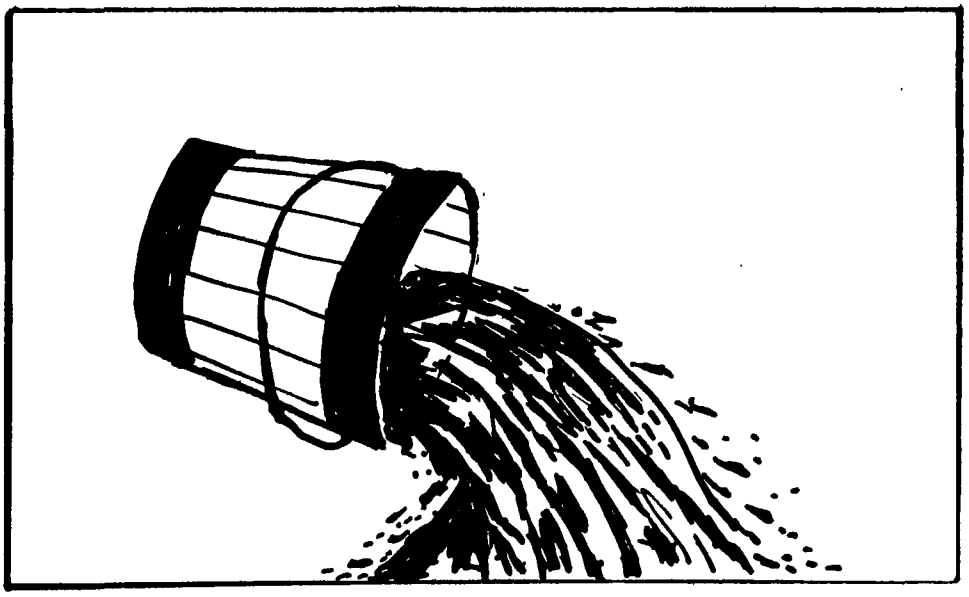
SCISSORS — COPPER PAN



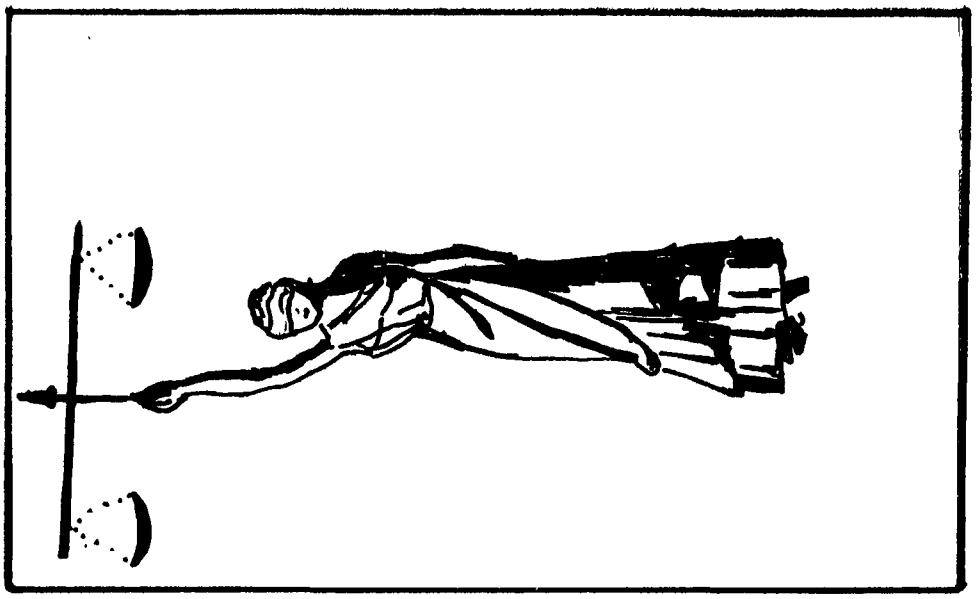


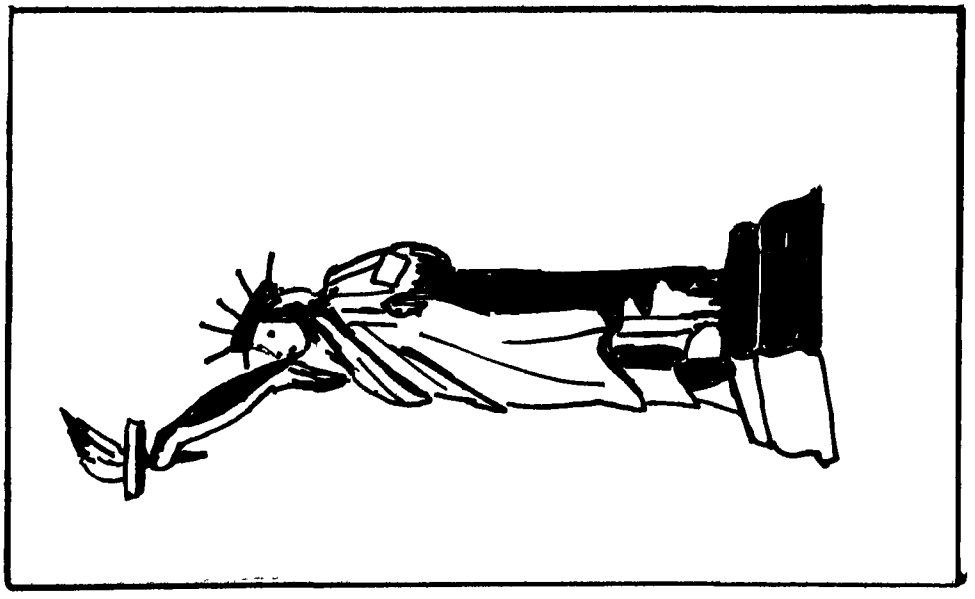
MOUNTAIN — LAKE



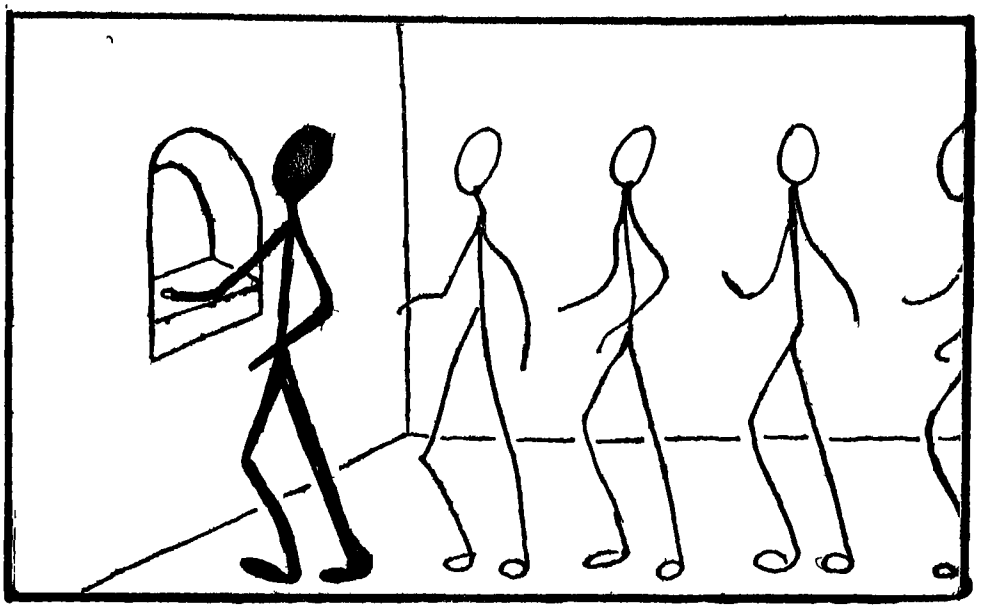


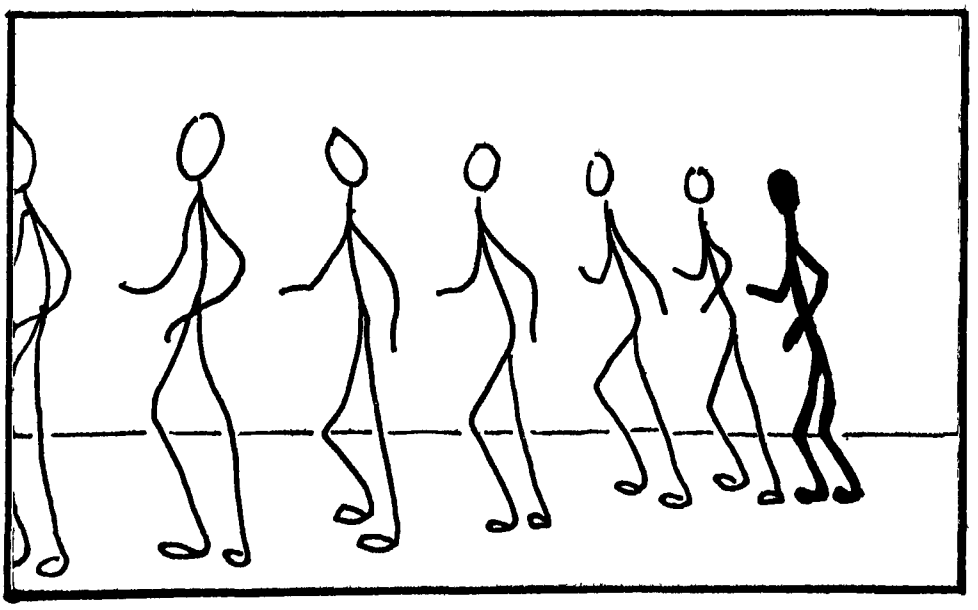
SALT — WATER



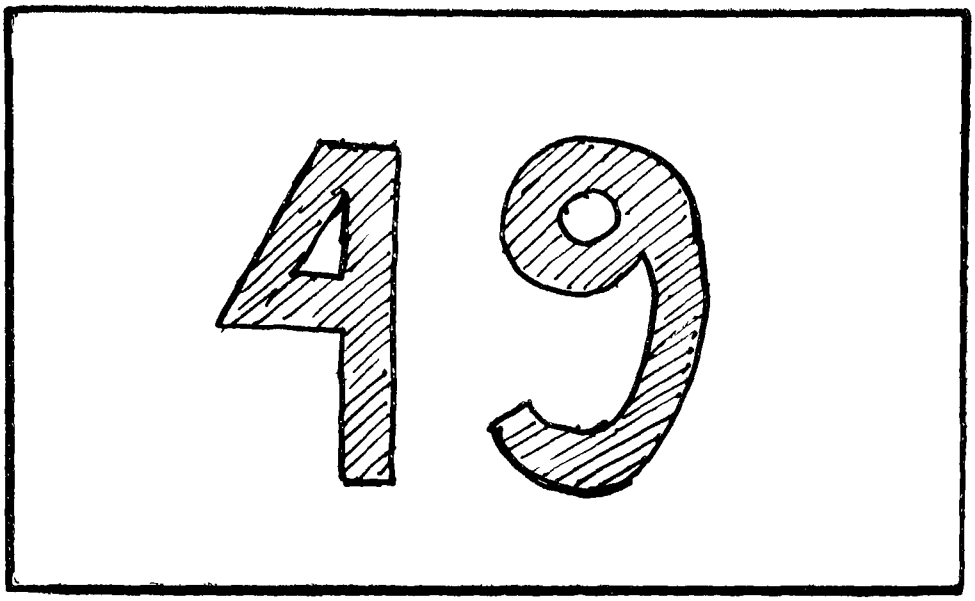


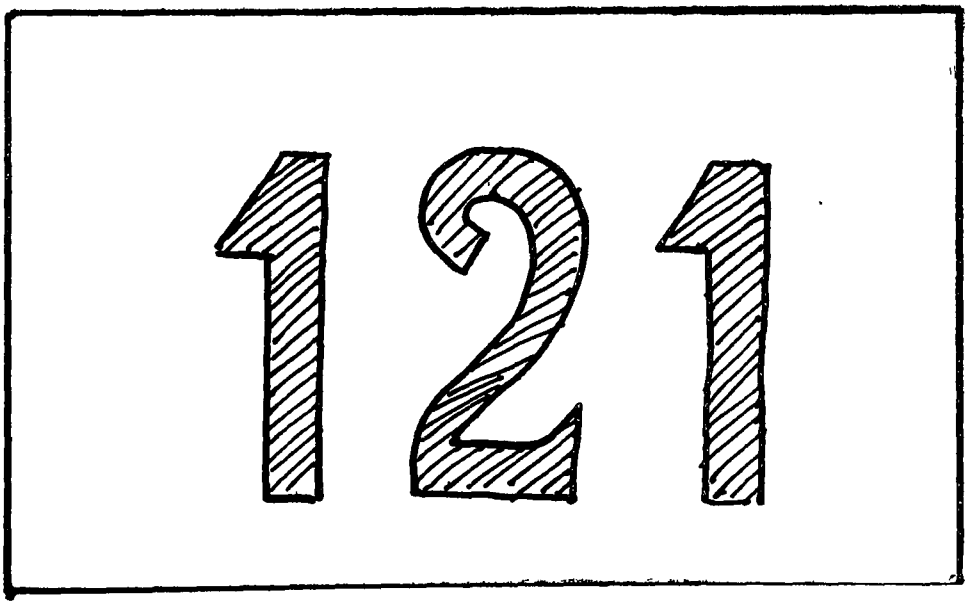
LIBERTY — JUSTICE





FIRST LAST





THE NUMBERS 49 AND 121

APPENDIX 2

A SAMPLE RECORD SHEET

NAME _____ EXAMINER _____

SCHOOL _____ ORAL - PICTORIAL: Yes No

GRADE _____ Year Month Day

