RESPONSE HIERARCHIES IN PRODUCTIVE THINKING

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

RICHARD PAUL STRATTON

1970

THESIS

This is to certify that the

thesis entitled

Response Hierarchies
in
Productive Thinking

presented by

Richard Paul Stratton

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

Donald M. Johnson

Date Aug. 15

O-169

ABSTRACT

RESPONSE HIERACHIES IN PRODUCTIVE THINKING

By

Richard Paul Stratton

In an effort to investigate processes involved in productive thinking, this experiment tested the response hierarchy theory of problem solving with a problem requiring subjects to write a series of sentences. This theory predicts that under free responding instructions, as more solutions are produced, the solutions will be less like initial solutions and will be judged to be of higher quality, i.e. more clever, etc. The quality dimension will be based on criteria which reflect prior learning. Since the sentence problem requires subjects to write many sentences which include four given words, there could be associative hierarchies involving pairs of the given words and word meanings. The theory would predict that strong adjective-noun associations and popular word meanings would appear early in a series of solutions. Remote associations and unusual meanings

would appear later and should be correlated with high quality. Under instructions which restrict responding by specifying criteria for high quality solutions, early solutions should be the same as later solutions in all respects, i.e. there should be no response hierarchy.

The present experiment used two naive judges to evaluate quality on a scale from one (low) to seven (high). Interjudge reliability was .86. With criteria-cued instructions and non-criteria-cued instructions regression analyses indicated that quality could be predicted by the number of word pairs (negatively correlated), unusual meanings, and sentence length. Quality increased over the response sequence for noncriteria-cued instructions when subjects wrote five or ten solutions, but not with criteria-cued instructions and ten solutions. Average quality was not significantly affected by quantity or criteria-cued instructions.

The experimental treatments involved subjects learning six sentences prior to producing solutions. These sentences included word pairs, unusual meanings, relevant words without pairs or unusual meanings, or irrelevant words. Results for subjects learning irrelevant sentences or none at all were identical. Learning word pairs increased the associative strength between the learned adjective-noun pairs such that (a) more word pairs were used, (b) learned pairs shifted upwards in the associative hierarchy such that they were given more frequently and earlier in the response sequence, (c) sentence length decreased, and (d) mean solution quality was lower. Learning sentences with relevant, but unpaired, words increased the use of word pairs, but relative position of pairs within the associative hierarchy did not change from that shown by control conditions. Also both conditions increased the associative strength between the words and their most common meanings, so that less unusual meanings were used. In both conditions quality increased as more remote associations were used.

Learning sentences with unusual meanings also increased the number of word pairs used, but the nouns acquired more new and unusual meanings. As more solutions were produced, unusual meanings became more unique, hence quality increased over the response sequence. Mean quality for this condition was higher than for other conditions, because unusual meanings were more highly weighed in judging quality than other variables.

These data were interpreted to be consistent with the response hierarchy theory and the general view that creativity involves a process of breaking up old associations and forming new ones, i.e. a divergent process. To the extent that externally or internally produced instructions provide cues to high quality solutions, the response hierarchy may be by-passed or may remain covert. Then no low quality solutions will be recorded, and no response hierarchy will be apparent in the solutions.

RESPONSE HIERARCHIES IN PRODUCTIVE THINKING

By

Richard Paul Stratton

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1970

GG-845
7-1-70

To Karin and Jason

ACKNOWLEDGMENTS

The good parts of this paper reflect the assistance of many people. To my wife, Karin, I owe a debt of gratitude for her consistent support and encouragement. Dr. Donald M. Johnson helped in many ways, so much so I am sure that he will never be aware of all that he has done. Drs. Gordon Wood, John Paul McKinney, and Charles Hanley gave much of that invaluable practical advice which is evident in every section of this paper. As a student I was privileged to be acquainted with many teachers who became models of what good teachers and researchers were like. I hope I may do the same for my students. As a son I was privileged to have parents who have helped in so many ways for so long. I hope that I can do that much for my son.

TABLE OF CONTENTS

Chapter		Page
	Quality and Quantity Instructions . .	52
	Effects of Learning Old Pairs	52
	Use of Word Pairs	54
	Other Response Measures	62
	Solution Quality	65
	Predicting Quality	66
	Word Pairs in Group NC10	68
	Effects of Learning Unusual Meanings. .	69
	Use of Unusual Meanings	70
	Solution Quality.	70
	Other Response Measures	76
	Analyses of Group TUM Solutions. . .	76
	Intersentence Associations	84
IV	DISCUSSION	90
	Divergent Processes Under Normal Conditions.	91
	Old Pairs.	95
	Unusual Meanings	105
	Instructional Effects.	111
V	SUMMARY	115
BIBLIOGRAPHY		118
APPENDICES		123

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

A "problem" exists as a problem, because the required response is initially not preeminent in the problem solver's response repertoire. Although this definition may be acceptable to most researchers in the field, the reason why the required response is not preeminent is a source of considerable disagreement.

Beyond this initial distinction problems may be classified as to the number of possible solutions. Many common problems, like Maier's two-string problem or an arithmetic problem, require a single solution which is either correct or incorrect by some obvious criterion. Other problems, like what title to use for a table of data, have several possible solutions, but only a few desirable solutions. Lastly, there are the problems which have an almost infinite number of possible solutions and a large number of desirable solutions. Writing a clever sentence which includes four specific words is an example of this type of problem. Furthermore, solutions to the latter may be easier to produce, but are certainly more difficult to score.

Guilford (1967) uses these differences between problems to separate convergent thinking (toward one solution) from divergent thinking (toward several solutions). Divergent problem solving has also been designated as productive thinking, orginality, or creative thinking by various researchers. The present study will investigate problem solving processes involved in solving divergent problems.

Problem solving may be viewed as being similar to learning. A pigeon in a Skinner box for the first time has many responses available to him. Learning is said to have occurred when the pecking response becomes dominant over grooming, fear and other irrelevant responses. When a problem is encountered by a human, the first solution omitted cannot be the best solution, or it would not be a problem. Other ineffective responses must be eliminated before the solution response can become preeminent. Perhaps it is this striking similarity between learning and problem solving that prompted Maltzman, Cofer and others to analyze problem solving processes in terms of learning theory. This introductory discussion will deal specifically with Maltzman's approach, and supportive evidence will be given first for recall problems, then for the more complex creativity problems.

Maltzman's Approach

Maltzman (1955) analyzed problem solving processes in terms of Hull's principles of learning. This extension focused upon Hull's concept of a spatial habit family hierarchy. In this type of hierarchy, groups (or families) of similar responses are arranged in hierarchical order in terms of the probability of occurrence, which is determined by habit strength, drive, etc. A spatial habit family hierarchy represents the hierarchical ordering of motor responses, and habit strength of motor responses varies as a function of spatial location relative to the goal.

In human problem solving, however, responses to changing spatial relations to a goal are rarely elicited. More commonly responses are verbal, and response changes are related to temporal proximity to the solution. Maltzman represented human thought as a compound temporal habit family hierarchy and assumed that principles operating in the spatial habit family hierarchy would also operate in temporally-based hierarchies as well.

As in simple trial-and-error learning, a problem stimulus is capable of eliciting a hierarchy of responses, and the correct response is initially low in the hierarchy superseded by incorrect responses of greater habit strength. According to Maltzman, the order of dominance in the response hierarchy may be changed as a result of three processes. (a) Dominant responses may be elicited, and the effective reaction potential of these may be reduced as the result of extinction. (b) The probability of occurrence of a habit family containing correct responses could increase due to reinforcement of individual members of that habit family. (c) The arousal of an anticipatory goal response would produce an increment in effective reaction potential for related responses. The antecedent condition for the anticipatory goal response would be commerce with the goal or a goal substitute, such as instructions which describe the goal.

The application of a response hierarchy interpretation to convergent problems with one correct solution is straightforward. Divergent problems would be expected to follow similar processes with "good" solutions behaving as correct solutions would in convergent problems. The unusual uses problem will serve as an example. In this problem _S_s are required to write many uses of a common object, such as a brick. The most desirable solutions are usually defined by the experimentor as solutions which are infrequent in a large sample of solutions obtained from many _S_s. For the uses of a brick an uncommon use would be to crush it and use it as a filter for a moonshine still.

A common use would be to use a brick for building a house. According to Maltzman common uses of an object are omitted and reinforced more frequently in everyday life, and subsequently common uses are more likely to be given for similar laboratory problems. Thus, it may be said that the many uses of an object form a response hierarchy based on frequency of occurrence and concomitant reinforcement. Such a response hierarchy would be evidenced when an individual produces a series of solutions and the common uses precede the uncommon ones.

Maltzman has been very careful to distinguish between orginality, where his theory was meant to apply, and creativity. The latter was said to conjure up the unfavorable connotations of individual differences, social influences, and the social value of solutions which cannot be controlled in the laboratory setting. Mednick (1962), on the other hand, assumes associative hierarchies, as does Maltzman, but uses them to account for individual differences in creativity. Although the present approach emphasizes solution processes and solution characteristics, it clearly does not deemphasize individual differences.

Response Hierarchies in Recall Problems

For the purposes of organization, divergent problems will be classified as to the degree to which problem solution depends upon simple recall. Verbal association problems may be seen as one end of the continuum where responses are stored in memory, and problem solution depends upon memory scanning. Initial solutions to the unusual uses problem are simply recalled, but, after these common solutions are exhausted, uncommon solutions must be constructed from what the problem solver knows about the object, its uses, etc. At the extreme are problems like the sentences problem where Ss are only given four words and must construct sentences which include all four. In this case every solution must be constructed from scratch, since the problem solver has never previously encountered the problem or its solutions. In accordance with other researchers, recall problems in this discussion will include verbal association, anagram and unusual uses problems. Creativity problems will include plot title, sentences and similar problems.

Initial evidence which may be cited in favor of Maltzman's approach was presented by Christensen, Guilford and Wilson (1957). In this study subjects were required to produce several solutions to a variety of problems. Unusual uses of a button and pencil were requested, and the number association problem consisted of writing down things commonly associated with the given numbers. An example of the latter would be "four-square." The frequency of occurrence of each solution was tallied, and each solution was given a value for uncommonness based on this frequency. It was found that, as Ss proceeded through the response sequence, the solutions became more uncommon.

Duncan has published several studies which present evidence for verbal response hierarchies based on association strength and frequency of occurrence in the language. Duncan (1966a) presented Ss with a stimulus word and a list of five associates to the stimulus. The task was to guess which associate was "correct." The order of their guesses proceeded from the strongest to the weakest associates. When Ss were told to guess a word with a particular association level to the stimulus, Ss were very accurate on the first guess and got more accurate with subsequent guesses (Duncan, 1967). In the unusual uses problem Manske and Davis (1968) and Turner (1967) note that responses were omitted in associative clusters (e.g., uses of a tire were "use as a wheel for a car, bike, boat trailer, etc."). They suggested that the order of association went from strongest to weakest.

Frequency of occurrence of a word in the language is also a powerful variable in other verbal problems. In Duncan's (1966b) study Ss were presented with the first and last letter of a five-letter word, and the task was to guess what word was intended. The several solutions for each problem varied in frequency, and the number of solutions omitted was greater for words of higher frequency. When the number of possible solutions was restricted to one high and one low frequency word, first guesses contained more high frequency solutions. In another experiment the task involved finding a word of a given length, beginning with a given letter and being in a given class, for example "trees." With only two possible solutions, when the high frequency solution was given as an example, more low frequency solutions were given. When the low frequency word was given as an example, there was no increase in the number of high frequency solutions.

In anagram problems frequency is one of the most powerful variables determining success and solution time (Johnson, 1966). Anagram solution time is inversely related to the frequency of the solution word when there is only one possible (Mayzner and Tresseit, 1958; Dominowski, 1965 and 1967). With more than one possible solution the order of emission of solutions will correspond to the frequency of the solution word (Johnson and Van Mondfrans, 1965). When the solutions vary in frequency and Ss are required to find both solutions, the high frequency solution is given first (Mayzner and

Tresseit, 1966). In these studies past experience in the form of frequency of usage determined the position of the responses in the verbal hierarchy.

The response hierarchy concept has been used by Geriach, Shutz, Baker and Mazer (1964) to explain the results of some brainstorming research. They used the unusual uses problem and hypothesized that a response hierarchy would exist with these problems such that Ss would give the most common response first and subsequently work down the hierarchy to less common responses. Their results showed that more "good" responses (based on judged uniqueness and value) were produced in the last two-thirds of the production sequence. Importantly, when Ss were informed of the criteria by which their solutions were to be evaluated (criteria-cued instructions), such a production order did not exist. Using the same type of problems Parnes (1961) and Manske and Davis (1968) have also found that quality was related to position in the response sequence without criteria-cued instructions, and with criteria-cued instructions that relationship was weaker if it existed at all.

The preceeding studies using the unusual uses problem and criteria-cued instructions found that these special instructions increased mean quality and disrupted the response hierarchy such that quality was unrelated to position within the response sequence. These results would be explained within the Maltzman (1955, 1962) framework as being due to instruction-induced anticipatory goal responses. The brick uses problem will serve as an example. When a long series of uses is elicited, the uses usually are recorded in associative clusters (Manske and Davis, 1968, Turner, 1967). Members of a cluster would have some property in common, such as a "weight use." Now let us say that the task instructions in addition to presenting the problem do not explicitly state, but implies, the desirability of weight uses. These instructions would be the external stimulus, and the internal representation (or the derived implication of "weight use") would be the anticipatory goal response which in turn would key off a series of weight use responses. Maltzman specifically discusses instructions as one source of the anticipatory goal response. In essence, the anticipatory goal response resembles a description of the desired solution and may be derived from task instructions or self-instruction. The result is that only a specific type of response is recorded, and the usual progression from common to uncommon responses is eliminated.

Response Hierarchies in Creativity Problems

The results of several experiments suggest that creativity problems may have response hierarchies associated with them. Christensen, _et al_. used the consequences problem and the plot title problem in addition to the ones already mentioned. The consequences problem asks such questions as "what would be the consequences if everyone in the world suddenly went blind?" Solutions were scored by judges on a three-point scale for the remoteness of the consequence from the initial situation. For example, "there would be a lot of stumbling" is less remote than "the congenitally blind would become world leaders." Solutions in the second half of the production sequence were significantly more remote than those in the first half. They also used the plot title problem. The plot titles were rated by judges for cleverness and appropriateness, but no improvement was found over the production sequence. Importantly Plot Title instructions asked for "clever" or "appropriate" solutions. With criteria-cued instructions the response hierarchy would be disrupted to the extent that one would not expect the normal ordering of responses. Instead _S_s were just as likely to produce good solutions in the first half of the production sequence as in the last half.

Johnson, Parrott and Stratton (1968) worked with the plot titles problem and four similar problems. The solutions were judged for "quality" based on criteria which varied with the problem. (a) They found that solutions above the 90th percentile in quality were more likely to occur in the last half of the production sequence for the plot title, sentence and cartoon caption problems. The ordering was found, however, only if instructions did not contain reference to evaluation (i.e. were not criteria-cued instructions). (b) An individual problem solver produced six to seven solutions, on the average, but the solutions varied widely in quality (the median range of quality was seven points out of 13). Thus it would have been possible for each individual to have produced some poor solutions before producing some good ones. (c) Finally, *S*s writing one solution had a higher mean quality than *S*s writing several solutions. Since all *S*s had the same time to write solutions, single-solutions *S*s could have covertly produced and rejected several inferior solutions while multiple-solution *S*s recorded every solution. By producing and rejecting some solutions, single-solution *S*s could have advanced further down the response hierarchy to better solutions. They would also have had more time to produce solutions, since they did not write down unacceptable solutions. These results, although based on judged quality, are similar to those obtained with statistical frequency (unusual uses problem) or judged remoteness (consequences problem).

Johnson (1968) reports a study, designed by Stratton, using four different problems where production time was limited. There was a significant increase in quality for _S_s writing one solution in two or four minutes over _S_s writing one solution in a half-minute or one minute. With criteria-cued instructions a half of a minute was as good as four minutes in producing one solution. Response hierarchies may be invoked as one explanation for these data. Without criteria-cued instruction one way to produce good solutions would be to advance down the response hierarchy by producing and rejecting several solutions. Limiting time then would limit the consideration of alternative solutions. Criteria-cued instructions would have made more time unnecessary for the construction of good solutions by providing cues to the nature of the desired solution.

Theories of Creativity

Mednick and Guilford present theories about creativity which stress individual differences, but which make assumptions about solution processes in creativity problems. Most theories of creativity assume the solution process to be a divergent one, so the ones discussed here will be more exemplary than exhaustive. The divergent nature of the creative process is quite similar to the notions of British associationists from Locke (1690) to Bain (1855). Psychologists whose work reflects the speculations of the associationists also argue for divergent processes in creativity. Freud (1938), Hollingworth (1928) and Binet (1899) will serve as examples. Golann (1963) presents an excellent summary of theories of creativity, and he concludes that divergent mechanisms are included in most theories.

Mednick (1962) stresses the associative nature of creativity as a basis for a theory about individual differences. He defines creative thinking as "the forming of associative elements into new combinations which either meet specified requirements or are in some way useful. The more mutually remote the elements of the new combination, the more creative the process of solution (p. 220)." Based on this assumption Mednick proposes to separate the creative from the noncreative person. When a creative person free associates to a stimulus word or produces a series of solutions to a problem, the associative strength between the stimulus and each successive response would be approximately equal. The associative strengths for a noncreative person would be very high for the first few responses and very low for the remainder of the responses. As a result the noncreative person does not produce as many uncommon solutions. Were one to plot associative strength as a function of position in the response sequence, a creative person would show a flat associative hierarchy, a noncreative person would show a steep associative hierarchy.

Mednick's definition is a reapplication of the old maxim that great discoveries are just old ideas combined in new ways. Relating it to the unusual uses problem one can see its possible validity. Let us assume that our intuitive notion of the associative strength between the word "tire" and its uses will suffice for empirical data. The use of a tire as "to put on the wheel of an automobile" may be said to have high associative strength. The use of a tire as "to burn for keeping fruit trees warm during the spring frosts" has a low associative strength. Thus, a creative person would be able to produce many solutions of all associative strengths, or perhaps to produce all solutions of low associative strengths. The noncreative person, on the other hand, would produce common uses with high associative strengths and few uncommon ones. Furthermore, Mednick assumes that an individual's organization of associations will influence the probability and speed of attainment of a creative solution. Creative people, then, are assumed to store experiences differently from noncreative people.

Guilford's (1967) approach is based on his familiar structure-of-intellect model. The plot title, remote associates, consequences, and unusual uses problems are included in the cell called "Divergent Semantic Transformations" (DMT). Previous researchers had called this the orginality factor. The assumptions underlying this factor are illustrated by the fact that tests loading on this factor emphasize either (a) ability to produce responses that are statistically rare in the population, (b) ability to produce remotely related responses, or (c) ability to produce clever responses. Verbal fluency, foresight, reasoning and other problem solving abilities are included in other factors. In the DMT factor the emphasis is on production of semantic units which differ from the original problem by being uncommon, remotely associated or clever responses (i.e. are transformations of stored information). His model for problem solving and creative production emphasizes the recall and transformation of information to create new informational products.

Response Hierarchies in the Sentence Problem

The above results and theories suggest that creativity problems may also have response hierarchies associated with them. To examine this question further it would be necessary to establish two things: (a) that responses are ordered along some dimension when there are no criteria-cued instructions, and (b) that this dimension is responsive to learning. Christensen, _et al_. found a progression of solutions to the consequences problem which went from less to more remote between halves of the response sequence. Johnson, _et al_. obtained the same results with Plot Titles, Sentences, and Cartoon Captions.

In the problems used by Johnson, _et al_. the dimension upon which the resposes were ordered was "quality," as defined by two judges. Unlike the "commonness" of Unusual Uses solutions, "quality" is more difficult to define, especially since the criteria on which quality is rated varies between problems. In the Johnson, _et al_. study the judges rated the quality of about a thousand solutions to a problem, then they constructed a Rating Guide, which expressed the characteristics of the solutions at each quality level. If learning does determine the order of solution production, it may be illustrated by the characteristics on which quality was evaluated.

The sentence problem will serve as an example. The task for this problem is to write many sentences, each of which includes the words _happy_, _expensive_, _horse_, and _lake_. The basic criteria for evaluation were that sentences include all four words and be grammatically well-constructed, and that the words appear unobtrusive in the sentence. The solutions were rated by each judge on a scale from one (bad) to seven (best).

Table 1 summarizes the criteria by which solutions to the sentence problem were evaluated. There are several criteria which may reflect learning. (a) Grammatical structure is learned prior to the experiment. On the other hand, the length and complexity of sentences could be increased if _S_s learned from one sentence to another, i.e., built each sentence upon elements of preceeding sentences. From the Rating Guide it is apparent that better sentences are those which are longer, more complex and grammatically well-constructed.

(b) The adjectives and nouns have varying degrees of associability which may be attributed to usage prior to the experiment. When two of the given words appear together, they are called an "old pair." For example, the old pair _happy_ _horse_ appears more frequently and makes more sense than _happy_ _lake_. Solutions which break up these pairs are given better quality ratings. Maier (1967) introduced this concept to measure the reorganization of experience in creative problem solving.

(c) The meanings of the given words are well established prior to the experiment. Each word normally has several meanings, but some are more common than others. A horse is initially "a four-legged mammal with

TABLE 1.--Rating Guide for Sentence Problem

For a rating of	The sentence must have these characteristics
1	does not use all four words.
2	lists the words.
3-5	a clumsy, mediocre sentence.
4-7	well-constructed sentence with words having novel meaning or usage (e.g., Horse Lake).
6-7	well-constructed sentences which are clever or humorous.

Additional considerations: any rating may be increased one point

(a) if the sentence is complex, and

(b) if happy and expensive modify nouns other than horse or lake.

N. B. "4" is an average rating.

solid hoofs and a flowing mane and tail," but it is also "slang for heroine, a sawhorse, gymnastics vaulting horse," or the verb "to horse around." These unusual meanings receive consistently higher ratings. Osgood (1953) proposed that meanings form associative hierarchies with their words such that some meanings are more highly associated with a word and would tend to be the first meaning recalled and used to solve a problem.

(d) Finally, good sentences incorporated the given words into their context so well that the judge had to look twice to make sure all four words were used. The obtrusiveness of the given words is reflected in the subject and direct object of the sentence, and better solutions use words other than the given ones for the subject and direct object clauses. This reflects learning in that, the more familiar a word, the easier it is to have it function in any part of the sentence, and _S_s become more familiar with a word as they use it in more sentences. This variable is called "topic freedom." Yngve (1960) described a similar measure to relate word depth to syntactic structure.

A Pilot Study

To illustrate the potency of these variables a pilot study will be briefly discussed (see also Appendix A). Before the Rating Guide for the sentence problem was constructed about a thousand solutions had been evaluated by two judges with interjudge correlations above. 80. The 215 complete solutions produced under standard instructions were analyzed in terms of 16 variables including the above. A multiple regression analysis showed that 50% of the variance in the judges' ratings could be accounted for by sentence length, sentence complexity, number of word pairs, unusual meanings, and topic freedom. These variables correlated with quality between .39 and .46 (word pairs correlate with quality -.38). The judges could see these features of the sentences when rating the sentence. If a response hierarchy in the sentence is based on quality and if it is to be controlled by learning, it would be good to look at the fate of these influential variables over the production sequence.

Summary

The foregoing discussion suggests that better solutions to creativity problems are uncommon in the sense that they contain word uses, meanings and associations which do not occur frequently in the normal sample of solutions. In other words, better solutions appear to diverge from the usual meanings, etc. normally associated with the words in this problem. If quality is related to these aspects of the solution, and if quality is to increase over the response sequence, initial solutions should be ones which most closely correspond to the normal properties associated with the problem elements. In the sentence problem the elements are the words. Later solutions should use less common meanings, etc. of the words and consequently receive higher quality ratings. The same progression of meanings and associations are assumed to occur if _S_s were to simply produce all possible meanings and associations. The first meanings produced, then, would also be the first to be used in solving a problem. The progression of solution quality would represent a response hierarchy in that there would be an orderly progression of solution quality which is determined by learning prior to the experiment or during an experimental treatment.

The Present Study

The major purpose of this study was to investigate the existence of response hierarchies in creativity problems. Two questions were investigated: (a) Does quality increase over the response sequence? (b) What conditions in the instructions and in the nature of the words affect this increase?

The sentence problem was used with the words _happy_, _expensive_, _horse_, and _lake_. Using a variety of words would increase the generality of the results, but it would be difficult to equate the quality ratings for sentences using different sets of words. Thus, two treatments were introduced to change the meaning and associative characteristics of the nouns horse and lake. These treatments also reflect the importance of learning in establishing response hierarchies in this type of problem.

The dependent variables were quality, sentence length and complexity, number of old pairs (or word pairs), number of unusual meanings, and topic freedom. These variables were defined in the previous discussion. The critical feature of the analyses was the variation of these solution characteristics over the response sequence.

Two groups of Ss were used to observe changes in the above variables under "normal," or unrestricted, production conditions. Group NC10 wrote ten solutions to the sentence problem and Group NC5 wrote five solutions. It was expected that quality would increase over the response sequence, and that changes in other highly correlated variables would indicate why the change in quality occurred.

Two other groups were used to investigate the effects of criteria-cued instructions. Group CC10 wrote ten solutions to the sentence problem, and Group

CC5 wrote five. The additional instructions should provide cues about desired solutions such that mean quality would be higher and there should be no increase in quality over the response sequence. This would agree with previous research. Other dependent variables would also show the solution characteristics which accounted for the increase in quality.

There were two treatments designed to illustrate the effects of the nature of the words on the observed changes in quality and thus to show the importance of learning in determining the order of solution production. The first treatment was designed to increase the associative strength between the words happy horse and expensive lake. The treatment for Group TOP consisted of learning six sentences which included both of these old pairs prior to reading the problem and writing ten solutions.

It was expected that the total number of old pairs used in solutions would be greater in this group than other groups. Furthermore, more old pairs should be used initially than later in production. Because quality is reduced by the use of old pairs, later solutions should increase in quality.

The second treatment was designed to alter the hierarchy of meanings associated with horse and lake by increasing the associative strength of unusual meanings.

Group TUM Ss learned six sentences which included novel meanings before reading the problem and writing ten solutions. It was expected that the total number of unusual meanings used in solutions would be greater in this group than in comparison groups. Because unusual meanings are highly weighted in quality ratings, mean quality should also be greater in this group. Furthermore, it was expected that other Ss would have to discover unusual meanings after the use of common meanings, but Group TUM Ss would start out using the same unusual meanings. Thus, Group TUM Ss should show no increase in unusual meanings over the production sequence and consequently no increase in quality.

There were three comparison conditions for the two learning treatments. The standard comparison was Group NC10 where ten solutions were produced with standard instructions. There were two controls for the learning treatment. To control for learning sentences which used the same words as the problem, Group TR Ss learned sentences which included the four given words but no old pairs or unusual meanings. Group TIR was a control for learning any sentences, and these Ss learned sentences with four other words. It was expeced that the comparison groups would not differ from one another.

To summarize, this research may be conceptualized as occurring in three phases: (a) A comparison of

instructional conditions of five or ten solutions produced with or without criteria-cued instructions.
(b) An attempt to change solution quality by increasing the associability of specific adjective-noun pairs.
(c) An attempt to change solution quality by increasing the associative strength of uncommon meanings of the nouns to be used.

The major hypotheses were: (a) Under instructions with no criteria-cues, sentences will increase in quality over the response sequence whether five or ten solutions are requested. There should be no such increase with criteria-cued instructions. (b) Learning old pairs will increase the use of old pairs in solutions, and the use of old pairs will decrease over the response sequence with a concomitant increase in quality. (c) Learning unusual meanings for nouns will increase the use of the unusual meanings in solutions to the sentence problem, but the use of unusual meanings will not increase over the response sequence.