SEX _____ DATE TESTED _____ _____ _____

AGE _____ DATE OF BIRTH _____ _____ _____

I.Q. _____ AGE _____ _____ _____

SIMILARITIES	SCORE
Plum - Peach	
Cat - Mouse	
Beer - Wine	
Piano - Violin	
Paper - Coal	
Pound - Yard	
Scissors - Copper Pan	
Mountain - Lake	
Salt - Water	
Liberty - Justice	
First - Last	

APPENDIX 3

ABSTRACT OF ORAL VERSUS ORAL-PICTORIAL PRESENTATION OF A SIMILARITIES TASK

ABSTRACT OF

ORAL VERSUS ORAL-PICTORIAL PRESENTATION OF A SIMILARITIES TASK[1]

The difficulty level of the similarities task of the WISC is such that, with young children, it fails to reflect individual differences in intelligence adequately. Lowering the difficulty level would raise the group mean, allowing more scope for variability in scores, which would thus be more parallel to the variability in intelligence.

To accomplish this, the nature of the stimulus task was varied by presenting a drawing of the items along with the usual oral question to one group. A comparable group was presented the oral question alone. The results upheld the hypothesis of no significant difference in the performance of the two groups.

As to the type of responses, few abstract-conceptual responses were given. Neither concrete nor functional responses predominated as a general trend. Further study of this question with different age groups was recommended.

Results suggest that the items are not in order of increasing difficulty. This could be investigated further.

---

1 Sheila D. Roberts, master's thesis presented to the School of Psychology and Education of the University of Ottawa, Ontario, May 1962, viii-59 p.