CHAPTER II

METHOD

Subjects

In each of the eight conditions there were 30 volunteer Ss. Males and females were roughly equated across conditions. There were 31 additional Ss who did not complete the task or were randomly discarded to equate the groups.

Materials

The sentence problem included the words happy, expensive, horse, and lake. Ss were required to produce five or ten sentences, each of which included all four words.

The standard instructions read: "Your task in this experiment will be to write ten (or five) sentences. Each sentence must include the words happy, expensive, horse and lake. You will have as much time as you need."

The criteria-cued instructions read: "Your task in this experiment will be to write ten (or five) good sentences. Each sentence must include the words happy, expensive, horse, and lake. Furthermore, a good sentence is grammatically well-constructed and reads smoothly. The four given words fit smoothly into the sentence structure of a good sentence. You will have as much time as you need to write ten (or five) good sentences."

The sentences used in the memory tasks were selected from previous experiments and were equivalent in judged quality and length between groups. The mean quality of each set of sentences was about average. The irrelevant sentences contained the words big, alone, noise and money, while relevant sentences included the words to be used later in the sentence problem. Three types of relevant sentences were used; without old pairs or unusual meanings, with old pairs, and with unusual meanings. Verbal instructions for memory task Ss posed the situation as consisting of two short unrelated experiments with the first being a memory experiment. In addition to the standard instructions for the sentence problems, Ss with memory tasks were instructed "Do not simply copy the sentences you have just learned. Think of your own." Memory task instructions and sentences for each group are included in Appendix B.

The memory task and the problem were assembled in booklet form. A cover sheet required S to record his name, student number, etc. The memory task followed if applicable. The next page was for recording the memorized sentences. The following page contained the problem and sufficient space for recording the required number of solutions.

Procedure

Three experimenters randomly assigned Ss to treatments such that each treatment was represented in almost every session. After filling out the cover sheets all groups were timed in subsequent activities. The memory task took ten minutes, and Ss were given 15 minutes to record ten solutions or ten minutes to record five solutions. Ss were informed of the time at various stages of production.

Group NC10 wrote ten solutions to the sentence problem and were given no criteria-cued instructions.

Group CC10 was the same as Group NC10 but had criteria-cued instructions.

Group TOP memorized six relevant solutions which included the pairs happy horse and expensive lake. Then, they read the problem and wrote ten solutions under Group NC10 instructions.

Group TUM memorized six relevant sentences which included several unusual meanings for horse and lake. Then, they wrote ten sentences under Group NC10 instructions.

Groups TR and TIR were control groups. Group TR Ss memorized relevant sentences which did not have any old pairs or unusual meanings. Group TIR Ss memorized irrelevant sentences. These groups controlled for familiarity with the problem or solutions. Both groups then read the problem and produced ten sentences under Group NC10 instructions.

Group NC5 was comparable to Group NC10, except that Ss were requested to write only five solutions.

Group CC5 wrote only five solutions and had criteria-cued instructions. This group is comparable to Group CC10.

CHAPTER III

RESULTS

Solution Scoring[1]

The obtained sentences were typed on IBM cards and coded on the reverse side. Coding included the group and subject numbers and the solution position. After shuffling the solutions, two judges who were naive as to the purpose of the experiment rated solution quality from 1 (bad) to 7 (good). The data for this experiment are the sum of these ratings. A Rating Guide established in previous experiments assisted in the rating (see Table 1). Spurious agreement was prevented by the first judge recording his rating on the back of the card, and the second judge recording his rating first on the front of the card, then on the back. The correlation between the two judges' ratings for the 2100 solutions in this experiment was .86, which is acceptable.

Prior to rating solutions for the present experiment the judges trained on a judgment training

[1]The assistance of John Jerome, Bill Gould, Jerald Wilbur and Karin Stratton in this stage of the analysis is gratefully acknowledged.

program, which was developed and validated in another experiment (Johnson, *et al.*, 1968). The training essentially presented the Rating Guide and practice in discriminating good and poor solutions with examples and feedback. It was designed to enable judges to rate solutions on the same judgment scale as the judges who developed the Rating Guide and on whose judgment styles this experiment is based. After training, the judges practiced with the seven-point scale until they established a high interjudge agreement.

After all solutions were scored for quality, each solution was scored on the following variables by different judges who were also naive: (a) *Length*--the total number of words in the sentence. (b) *Complexity*--a three-point rating based on the grammatical complexity of the sentence. (c) *Old Pairs*--the number of times *happy horse* and *expensive lake* occur in the sentence. *Total pairs* or *word pairs* refers to all combinations of the given words. (d) *Unusual Meanings*--meanings or uses of the given words which differ from the ordinary, e.g. "Happy Lake Dude Ranch" or "to horse around." (e) *Topic Freedom*--the number of times the given nouns are used as other than the subject or direct object of the sentence.

The identification codes and scores were then punched onto IBM cards for computer processing.

With respect to the assumptions underlying the analysis of variance tests to be used, it should be noted that nowhere have the assumptions been violated to the extent that the conclusions should be suspect (Boneau, 1960). With 30 Ss per group, n = 300 for comparisons between groups with ten solutions per S, and n = 150 for comparisons between groups with five solutions per S. Variances for these comparisons were inspected and there were no differences as large as 3:1. The distributions are without marked skewness, except for complexity and unusual meanings. Complexity has a skewness of about -2.0 (zero for a normal distribution) in each group. This results from a ceiling effect of the three-point scale. Unusual meanings are generally positively skewed at a value of about 2.0. This is due to the fact that most solutions have no unusual meanings in all but Group TUM. Boneau states that with large samples and equivalent variances such deviations from normality should have little effect on obtained probability values, even when a J-shaped distribution is tested against a rectangular distribution.

Standard Instructions

This section is concerned with two questions; "Does quality increase over the response sequence under standard instructions?" and "How are changes in quality mediated by response variables which reflect learning?" This section will also present data on the effects of quantity instructions on the solution process.

Quality and Position

Figure 1 presents the mean quality for each position in the response sequence for Group NC10 with ten solutions per S and Group NC5 with five solutions per S. Both groups show a general increase in quality, but it is more pronounced for Group NC5. Single-factor analyses of variance for repeated measures across positions tested the statistical significance of changes in quality. Quality increases over the response sequence in Group NC10 ($F = 2.76$; df9, 261:$p<.01$) and in Group NC5 ($F = 3.15$; df4, 116:$p<.05$). Newman-Keuls comparisons for repeated measures across positions show for Group NC10 that only position 5 solutions are of higher quality than solutions in positions 1 or 3 ($p<.05$). For Group NC5 the same comparisons show that the solutions in position 4 and 5 are better than those in the first position ($p<.05$).

The question posed initially may be better answered by analyzing the variance between positions for a linear trend component. The curve of best fit should be linear and have a positive slope. A trend analysis for

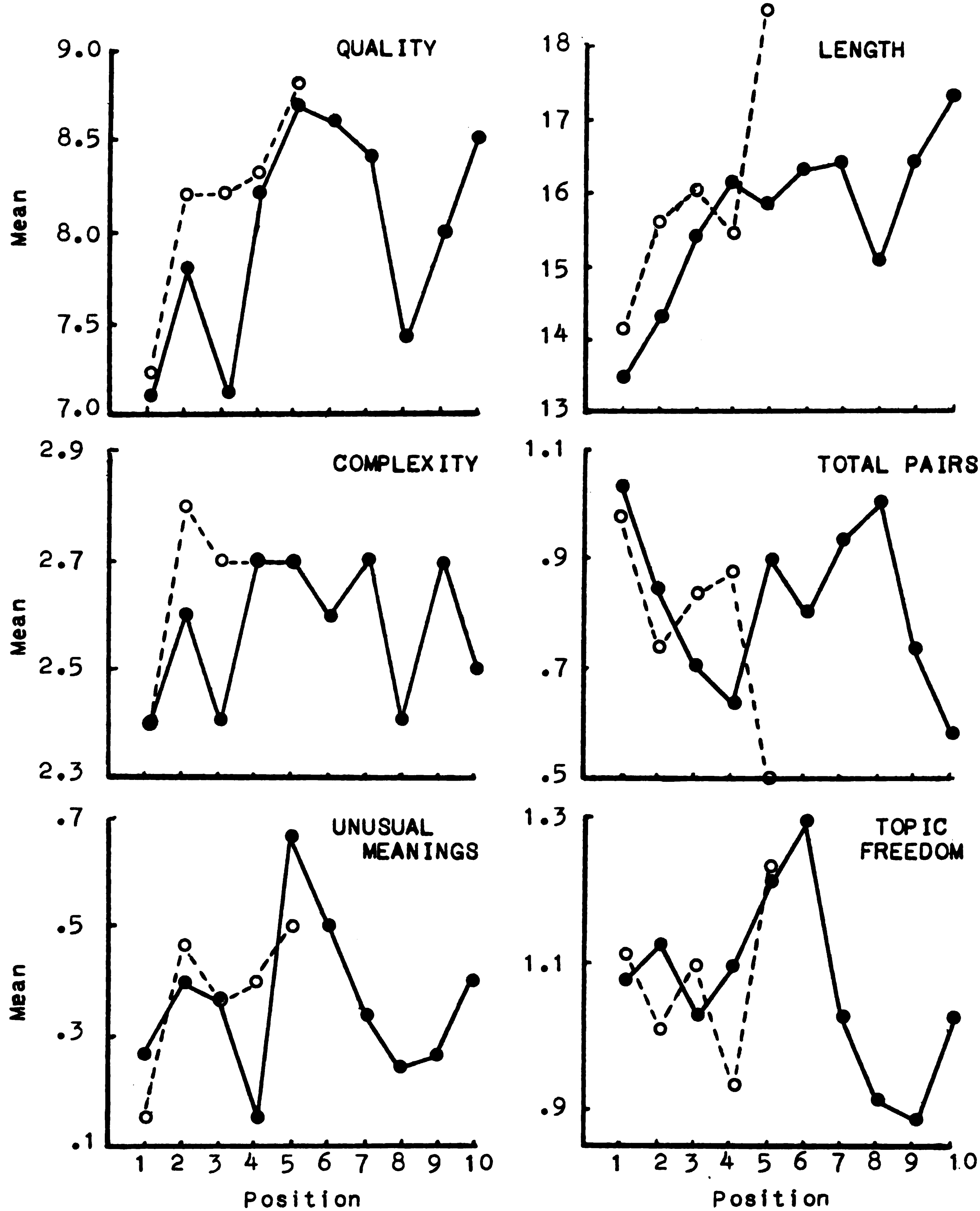

Fig. 1. Group NC10, ten solutions (solid line) and Group NC5 (broken line). Means for response measures as a function of position.

repeated measures indicates for Group NC10 that 22.6% of the variance between positions may be predicted by a linear regression. Nonlinear trends are nonsignificant. The linear equation giving the best fit for quality in Group NC10 is: $Y = .095X + 7.47$, where Y is quality and X is position. The slope is small but positive, and the linear correlation between position and quality is .128. Since this correlation takes into account the within-position variance ($n = 300$), this correlation is statistically significant at the .05 level, but it accounts for only a meager portion of the total variance (1.6%).

For Group NC5 the linear regression accounts for 79.8% of the variance between position with no significant nonlinear trends. The curve of best fit is: $Y = .343X + 7.12$. The linear correlation ($n = 150$) between position and quality is .232 which is significant beyond the .01 level and accounts for roughly 5.4% of the total variance.

Predicting Quality

In both groups there is a linear increase in quality over the response sequence. The next question is, "How can learning variables account for this change?" This section will probide a partial answer by looking at multiple linear regression analyses and analyses of variance.

The present experiment emphasizes the importance of quality-related response measures such as unusual meanings and old pairs, because in judging solutions from previous experiments these variables were found to be influential in determining solution quality (see Appendix A). Table 2 gives the correlations between response measures for Groups NC10 and NC5. Correlations between single solutions are confounded by the covariance due to each S writing several solutions. To prevent this, response measures were summed to give a total score for each S on each measure. Thus, the high negative correlation between total pairs and sentence length in Group NC10 indicates that Ss writing short sentences also used many pairings of the given words. Subsequent regression analyses of necessity must predict performance for Ss rather than for individual solutions.

Correlations ($n = 30$) above .349 and .449 are significant beyond the .05 and .01 levels respectively. Both groups show significant correlations between quality and unusual meanings and topic freedom. For Group NC5 length and total pairs are also correlated with quality. Note also that total pairs and length are correlated in Group NC10 but not in Group NC5.

Multiple linear regression analysis gave partial correlations between quality and each response measure.

TABLE 2.--Intercorrelation Matrices for Response Measures of Solutions from Group NC10 (below diagonal) and Group NC5 (above diagonal).

Response Measure	1.	2.	3.	4.	5.	6.
1. Quality	-	.40	.08	-.46	.53	.42
2. Length	.26	-	.24	-.01	.12	.02
3. Complexity	-.04	.05	-	-.01	.12	.02
4. Total Pairs	-.05	-.69	.07	-	.26	-.08
5. Unusual Meanings	.65	-.09	.11	.27	-	.39
6. Topic Freedom	.45	-.07	-.21	.24	.53	-

Darlington (1968) states that such an analysis cannot give an accurate estimate of the relative importance of predictor variables when the predictor variables themselves are intercorrelated. Thus, the results cannot be interpreted to indicate the amount of the independent contribution to variance in quality. It only indicates which variables are useful in predicting solution quality, although relative usefulness still cannot be inferred from the data. The present results were obtained after a stepwise deletion of nonsignificant variables. For Group NC10 the only variables which predicted quality were sentence length (partial correlation = .43) and unusual meanings (.71) accounting for 53% of the variance in quality. For Group NC5 total pairs (-.74) and unusual meanings (.77) were the only useful variables and accounted for 68% of the variance.

Response Measures and Position

Unusual meanings and total pairs were hypothesized to reflect learning and do accurately predict quality. That is to say, those Ss who write few total pairs and many unusual meanings will write higher quality solutions. The next step, then, is to see which measures change over the response sequence as does quality. Figure 1 presents this information. (a) The quality of solutions in Group NC10 appears to increase for the first five solutions for length, complexity, unusual meanings and topic freedom. Total pairs appears to decrease for these solutions. (b) Quality drops off from solutions five to eight, and decreases are also found in unusual meanings and topic freedom. Total pairs appears to increase from solutions four to eight. (c) Quality also seems to increase for solutions nine and ten. Length, unusual meanings and topic freedom increase, and total pairs decreases for these solutions.

Sincle-factor analyses of variance (df9, 261) test the statistical significance of these changes over the response sequence. For Group NC10 quality ($F = 2.76$, $p < .01$), and length ($F = 3.23$, $p < .01$) show the only significant increases. Unusual meanings with an F-value of 1.66 (df9,261) comes the closest of the other variables to being statistically significant ($p < .1$).

Figure 1 shows that in Group NC5 solutions there is an increase in quality for all five solutions with the greatest increase between the first two solutions. Length, complexity, and unusual meanings show similar increases. Total pairs shows a decrease over the response sequence. Single-factor analyses of variance (df4,116) show that the increases in quality ($F = 3.15$, $p < .05$) and length ($F = 77.43$, $p < .01$) are statistically significant.

Effects of Quantity Instructions

Groups NC10 and NC5 differ only in that one wrote ten solutions and the other wrote five solutions. In terms of the increase in quality over the response sequence Figure 1 shows that Ss writing five solutions write better solutions earlier in the response sequence than those Ss writing ten solutions. But, as Table 3 shows, the means for all solutions in each group do not differ for quality or any other response measure. Figure 1 indicates that each group produces solutions of about equivalent quality over the first five solutions apparently decrease in quality.

Comparing the first five solutions of Groups NC10 and NC5 by post hoc analyses will illustrate changes in the solution process as a result of differing quantity instructions. The two-factor analyses of variance with repeated measures on position (Table 4) show that the first five solutions of Groups NC10 and NC5 do not differ on any response measure. There are increases over position for quality ($p<.01$), length ($p<.01$), complexity ($p<.05$) and unusual meanings ($p<.05$). The absence of significant interactions reinforces the decision that the first five solutions are the same whether produced as the entire assignment or as only half of the assigned task.

Furthermore, a linear regression analysis of the first five solutions in Group NC10 produces a curve of

TABLE 3.--Means and SD's for Response Measures of Solutions in Groups NC10 and NC5.

Response Measures		Group NC10	Group NC5
Quality	$\bar{X}$	7.99	8.15
	SD	2.14	2.10
Length	$\bar{X}$	15.64	15.93
	SD	4.08	4.35
Complexity	$\bar{X}$	2.57	2.65
	SD	.74	.59
Total Pairs	$\bar{X}$	.81	.77
	SD	.94	.94
Unusual Meanings	$\bar{X}$	.36	.37
	SD	.72	.80
Topic Freedom	$\bar{X}$	1.08	1.10
	SD	.60	.56
N		300	150

TABLE 4.--ANOVA Summary for Response Measures of the First Five Solutions of Groups NC10 and NC5.

Source	df	MS	F	MS	F	MS	F
		Quality		Length		Complexity	
Treatment	1	9.36	1.53	64.40	1.82	.65	1.80
Error	29	6.12		35.38		.36	
Position	4	23.24	6.29**	92.18	6.66**	.99	2.45*
T x P	4	3.16	.85	25.30	1.83	.33	.82
Error	261	3.69		13.83		.41	
		Total Pairs		Unusual Meanings		Topic Freedom	
Treatment	1	.21	.14	.00	.00	.05	.12
Error	29	1.57		1.05		.43	
Position	4	.92	1.09	1.40	2.45*	.40	1.24
T x P	4	1.09	1.28	.45	.80	.14	.45
Error	261	.85		.57		.32	

*p<.05 **p<.01

best fit which is: Y = .360X + 6.72. Nonlinear trends are nonsignificant. This is very comparable to the same curve for Group NC5 which is: Y = .343X + 7.12. Groups NC5 and the first five solutions of Group NC10 appear to be very similiar in terms of the rate of quality increase and in the fact that for both conditions quality and length increase over the response sequence. Furthermore, quantity instructions do not significantly influence production in terms of quality or any other response measure.

Criteria-cued Instructions

This section addresses itself to the effects of criteria-cued instructions on the increase in quality and the change in response measures over the response sequence. The interaction of criteria-cued instructions and quantity instructions will also be presented.

Quality and Position

Figure 2 presents the mean quality for each position for Group CC10 with ten solutions per S and for Group CC5 with five solutions per S. It is apparent that quality increases in Group CC5 and not in Group CC10. Single-factor analyses of variance for repeated measures on position show that there is no increase in quality over the response sequence for Group CC10 ($F<1$). For Group CC5 the increase in quality is statistically significant

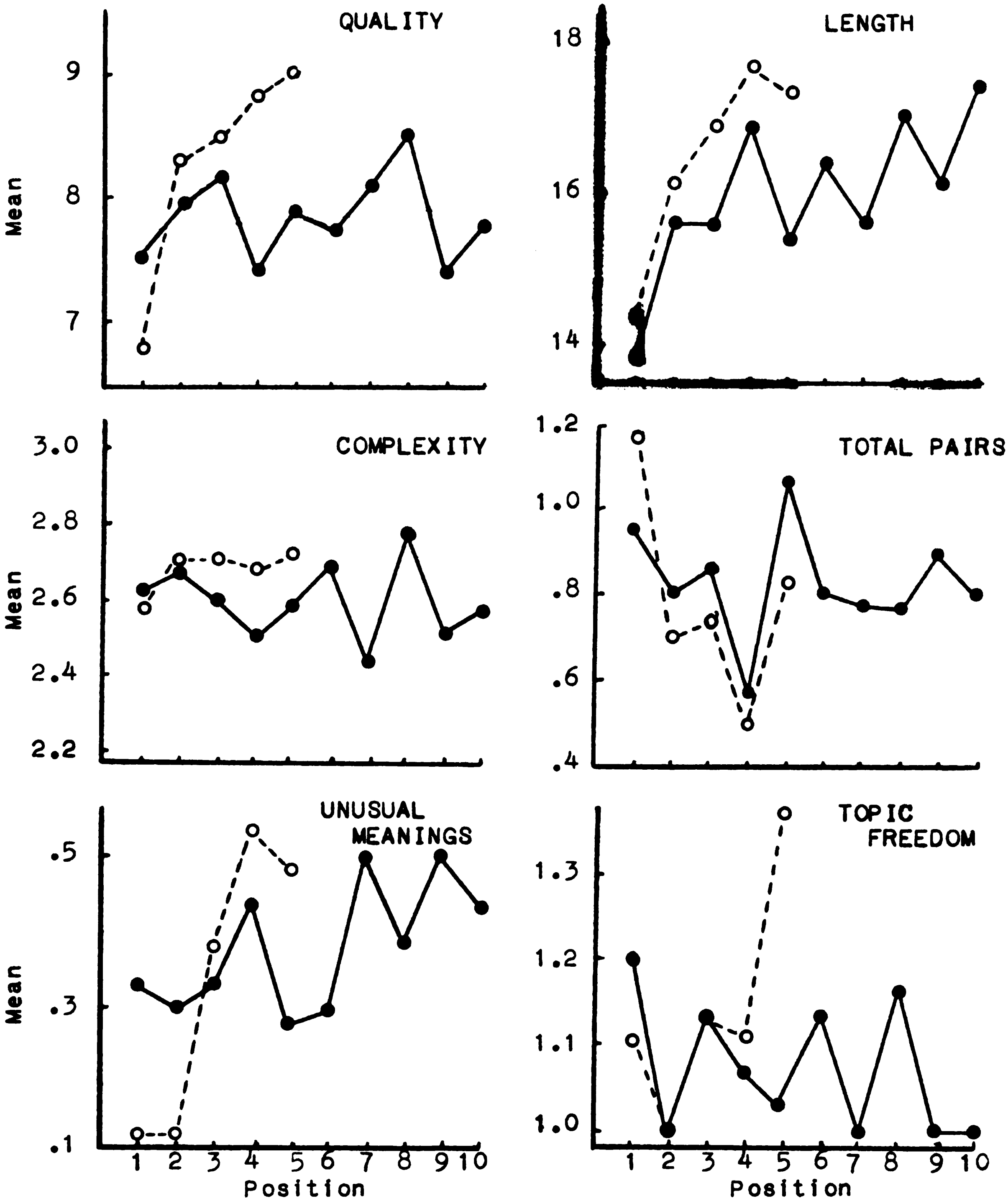

Fig. 2. Group CC10, ten solutions and criteria-cues (solid line) and Group CC5, five solutions and criteria-cues (broken line). Means for response measures as a function of position.

($F<1$). For Group CC5 the increase in quality is statistically significant ($F = 6.12$; df 4, 116; $p<01$). For the latter group Newman-Keuls comparisons, adjusted for repeated measures, indicate that solutions in the first position are significantly lower in quality than those in other positions ($p<.01$).

The linear trend in the increase in the quality of Group CC5 solutions accounts for 81.27% of the variance between positions. Nonlinear trends are nonsignificant. The regression line is: $Y = .346X + 7.25$. The linear correlation between position and quality ($n = 150$) is .335 which accounts for about 11.2% of the total variance.

Predicting Quality

The correlations between solution quality and response measures are given in Table 5. The correlations are based on the sum for each S across all solutions ($n = 30$). Values above .349 and .449 are significant beyond the .05 and .01 levels respectively. In Group CC10 quality is correlated with sentence length, unusual meanings, and topic freedom. For Group CC5 quality is correlated with sentence length and complexity. Notice that sentence length and total pairs show high negative correlations in both groups.

Multiple linear regression analysis using a stepwise deletion procedure was used to determine which variables were useful in predicting quality. For Group

TABLE 5.--Intercorrelation Matrices for Response Measures of Solutions in Group CC10 (below diagonal) and Group CC5 (above diagonal).

Response Measure	1.	2.	3.	4.	5.	6.
1. Quality	-	.61	.70	-.36	.32	.32
2. Length	.36	-	.43	-.63	-.14	.23
3. Complexity	.06	.42	-	-.12	.26	.37
4. Total Pairs	-.02	-.63	-.29	-	.21	.13
5. Unusual Meanings	.61	-.16	-.07	.18	-	.05
6. Topic Freedom	.64	.04	.11	.02	.55	-

CC10 sentence length (partial correlation = .59), unusual meanings (.57) and topic freedom (.48) account for 68% of the variance in solution quality. For Group CC5 sentence length (.57), unusual meanings (.40) and complexity (.54) account for 67% of the variance.

<u>Response Measures and Position</u>

According to the original predictions, criteria-cues should eliminate any increase in quality over the response sequence. This is true for Group CC10 but not for Group CC5, which differed only in the number of solutions written. Changes in influential response measures should also reflect this difference between groups. Single-factor analyses of variance with repeated measures over position show that for Group CC10 only sentence length changes over position ($F = 2.24$; df9, 261:$p<.05$). Newman-Keul's comparisons show only solution 10 to be longer than solution 1 ($p<.05$). For Group CC5 position differences (df 4, 116) are significant for length ($F = 4.33$, $p<.01$), total pairs ($F = 3.02$, $p<.05$), and unusual meanings ($F = 2.73$, $p<.05$). For length, individual comparisons show the first solutions to be shorter than the rest ($p<.05$). Solutions in the fourth position have fewer word pairs than the first solution ($p<.01$), and the fourth solution also has more unusual meanings than the first two solutions ($p<.05$). Apparently the increase in quality in Group CC5 is also accompanied by increases in the use of unusual meanings and a decrease in the number of word pairs. On the other hand, Group CC10 which does not increase in quality does not show changes in the use of unusual meanings or word pairs.

Effects of Quantity Instructions

Previously comparisons between solutions produced under different quantity instructions proved to be valuable. Figure 2 indicates that such a *post hoc* analysis would also point out some meaningful differences between Groups CC10 and CC5. Table 6 gives the means and standard deviations based on all solutions in each group. Two-tailed t-tests show that solution quality for Group CC5 is slightly higher than Group CC10 ($p<.05$; df448). Other differences are not statistically significant.

Table 7 presents the results for two-factor analyses of variance with repeated measures across the first five solutions of each group. There are no significant differences between the means of the groups for the first five solutions. Only the treatment-by-position interaction for quality approaches significance ($p<.06$). From Figure 2 the quality of Group CC5 solutions appears to increase faster than the quality of Group CC10 solutions. Increases across position are significant for quality and length ($p<.01$). Total pairs significantly decreases over the

TABLE 6.--Means and SD's for Response Measures of Solutions in Groups CC10 and CC5.

Response Measures		Group CC10	Group CC5	t
Quality	$\bar{X}$	7.85	8.29	1.99*
	SD	2.29	2.09	
Length	$\bar{X}$	15.98	16.49	1.14
	SD	4.55	4.38	
Complexity	$\bar{X}$	2.59	2.67	1.14
	SD	.76	.62	
Total Pairs	$\bar{X}$	.83	.79	.48
	SD	.91	.83	
Unusual Meanings	$\bar{X}$	.38	.33	.79
	SD	.67	.66	
Topic Freedom	$\bar{X}$	1.07	1.14	1.30
	SD	.58	.54	
N		300	150	

*$p<.05$

TABLE 7.--ANOVA Summary for Response Measures of the First Five Solutions in Groups CC10 and CC5.

Source	df	MS	F	MS	F	MS	F
		Quality		Length		Complexity	
Treatment	1	19.76	3.71	82.16	2.74	.48	.57
Error	29	5.32		29.96		.85	
Position	4	16.20	3.88**	82.54	4.89**	.10	.23
T x P	4	9.57	2.29	5.52	.33	.15	.33
Error	261	4.17		16.87		.44	
		Total Pairs		Unusual Meanings		Topic Freedom	
Treatment	1	.33	.36	.00	.00	.21	.44
Error	29	.93		.49		.49	
Position	4	2.48	3.68**	.71	1.70	.34	1.17
T x P	4	.39	.58	.44	1.06	.41	1.38
Error	261	.67		.42		.29	

**p<.01

response sequence ($p<.01$). Figure 2 shows that other than for quality the first five solutions for each group are very similar for all response measures.

Quality and Quantity Instructions

Up to now solutions produced with criteria-cued instructions have not been compared to solutions produced with standard instructions. Table 8 compares the means for all solutions produced under each condition of quality and quantity instructions. A 2 x 2 factorial analysis of variance compares these means which are based on a mean quality per S. The results show that there is no significant difference due to criteria cued instructions ($F<1$) and no significant difference due to quantity instructions ($F = 2.56$; df 1, 116). The interaction was also nonsignificant. Apparently when each S is specifically assigned a number of solutions to produce, mean quality does not depend on how many solutions he is required to write or what type of instructions he is given.

Effects of Learning Old Pairs

In Group TOP Ss learned sentences with the old pairs happy horse and expensive lake. Group TR S learned sentences with the same words, but they were unpaired. Group TIR Ss learned sentences with four other words. and Group NC10 Ss had no learning experience during the

TABLE 8.--Means and SD's for Quality of Solutions in Groups NC10, NC5, CC10 and CC5.*

			Quality Instructions		
			No Criteria-Cues	Criteria-Cues	Total
Quantity Instructions	5 Solutions	$\overline{X}$	8.15	8.29	8.22
		SD	1.17	.94	1.07
	10 Solutions	$\overline{X}$	7.99	7.85	7.92
		SD	.87	1.04	.96
	Total	$\overline{X}$	8.07	8.07	8.07
		SD	1.02	1.04	1.03

*Based on mean quality per S.

n = 30 per cell.

experiment. The hypothesis is that learning old pairs will make Group TOP Ss use more old pairs in sentences but that such use will decrease as more solutions are produced. As a result of the reduction in the use of old pairs, one would also expect quality to increase over the response sequence in Group TOP.

Use of Word Pairs

Figure 3 presents the mean number of old pairs in each position for each group.[2] (Group means are given in Table 9). The results of a two-factor analysis of variance with repeated measures on position are given in Table 16. Group TOP Ss used more old pairs than other Ss ($p<.01$) and Ss in Group TR used more old pairs than Ss in Groups NC10 or TIR ($p<.05$) by Newman-Kuels tests. Position has a significant effect ($p<.05$) with solutions in position 1 having more old pairs than those in position 6 ($p<.05$) by Newman-Keuls tests. The significant interaction between treatment and position ($p<.01$), which can be seen in Figure 3, can be analyzed by single-factor analyses of variance for each group separately with position being the repeated measure. Only Group TOP shows a significant decrease in the use of old pairs

[2]Learning sentences with old pairs lowered the variance of total pairs in the first solution from .87 for Group 1 total pairs to .46 for Group 3. For quality the variance in the first solution was 3.70 for Group 1 and 2.96 for Group 3.

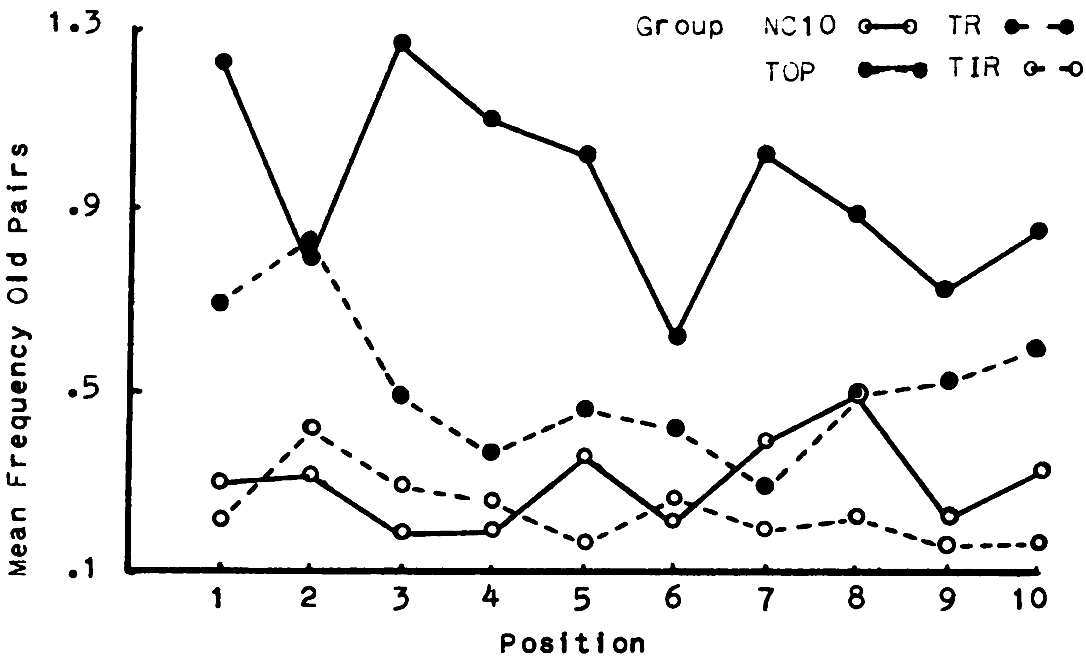

Fig. 3. Groups NC10, TOP, TR and TIR. Mean number of old pairs as a function of position.

TABLE 9.--ANOVA Summary for Old Pairs in Solutions of Groups NC10, TOP, TR and TIR.

Source	df	SS	MS	F
Treatment	3	94.14	31.38	25.72*
Error	116	141.26	1.22	
Position	9	5.72	.64	1.89*
T x P	27	17.07	.63	1.88**
Error	1044	351.71	.34	

*p<.05 **p<.01

(F = 2.92; df 9, 261; p<.01). Newman-Keuls comparisons indicate that in Group TOP solutions 6 and 9 have significantly less old pairs than solutions 1 and 3 (p<.05).

Old pairs are <u>happy</u> <u>horse</u> and <u>expensive</u> <u>lake</u>, and total word pairs are all combinations of these four words. The effect of learning sentences with the four words to be used later can also be seen in the frequency of use of total word paris. Table 10 shows that Groups TOP and TR used more word pairs than other groups. Table 11 gives the results of a two-factor analysis of variance with repeated measures on position. The difference between treatments (p<.01) is attributable to Group TOP using more word pairs than all other groups (p<.01), and Group TR using more than Groups 1 and 6 (p<.05), which do not differ. The use of word pairs decreases over the response sequence (p<.01). Individual comparisons indicate the first two solutions have more word pairs than other solutions (p<.01), and all solutions have more word pairs than the last solution (p<.01).

Single-factor analyses of variance with repeated measures (df 9, 261) shows that the use of word pairs decreases over the response sequence for Group TOP (F = 2.28; p<.05) and Group TR (F = 2.55; p<.01), but not for Groups NC10 (F = .91) or TIR (F = 1.78). Apparently learning sentences with the four words to be used later in solving provlems has the effect of increasing the use

TABLE 10.--Means and SD's for Response Measures of Solutions in Groups NC10, TOP, TR, and TIR.

Response Measures		Group TOP	Group TR	Group NC10	Group TIR
Quality	$\bar{X}$	6.89	7.56	7.99	8.08
	SD	2.02	1.96	2.14	2.12
Length	$\bar{X}$	12.79	14.26	15.64	15.53
	SD	3.72	3.66	4.08	4.74
Complexity	$\bar{X}$	2.47	2.63	2.57	2.63
	SD	.65	.55	.74	.62
Old Pairs	$\bar{X}$	.96	.52	.31	.24
	SD	.84	.67	.55	.48
Total Pairs	$\bar{X}$	1.51	1.10	.81	.80
	SD	.94	1.00	.94	.89
Unusual Meanings	$\bar{X}$	.13	.11	.36	.29
	SD	.44	.42	.71	.63
Topic Freedom	$\bar{X}$	.86	.97	1.08	1.04
	SD	.54	.56	.60	.58

TABLE 11.--ANOVA Summary for Response Measures of Solutions in Groups NC10, TOP, TR and TIR.

Source	df	MS	F	MS	F	MS	F
		Quality		Length		Complexity	
Treatment	3	89.20	7.14**	533.66	8.36	1.86	3.34*
Error	116	12.49		63.80		.56	
Position	9	19.89	6.17**	199.85	20.08**	1.13	2.89**
T x P	27	3.15	.98	7.56	.76	.42	1.10
Error	1044	3.23		9.95		.39	
		Total Pairs		Unusual Meanings		Topic Freedom	
Treatment	3	32.75	16.52**	4.49	5.29**	2.89	5.81**
Error	116	1.98		.85		.50	
Position	9	2.92	3.87**	.41	1.21	.47	1.50
T x P	1044	.75		.34		.31	

*p<.05 **p<.01

of pairs of these words in problem solutions. Learning sentences with specific pairs increases the use of those specific pairs, but in both cases the use of word pairs decreases over the response sequence.

It is important to ask about individual word pairs, because it provides information about how the treatments change the associative hierarchy. The solutions were reread and frequency of use was tallied for each word pair. The obtained totals may be slightly different, because the present judge may have used a stricter criterion for a word pair than used by the original judge.

Table 12 gives for each word pair in Groups NCIO, TOP and TR the frequency of occurrence and average positions of occurrence. In every group the average position of occurrence of a word pair is inversely related to the frequency of occurrence. That is, the more popular word pairs occur earlier in the response sequence than less popular pairs. Expensive horse is the most popular word pair in the groups which did not learn old pairs. Group TOP learned the pairs happy horse and expensive lake, and they are the most popular word pairs in this group. Apparently the most common word pairs are used before less common ones. Learning specific word pairs makes them more common and makes them occur earlier than in other groups. When relevant sentences without word pairs are learned, the

TABLE 12.--Frequency of Use and Position for Each Word Pair in Solutions for Groups NC10, TOP and TR.

	Mean Position	Median Position	Frequency
		Group NC10	
1. Expensive Horse	5.01	4.50	106
2. Happy Horse	5.83	5.75	53
3. Expensive Lake	7.10	7.40	21
4. Horse Lake	8.18	8.78	11
5. Happy Lake	7.86	8.50	7
Total	6.80	6.99	198
		Group TOP	
1. Happy Horse	4.92	3.80	136
2. Expensive Lake	5.03	4.00	106
3. Expensive Horse	5.60	5.10	91
4. Happy Lake	6.16	5.75	19
5. Horse Lake	6.75	7.00	4
Total	5.69	5.13	356
		Group TR	
1. Expensive Horse	4.54	4.00	114
2. Happy Horse	4.90	4.00	82
3. Expensive Lake	6.57	6.80	42
4. Happy Lake	7.67	7.00	6
5. Horse Lake	7.00	7.00	3
Total	6.00	5.76	247

word pairs are used more frequently but in the same associative hierarchy as in the group with no learning.

<u>Other Response Measures</u>

Treatment conditions also produce significant differences on other response measures as well. Table 11 also gives the results of these analyses of variance. Individual comparisons show that Group TOP solutions are shorter in length ($p<.05$), have less topic freedom ($p<.05$), and use more word pairs ($p<.01$) than all other solutions. They are as complex as Group TR and TIR solutions ($p<.05$). They have as few unusual meanings as Group TR solutions, but both have less than solutions in Groups NC10 and TIR ($p<.05$). Group TR solutions differ from those in Groups NC10 and TIR only in having more word pairs ($p<.05$) and less unusual meanings ($p<.05$). Group NC10 and TIR solutions do not differ on any response measure. Learning sentences with old pairs not only decreases mean quality, but also increases the use of word pairs and produces shorter and less complex solutions which have fewer unusual meanings and less topic freedom than most other groups. Learning sentences which use the words to be later incorporated in solutions seems to increase the use of word pairs and decrease the use of unusual meanings.

Solutions in Group TOP use word pairs less often as more solutions are produced, and solution quality increases. Figure 4 shows this increase in quality. A single-factor

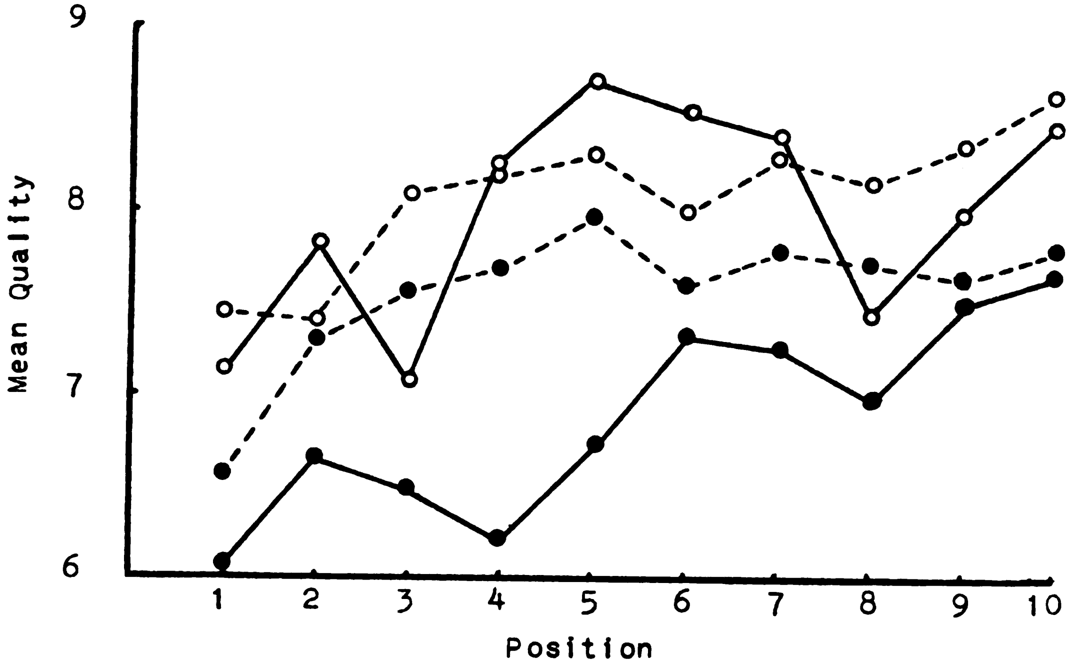

Fig. 4. Groups NC10, TOP, TR and TIR. Mean quality as a function of position.

repeated-measures (position) analysis of variance of quality yielded an F-value of 3.17 for position (df 9, 261; $p<.01$). A linear trend accounts for 80.36% of the variance between positions with no significant nonlinear trends. The curve of best fit is: $Y = .164X + 5.99$. The linear correlation between position and quality is .234 ($n = 300$) which is statistically significant at the .01 level. With the decrease in the use of old pairs Group TOP also shows a linear increase in solution quality over the response sequence.

The same analysis of variance for other response measures (df = 9, 261) for Group TOP shows that other response measures do change over the response sequence. Length ($F = 8.43$; $p<.01$) and complexity ($F = 2.12$; $p<.05$) increase over the response sequence. The use of all word pairs, of course, decreases over the response sequence ($F = 2.28$; $p<.05$). The frequency of use of unusual meanings and topic freedom do not change. For Group TOP the increase in solution quality over the response sequence is accompanied by a decrease in the use of all word pairs and increases in solution length and complexity.

The same analysis (df 9, 261) can be applied to Group TR, since these solutions so closely resemble those produced by Group TOP. Quality increases slightly over the whole response sequence ($F = 1.84$; $p = .06$) and significantly over the first five solutions ($F = 5.03$; df 4,116; $p<.01$).

Sentence length ($F = 5.62$; $p<.01$) increases and use of all word pairs decreases ($F = 2.55$; $p<.01$). There are no significant changes in sentence complexity, unusual meanings or topic freedom. For Group TR a slight increase in quality is also accompanied by large increases in sentence length and decreases in the use of word pairs.

Solution Quality

Learning sentences with old pairs affects the subsequent use of old pairs, but it also affects solution quality and other response measures. Table 10 presents the means and standard deviations for quality in each group. Table 11 presents the results of two-factor analyses of variance with one repeated measure for quality. The main questions under investigation are whether quality increases over the response sequence and under what conditions is this the case.

As Figure 4 and Table 11 indicate, quality does vary between treatments ($p<.01$) and between positions ($p<.01$). Newman-Keuls comparisons show that Group TOP solutions are of inferior quality to other groups ($p<.01$), which do not differ from one another. Totaling across all groups shows that quality does increase over the response sequence with the lowest quality solutions occurring in the first three positions ($p<.05$). Solutions in positions 2, 3, 4, and 8 do not differ from one another but are of lower quality than those in positions 5, 6, 7, 9, and 10 ($p<.05$).

Single-factor analyses of variance with position as a repeated measure (df 9, 261) show the change in quality over position for individual treatments. Quality increases in Group NC10 ($F = 2.76$; $p<.01$), Group TOP ($F = 3.17$; $p<.01$) and Group TR approaches significance ($F = 1.84$; $p = .06$). The quality of the first five solutions in Group TR does increase significantly ($F = 5.03$; df 4, 116; $p<.01$). It is apparent that learning sentences with old pairs increases the use of old pairs and reduces solution quality beyond the effects of control conditions. Learning sentences with the four words to be later used in solutions increases the use of word pairs, but not as much as learning specific pairs. Solution quality for the latter condition is not significantly reduced. For most conditions solution quality increases over the response sequence.

Predicting Quality

Group TOP shows a large increase in quality over the response sequence and concomitant increases in length and decreases in the use of total pairs. The next question is whether the decrease in word pairs has any effect on quality. The results for Group NC10 indicated that total word pairs did not have significant influence on the judges' rating of solution quality. Table 13 presents the intercorrelation matrix for relevant response measures in Group TOP solutions. The correlations are based on total scores per S ($n = 30$). Thus, correlations above .349 and .449 are significant beyond

TABLE 13.--Intercorrelation Matrix for Group TOP.

Response Measure	1.	2.	3.	4.	5.	6.
1. Quality	-					
2. Length	.79	-				
3. Complexity	.50	.52	-			
4. Total Pairs	-.87	-.71	-.43	-		
5. Unusual Meanings	.71	.54	.33	-.53	-	
6. Topic Freedom	.30	.30	-.02	-.41	.24	

the .05 and .01 levels respectively. Note that quality is highly correlated with length, complexity, total pairs and unusual meanings. Note also that total pairs and length are highly negatively correlated. A stepwise deletion procedure was used for the multiple linear regression analysis to predict solution quality from the other response measures. Partial correlations with quality were -.65 for total pairs, .51 for length and .54 for unusual meanings. Together these variables account for 85% of the variance in solution quality. It cannot be concluded, however, that the decrease in the use of word pairs is the only reason for the increase in quality over the response sequence in Group TOP.

Word Pairs in Group NC10

The validity of learning sentences with old pairs could be questioned as representing a condition which would not happen in an unselected sample. That is, in Group NC10 *Ss* who wrote solutions with different numbers of old pairs would not differ in quality, even if the number of unusual meanings were equated. This would imply that the low correlation between quality and old pairs in Group NC10 was true and not due to the cancelling effect of unusual meanings which included word pairs. To test this the 30 *Ss* in Group NC10 were separated into the third with the highest and the third with the lowest number of word pairs. Five *Ss* in each sample could be matched exactly for the number of unusual meanings used. The high *Ss* used an average of 1.38

word pairs per solution (compared to 1.51 for Group 3), and the low Ss used an an average of .38. The mean quality for the high Ss was 7.22 (SD = .68) and for the low Ss 8.24 (SD = .63). The t-value is 2.46 (df = 8) which is significant at the .05 level for a two-tailed test. Thus, quality is determined, at least in part, by the use of word pairs for Ss under normal conditions.

Effects of Learning Unusual Meanings

Previous analyses have shown unusual meanings to be one of the most, if not the most, influential variable in determining solution quality. Group TUM Ss learned sentences which used the four words in unusual ways, sometimes giving them novel meanings and sometimes giving them a different grammatical function. Using a noun as a modifier would be an example of the latter use. The comparison groups are the same as for the previous treatment. Group NC10 Ss had no learning, Group TIR Ss learned irrelevant sentences, and Group TR Ss learned relevant sentences with no unusual meanings and word pairs. The hypotheses are that Ss learning unusual meanings will use more of them in solutions to the sentence problem, and that the use of unusual meanings will not increase over the response sequence. Based on the use of unusual meanings alone, one would not expect quality to increase over the response sequence in Group 4.

Use of Unusual Meanings

Figure 5 presents the mean frequency of unusual meanings for each position in each group. Table 14 presents the results of a two-factor analysis of variance with position as the repeated measure. Treatment groups differ in the frequency of use of unusual meanings ($p<.01$). There is no significant change in the use of unusual meanings over position, nor is there a significant interaction. More unusual meanings are used by Group TUM than by other groups ($p<.01$). But Group TR uses significantly less than Group NC10 ($p<.05$). Groups NC10 and TIR do not differ by Newman-Keuls tests. Apparently learning sentences with unusual meanings does increase the subsequent use of unusual meanings in solutions to the sentence problem. But learning sentences with the words to be used later, decreases the use of unusual meanings.

Solution Quality

Regression analyses have shown unusual meanings to be about the most important variable in determining solution quality. And Group TUM seems to have solutions of better quality as shown by Figure 6 and Table 15. The results of two-factor analyses of variance with position as the repeated measure are given in Table 16. They show that groups differ also on quality, total pairs and topic freedom. For quality, individual comparisons indicate that the best solutions are in Group TUM ($p<.01$), and other groups do not

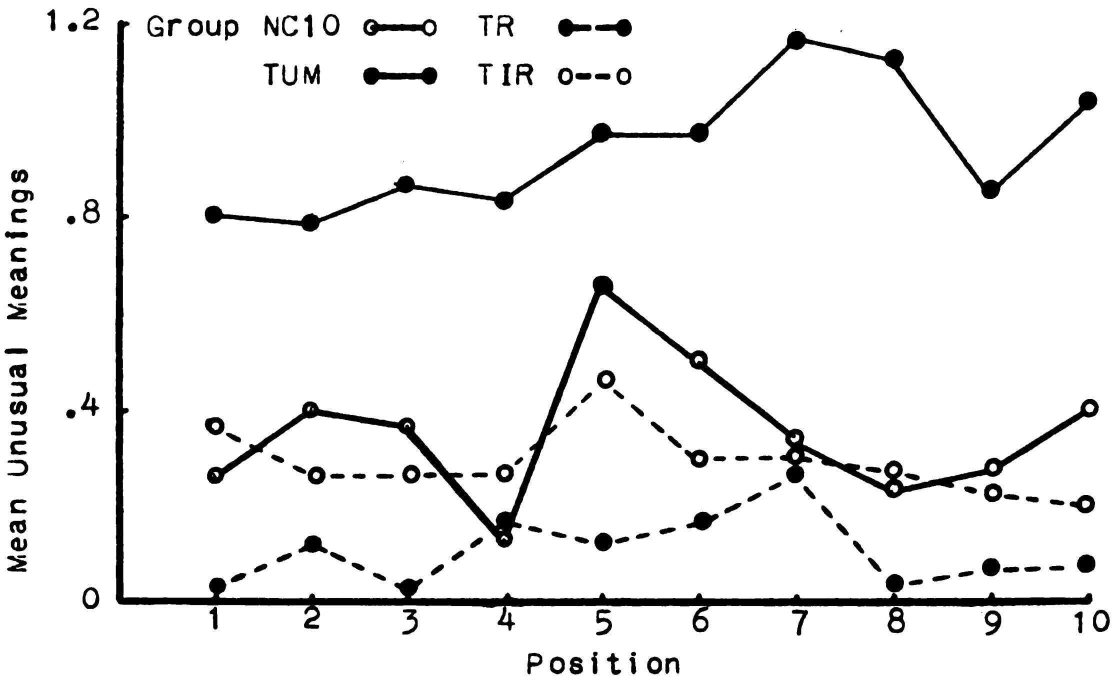

Fig. 5. Groups NC10, TUM, TR and TIR. Mean number of unusual meanings as a function of position.

TABLE 14.--ANOVA Summary for Unusual Meanings in Solutions of Groups NC10, TUM, TR and TIR.

Source	df	SS	MS	F
Treatment	3	113.89	37.96	24.02**
Error	116	183.46	1.58	
Position	9	5.81	.65	1.51
T x P	27	8.91	.33	.77
Error	1044	450.88	.43	

**p<.01

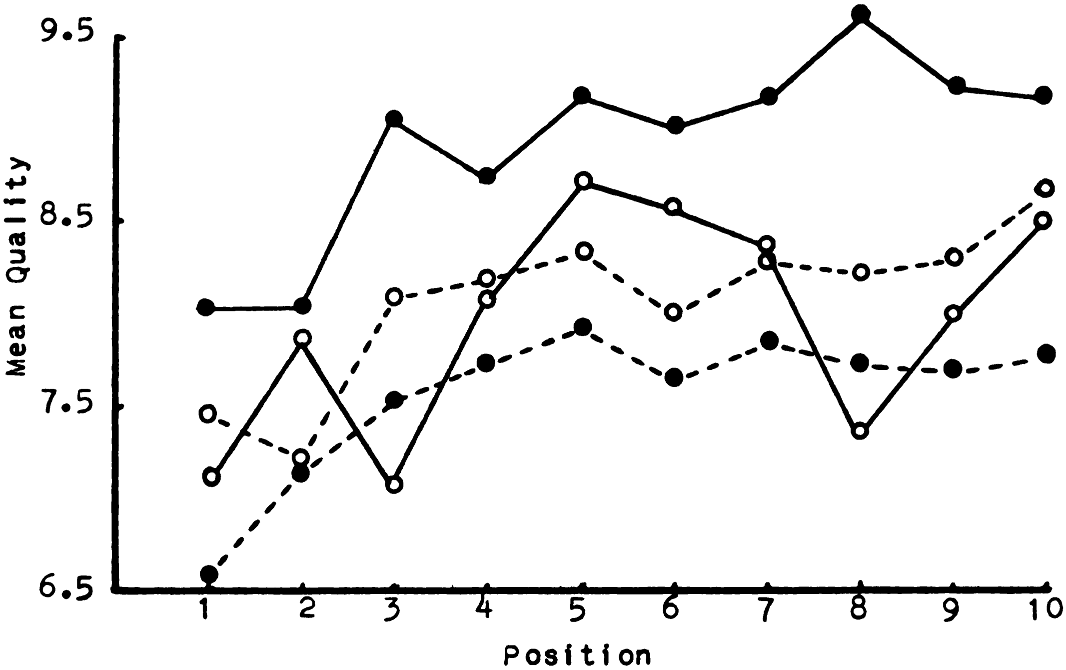

Fig. 6. Groups NC10, TUM, TR and TIR. Mean quality as a function of position.

TABLE 15.--Means and SD's for Response Measures of Solutions in Groups NC10, TUM, TR and TIR.

Response Measures		Group TUM	Group TR	Group NC10	Group TIR
Quality	$\bar{X}$	8.94	7.56	7.99	8.08
	SD	2.09	1.96	2.14	2.12
Length	$\bar{X}$	14.46	14.26	15.64	15.53
	SD	4.19	3.66	4.08	4.74
Complexity	$\bar{X}$	2.60	2.63	2.57	2.63
	SD	.57	.55	.74	.62
Total Pairs	$\bar{X}$	1.04	1.10	.81	.80
	SD	.99	1.00	.94	.89
Unusual Meanings	$\bar{X}$	.93	.11	.36	.29
	SD	1.04	.42	.71	.63
Topic Freedom	$\bar{X}$	1.19	.97	1.08	1.08
	SD	.59	.56	.60	.58

TABLE 16.--ANOVA Summary for Response Measures of Solutions in Groups NC10, TUM, TR and TIR.

Source	df	MS	F	MS	F	MS	F
		Quality		Length		Complexity	
Treatment	3	100.68	8.36**	153.04	2.25	.29	.48
Error	116	12.05		68.10		.59	
Position	9	19.74	5.86**	190.43	17.81**	.92	2.55**
T x P	27	2.99	.89	6.45	.60	.34	.93
Error	1044	3.37		10.69		.36	
		Total Pairs		Topic Freedom			
Treatment	3	6.86	3.64*	2.45	5.08**		
Error	116	1.89		.48			
Position	9	3.29	4.19**	.37	1.17		
T x P	27	.88	1.1]	.38	1.20		
Error	1044	.78		.32			

*p<.05 **p<.01

differ. Furthermore, quality increases over the response sequence ($p<.01$) with the worst solutions occurring in the first position ($p<.01$). Second position solutions are significantly inferior to all subsequent solutions ($p<.05$), except those in positions 3 and 4. Solutions in positions 3 to 10 do not differ by Newman-Keuls individual comparisons. Learning unusual meanings, then, increases solution quality above solutions produced under control conditions.

Other Response Measures

Groups also differ on the total number of word pairs used ($p<.05$) and the amount of topic freedom in solutions ($p<.01$). Newman-Keuls comparisons indicate that Group TUM and TR solutions use more word pairs than Groups NC10 and TIR ($p<.05$), which do not differ. There is also more topic freedom in Group TUM solutions than in other solutions ($p<.05$). Solutions produced after learning unusual meanings are of a higher quality and have more unusual meanings, more word pairs and more topic freedom than solutions produced under most other conditions. Learning sentences with the words to be used later has the affect of decreasing the number of unusual meanings and increasing the number of word pairs used in subsequent solutions.

Analyses of Group TUM Solutions

A single-factor repeated measures (position) analysis of variance for Group TUM indicates that quality changes over position ($F = 2.46$; df 9, 261; $p<.01$). Figure 6 shows only a linear trend is significant, and this accounts for 60.37% of the variance between positions. The curve of best fit is: $Y = .137X + 8.19$. The linear correlation between position and quality is .188 which is statistically significant at the .01 level ($n = 300$).

Since solution quality in Group TUM increases over the response sequence, the next question is whether the use of unusual meanings also increases over position. A single-factor analysis of variance with position as the repeated measure shows that there is no significant increase in the use of unusual meanings ($F = .72$; df 9, 261). From Table 17, however, it is apparent that judges did use unusual meanings to determine solution quality. A multiple linear regression analysis using a stepwise deletion procedure indicates that unusual meanings (partial correlation = .91) and sentence length (.61) are useful in predicting solution quality and account for 81% of the variance in quality.

If judges were aware of the use of unusual meanings, they must have been aware that some meanings were used more frequently than others. Thus, the frequent use of <u>Horse Lake</u> would reduce its effectiveness in a sentence. The solutions of Group TUM were reread by another naive judge for <u>post hoc</u> comparisons. Since the judges were different between the <u>a priori</u> and the <u>post hoc</u> comparisons, the obtained values may differ slightly. Table 18 presents the

TABLE 17.--Intercorrelation Matrix for Group TUM.

Response Measures	1.	2.	3.	4.	5.
1. Quality	-				
2. Length	.21	-			
3. Complexity	.04	.33	-		
4. Total Pairs	-.19	-.74	-.04	-	
5. Unusual Meanings	.70	-.44	-.16	.40	-
6. Topic Freedom	.58	-.07	.22	-.04	.59

TABLE 18.--Use of Unusual Meanings in Group TUM Solutions

	Freq.		Freq.
		Learned Meanings	
Names of Animate Objects		Names of Inanimate Objects	
Happy	30	Happy Horse Dude Ranch	1
Horse	3	Happy Hours Resort	0
Total	33	Lake Park	0
		Total	1
Unusual Word Meaning		Unusual Word Function	
Horsing around	10	Lake front	10
Horseback riding	6	Lake shore	2
Horse (heroin)	2	Horse laugh	2
Total	21	Total	14
Total learned meanings used	69		
		Created Meanings	
Names of Animate Objects		Names of Inanimate Objects	
Happy Horse	2	Horse Lake	22
Mr. Horse	1	Horse & Lake Liquor	1
Expensive	1	Horse Barn Road	1
Expensive Horse	1	Happy Horse Lake	4
Mr. Lake	3	Happy Horse Resort	2
Happy Lake	3	Happy Horse Stables	2
Joe Lake	1	Happy Horse Riding Stables	2
Lake . . .	3	Happy Horse	2
Total	14	Happy Horse Point	1
		Happy Horse Lake Hotel	1
Unusual Word Meaning		Happy Lake	14
Horseshoe	1	Happy Lake Haven	1
Horse face	1	Happy Lake Ranch	1
Ate like a horse	1	Happy Lake Resort	1
Lake of tears	1	Happy Lake Trading Post	1
Lake of beer	1	Happy Lake Horse Farms	1
		Happy Lake Horse Ranch	1
Total	5	Happy Lake Farms	1
		Happy Haven for Horses	1
Unusual Word Function		Happy Club	1
Lake cottage	5	Lake Happy	5
Lake resort	4	Lake Farms	3
Lake ranch	2	Lake Stables	2
Lake lots	2	Lake Resort	2
Lake water	2	Lake of Horses	1
Lake home	1	Lake Horse	1
Lake supervisor	1	Lake Happy Horse	1
Lakeside cottage	1	Lake-View Stables	1
Horse farm	2	Lake-View Races	1
Horse ranch	1	Lake Club	1
Horse barn	1	Lake Estates	1
Horse trailer	1	Lake Dude Ranch	1
Horse lovers	1	Lakeshore town	1
Horse hide purse	1	Lake Expensive	1
Horse skin bathing suit	1	Lakeville	1
Horse show	1	Expensive Lake	1
Total	27	Total	88
Total created meanings used	134		

frequency of use of each unusual meaning for Group TUM solutions. Notice that Happy as a person's name and Horse Lake were used 52 times in the 300 solutions of Group TUM (39% of the obtained unusual meanings). These particular meanings were also used in about 8% of the solutions in other groups (about 30% of the obtained unusual meanings). These meanings, then, may be called "common-unusual meanings." The question is whether the frequency of use of an unusual meaning would affect solution quality. The mean quality of Group TUM solutions with these common-unusual meanings is 9.44 (SD = .98), and for those with uncommon-unusual meanings the mean quality is 10.37 (SD = 1.02). The difference between these means is significant (t = 4.51; df 164) beyond the .01 level. Apparently those solutions with common-unusual meanings do have a lower quality than those with common meanings.

The position within the response sequence in which the unusual meanings occur is also a function of the frequency of occurrence of that meaning. Table 19 presents the mean

TABLE 19.--Mean Position of Unusual Meanings as a Function

	Frequency of Occurrence				
	1-2	3-4	5-6	10-14	22-30
Mean Position	6.38	6.20	5.48	6.06	5.06

position of unusual meanings as a function of frequency of occurrence (grouped in pairs to remove some irregularities). It is apparent that the unique meanings occur later than the more popular ones. The common-unusual meanings (22-30) occur much earlier in the response sequence than any other unusual meaning.

Figure 7 gives the cummulative percentages for the two types of unusual meanings. Note that over 50% of the common meanings are produced before the fourth solution. Whereas, for uncommon meanings the median position is 6.0. A Kolmogorov-Smirnov two-sample test for large, unequal n's shows that the two distributions are in fact different beyond the .01 level of confidence. Chi Square tests (df 9) confirm this analysis with a value of 25.26 for uncommon meanings ($p<.01$) and a value of 30.70 for common meanings ($p<.01$). This indicates that the majority of common meanings do come in the first portion of the response sequence and that the majority of uncommon meanings come in the last portion. Furthermore, the correlation between quality and unusual meanings (.635) is increased to .688 by not counting those solutions with common-unusual meanings.

These analyses justify separating the common-from the uncommon-unusual meanings. When common-unusual meanings are counted as not being unusual meanings, a single-factor analysis of variance for repeated measures on position gives an F of 5.50 (df 9, 261) which is significant beyond the .01

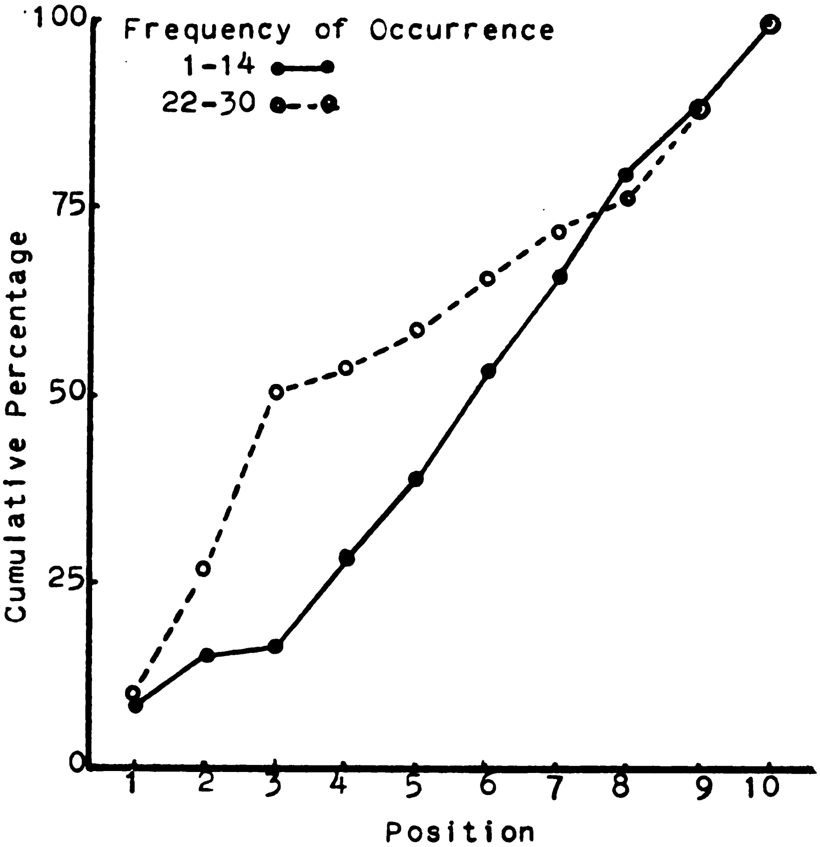

Fig. 7. Group TUM. Cumulative percentage of unusual meanings as a function of position.

level. Individual comparisons show that more unusual meanings are found in solutions 6, 7, and 8 than in solutions 1, 2, and 3 ($p<.05$). Other solutions do not differ. Thus, for Group TUM it may be concluded that the increase in quality over position is also accompanied by an increase in the use of unusual meanings.

Single-factor analyses of variance (df 9, 261) for repeated measures on position show changes in other response measures in Group TUM. Length generally increases ($F = 4.90$; $p<.01$), and topic freedom increases ($F = 1.96$; $p<.05$) as more solutions are produced. The use of word pairs decreases ($F = 2.20$; $p<.05$). Increases in solution quality for Group TUM, then, are accompanied by increases in the use of unusual meanings and in topic freedom and sentence length, and by decreases in the use of word pairs.

Table 18 presents the frequency of use of each unusual meaning. The unusual meanings which were included in the sentences Ss learned are under the heading "learned meanings." Those meanings which differed from the learned meanings are under the heading "created meanings." It is apparent that created meanings are used more frequently in solutions than learned ones (134 vs. 69). There are more different created meanings than learned ones (66 vs. 12). Also learned meanings are used earlier in the response sequence than created meanings (mean position 5.50 vs. 6.07).

Although Ss did not repeat the meanings which were learned, the meanings which were used did closely resemble the learned ones. Table 18 presents two striking similarities. (a) Frequency of usage in sentences corresponds to frequency of exposure during learning. (b) Created meanings often contain only simple changes of the learned meanings, e.g. from the learned Happy Horse Dude Ranch to Happy Horse Stables.

In summary, learning sentences with unusual meanings enables Ss to use more unusual meanings in sentences. Such a use increases the judged quality, and novel meanings, which occur later in the production sequence, increase quality even further. These combined effects produce a higher average quality and an increase in quality over the production sequence.

Intersentence Associations

This section will report an attempt to measure the strength of association between each S's solutions. After several attempts, a naive judge was able to classify the ways sentences could be similiar into five categories. If a theme was developed from sentence to sentence and more information was added, it was called a "develop" association. An example would be where a story is developed from when a person is thinking about going to camp, to what he does at camp, and finally to what he is thinking about as he comes home. This is a sort of thematic continuity. The second major type of association is "rework." Here the basic sentence remains the same, and the new sentence is only slightly reworded; no new information is added. For example, "the happy horse drank from the expensive lake" could be followed by "the happy horse jumped into the expensive lake." The remainder of the associations are based on the repetition of only part of the sentence. An association would be a "run on" if the last word or phrase of one sentence were the first word or phrase of the next sentence. "Repeat" is the simple repetition of a particular word or phrase, and "position" is a "repeat" in the same position.

The solutions in Groups NC10, TOP and TUM were reread, and the associations between an S's solutions were tallied. The first solution and incomplete solutions were not counted in any of the calculations. Table 20 gives the mean number of associations per S. Notice the large number of "rework" associations as opposed to "develop" associations. From this information it can be concluded that there is some degree of association between some solutions, but not between all solutions. The only difference between groups is that Group TOP solutions had more "rework" associations.

The next question is whether an association between sentences helps the quality of the second sentence. Table 21 gives the correlations between associations and quality for solutions in each group. "Rework number" and "develop number" are the number of sentences to which the present

TABLE 20.--Mean Number of Associations per Subject.

Association	Group NC10	Group TOP	Group TUM
Rework	2.92	4.27	2.59
Develop	.66	.84	1.15
Run on	2.25	1.79	1.95
Repeat	8.00	7.48	6.90
Position	9.14	8.74	8.39
Total	22.79	23.50	21.12

TABLE 21.--Correlations between Solution Quality and Associations.

Association	Group NC10	Group TOP	Group TUM
Rework	-.21	-.30	-.17
Rework number	-.16	-.22	-.16
Develop	-.03	.10	.02
Develop number	-.02	.13	.08
Run on	-.04	-.10	-.04
Repeat	-.16	-.30	.05
Position	-.18	-.25	-.24
Total	-.30	-.38	-.16

sentence relates. For example, if a sentence is the fifth in a series of sentences which all relate to the same theme, the "develop number" value will be "5." If quality is correlated to associations, one would also want to know if quality is also a function of the length of the associative chain. With n = 300 a correlation of .113 is significant at the .05 level. Because solutions are the object of this analysis, covariance due to subject abilities cannot be removed, nor would it be desirable to do so. For Groups NC10, TOP and TUM the correlations are generally low and negative. Apparently the more associations a sentence has with the preceeding one, the lower the quality will be. This will be especially true if the sentence has been changed only slightly, i.e. is a "rework" association.

One could also ask if these correlations would be changed were only sentences with at least one association counted. The results are changed in two ways. For "rework" and "develop" associations the correlations are increased. Thus, when a sentence has been reworked from previous sentences, more related sentences will give lower quality (median r = -.19). When a sentence has been developed along a theme, more related sentences will increase quality (median 4 = .18).

Other results have suggested that length increases as more solutions are produced because of the same sort of response chaining; i.e., building from one solution to another.

Correlations between intersentence associations and solution length are generally low and negative for Group NC10 (range -.10 to -.01). Group TOP correlations are somewhat higher and negative (range -.46 to .08). Correlations are highest for rework, (-.26), repeat (-.37) and position (-.31). Group TUM correlations reflect the same conclusions (range -.16 to .04) with the only significant correlation being for repeat (-.16).

For the type of intersentence associations that were measured, if there is a relationship to solution quality or length at all, it is a negative one. In terms of solution quality there appears to be no advantage in reworking a sentence or in building a theme from one sentence to another.

CHAPTER IV

DISCUSSION

The rationale for the present study may be summarized in the following way. There is a need to understand the processes which enable a person to produce good solutions to a problem. Good solutions may be produced as a result of instructions or training, but such devices have not successfully focused on problem solving processes. A person may also produce good solutions after several less successful tries. The literature indicates that the process involved in producing poor-then-good solutions may be based on breaking up the elements of the problem and rearranging them into new forms. The sentence problem is a complex productive thinking problem, the solutions of which can be analyzed into several response measures which could reflect this process. The divergent process, of course, is assumed to operate in different types of problems as well. The first level of analysis focuses upon the correlation of solution quality and the response measures. The second level of analysis is experimental and involves changing the characteristics of the words used in the sentence problem and noting changes in the response measures. In this way solution processes may be illustrated for the sentence problem with four specific words and with slightly different words.

The results of the present experiment generally point to a divergent problem solving process; i.e., good solutions depend upon being different from what one would expect and, the more solutions one produces, the more different are the solutions from the usual first response. Correlational analyses show that divergent response measures are highly correlated with quality. The four original words were selected only for being two adjectives and two nouns. Casual observation will show that the adjectives and nouns are only moderately associated and the nouns have few unusual meanings. Making the word pairs more highly associated tends to make responses more stereotyped, hence solutions are of poorer quality. As more solutions are produced, however, solutions become better as a result of breaking up the word pairs. Giving the words unusual meanings tends to make responses less stereotyped by allowing subjects to use unusual meanings, hence solutions are of higher quality. As more solutions are produced, however, even more unusual meanings are created and solution quality increases even more. The following discussion will present detailed explanations of each finding.

Divergent Processes Under Normal Conditions

The production of solutions in Groups NC10 and NC5 may be considered "normal" in two ways. First, the instructions were standard instructions giving no cues to desired performance other than the number of solutions desired. Second, the words used were not selected for any special reason. As a result two adjectives and two nouns may as well have been randomly selected from a dictionary. The two measures of importance for this study seem to have moderate values for these words. The associability of the word pairs does not appear to be very strong. Happy horse or expensive lake certainly are not as strongly associated as red barn or yellow canary. Thus, there would be no strong need to use the given words together in a sentence. Furthermore, neither horse or lake have many unusual meanings. Because of their contemporary usage, words such as ball or cat would have more unusual, and clever, meanings which would be used to better solution quality.

The principal questions under investigation are (a) does quality increase over the response sequence, and (b) how do variables which reflect learning change as more responses are produced.

When producing five or ten solutions, solution quality increases as more solutions are produced, and the increase is linear. These results are in agreement with those obtained with less complex problems. Research with the unusual uses problem (e.g., Christensen, et al., Gerlach, et al. and Turner) show that rated creativity of uses increases over the response sequence, later uses are statistically more novel, and more novel categories of uses occur in later solutions. These results have been interpreted to indicate a divergent process which includes breaking up old associations between object and uses and the construction of new ones. The question is whether the same analysis can be applied to the sentence problem.

Response measures, other than length and complexity represent measures of the degree to which part of a sentence corresponds to (or is associated with) the strongest or most dominant response. For example, unusual meanings reflect a breaking away from the conventional meanings. Quality is correlated to unusual meanings, word pairs (negatively), topic freedom and length and complexity. Regression analyses show that quality may be most accurately predicted from unusual meanings, word pairs and length. Thus, the importance of two divergence measures is verified.

The analyses of variance over position represent another form of correlational analysis. In this case the hypothesized relationships take the form of a multiple correlation. If quality is correlated to position, is a variable which is correlated with quality also related to position? And, like the correlational analyses, this analysis is more suggestive than convincing. Graphically fluctuations in solution quality appear to be matched by changes in unusual meanings, topic freedom, length and complexity. The use of word pairs appears to be inversely related to quality in Group NC5. Statistical analyses, however, weigh heavily the subject variance and show that only solution quality and length change over position. The use of unusual meanings increases slightly.

To explain the increase in quality several explanations may be invoked. (a) Warm-up is evident in most extended exercises. Perhaps it takes a few solutions to find out enough about the problem and its solutions to produce a good solution. (b) The first few solutions could be hurried efforts which are later revised and made, for example, more grammatically correct. (c) Initial solutions could be more similar to one another, hence are more common and of lower quality. Later solutions are less alike and different from the initial solutions, hence are less stereotyped and of higher quality. Or (d) initial solutions represent the response which is most strongly associated to the problem. Later solutions have a lower associative strength, because they are further down the associative hierarchy.

It is apparent that additional data will be needed to discriminate between these possible explanations. The following considerations led to the treatment conditions represented by Groups TOP and TUM. Solution quality is determined by several variables. The exact relationship of any one variable to quality is obscured by the fact that there are many variables correlated to quality and that most of them are intercorrelated. For example, unusual meanings frequently use old pairs, such as Horse Lake. This obscures the relationship between word pairs and quality by making the correlation more positive. Furthermore, the relationship between position and quality is so small that it would be difficult to show that a quality-related variable had the same relationship to position as quality. The treatment conditions represent an effort to resolve this problem by increasing the importance of a variable in determining quality so that changes in quality would be accompanied by changes in this variable. The influence of a variable would be increased if the frequency of its occurrence were increased and the use of confounding variables decreased.

Since the words happy, expensive, horse and lake essentially represent moderate associability and meaningfulness values, it is reasonable to ask what would happen if they had more extreme values. By increasing the frequency of occurrence of word pairs and unusual meanings the extreme values are better approximated.

Lastly, *S*s differ greatly in the frequency of use of word pairs and unusual meanings. It was felt that the observed relationships would be more evident if *S*s were at about the same level to begin with and were free to become different as more solutions were produced.

Old Pairs

Learning sentences with old pairs alleviated some of the problems with the normal case. First, the variance between Ss for total word pairs used and for quality was reduced below that of Group NC10 for the first solutions. Thus, Ss began at about the same point. Second, the associative strengths for word pairs were increased such that word pairs were used more frequently and hierarchy positions were adjusted in line with the learning conditions. Third, by the increase in the use of word pairs and the decrease in the use of unusual meanings, the correlation between word pairs and quality was increased greatly.

It was hypothesized that learning sentences with old pairs would (a) increase the mean frequency of old pairs in solutions, (b) increase the initial use of old pairs, but that use would decrease over position, and (c) allow for more of an increase in quality over position. These hypotheses were confirmed and will be discussed in order.

When associative hierarchies are assessed by the method of continued association, associative strength is inferred from the average position in the response sequence and the frequency of occurrence of a particular response. Taking Group NC10 as an example, the relative associative strengths of word pairs may be inferred by the average position of occurrence in the response sequence, and frequency of occurrence agrees almost perfectly. The associative hierarchy for five word pairs may be said to go from strongest to weakest: expensive horse, expensive lake, happy lake and horse lake.

Why should such an associative hierarchy exist at all for the normal case? Or why should Ss use word pairs in sentences? Without adequate associative norms one can only speculate as to the cause or the strength of associative links. First, the motivation to use word pairs could come from the fact that it makes the problem easier by using two words in one thought or language unit. Secondly, conceptually it makes sense to write about an expensive horse, more so than a happy lake. Furthermore, one has encountered more expensive horses than happy lakes. Thus, one could attribute the motivation to initially use word pairs to Ss, and account for the order of their use by established language habits.

After learning sentences with old pairs, Ss in Group TOP used vastly more old pairs and other word pairs than Ss in any other group. And the learned old pairs shifted to the top of the associative hierarchy. It is not surprising that Ss learned old pairs or that they used them in solutions, but it is somewhat surprising that they were used earlier in the response sequence that other word pairs in their own solutions and earlier than word pairs in other groups. Voss (1968) assessed individual Ss' associative hierarchies for eight words, then changed the hierarchies with paired-associate learning of the stimulus word and low level response words. A second free association session verified the hierarchy shifts. Such a shift persists up to 48 hours, although slight attenuation does occur (Bruder, 1968). These studies used associative hierarchies as assessed for individual Ss. McConkie (1969) obtained similar results using the Minnesota and Connecticut free association norms. By learning sentences with old pairs, then, the specific word pairs accumulate a higher associative strength between the member words. As a result, when the words are recalled for a solution, the pairs with stronger associative links are emitted first. Apparently the learning was sufficient to override normal language habits so that more word pairs were used and old pairs were used prior to "natural" pairs in the solutions of Group TOP and prior to "natural" pairs in other groups.

A more difficult finding to explain is the word pairs in Group TR. This group used word pairs more frequently than Group NC10 Ss, but Group TR Ss learned relevant sentences without word pairs. A second result of importance is that the associative hierarchy is the same as the normal case, except for the reversal of the weakest members.

Several investigators have found paired-associate learning to be facilitated by pre-exposure to contextual phrases (Epstein, Rock and Zuckerman, 1960), by pre-exposure to meaningful syntactically structured verbal strings (Rohwer, 1966) or by instructions to form sentences on the first study trial (Jensen and Rohwer, 1965). These investigators used response learning and verbal mediation to explain the facilitation, and they point out several characteristics of the material which influence the amount of facilitation.

In general, the sentences as used in the present experiment would provide the situation for the greatest facilitation in subsequent learning. To some degree response learning must be involved, especially since Ss also learned and subsequently used only conventional meanings for the given words. Secondly one would expect some sort of special effort on Ss' part to learn the only words common to all sentences. If this were the case, it would not be surprising if the words were memorized according to pre-existing language habits; i.e., word pairs that were familiar or were meaningful were memorized first. Also, in the first two sentences the phrase "the horse was happy" was included, and in several sentences expensive and lake were connected by prepositions or things around a lake were labeled expensive. Rohwer (1966) found verb connectors facilitated learning the most and prepositions were almost as effective. In this way, the given words would be salient and would be stored according to existing language habits to be recalled later in solving the problem.

It would be interesting to see if syntactic facilitation in the Jensen and Rohwer paradigm would be influenced by the associative strength of pairs prior to learning sentences. In other words, if a noun was learned in a sentence with one highly associated adjective and one lowly associated adjective, would the increment in associative strength be only in the highly associated pair? The present results indicate that naturally occurring (or strong) associations would be strengthened first.

The data show that the majority of word pairs are in solutions of Groups TOP and TR. In each of these groups the number of word pairs decreases over the response sequence? A simple explanation would suggest that Ss simply forgot the word pairs. This is doubtful in 15 minutes. Another explanation would suggest extinction or nonreinforcement as a reason for suppression of the response, even though the pairs may still be available. This is possible since sentences using word pairs are commonly very poor and unexciting. As these data show Ss then use the next strongest associated pair, then the next, until they use different pairs entirely. Furthermore, as in free association trials, Ss do not repeat solutions, so they are forced to move to different word combinations. Usually one can also note thematic shifts between sentences, just as if Ss were instructed not to repeat the same idea.

It is one thing to describe associative hierarchies and shifts in them, but it is another thing entirely to use associative hierarchies to explain changes in solution quality. In Group NC10 quality increased over the response sequence in a linear fashion. But the use of word pairs did not change significantly, and the correlation between word pairs and quality was very low. One could argue that the associative strengths between members of those word pairs were not high enough to demonstrate a decrease over ten solutions. This would be true especially if one considered all the possible adjectives which could be used in pairs, such as <u>brown horse</u> or <u>smelly horse</u>. Surely some would be stronger associates than <u>expensive horse</u>. Furthermore, the lack of a correlation between solution quality and word pairs could be due to the confounding effect of unusual meanings which were frequently word pairs. Whereas the use of word pairs alone will reduce the quality ratings, unusual meanings, even with word pairs, will increase quality. Thus, a correlation would include these canceling effects.

On the other hand, solutions in Groups TOP and TR contain more word pairs and far less unusual meanings than solutions in other groups. As a result there is a strong negative correlation between the use of word pairs and solution quality. Group TOP represents an almost ideal case. With the increased use of word pairs the mean solution quality is lower than the quality in any other group. Furthermore, as the use of word pairs decreases over the response sequence, the solution quality increases. The correlation between quality and number of word pairs was -.87. Learning sentences with old pairs increases the associative strength between the words such that more word pairs are used in solutions. After using the pairs with the highest associative strength, Ss move down the associative hierarchy to other word pairs. Sentence length and complexity and word pairs are the only variables which significantly change over the response sequence. The additive influence of increased length, complexity and decreased use of word pairs may be said to cause the increase in solution quality in Group TOP.

It is possible to speculate now on the relationship between the associative structure of thought and the length and structure of sentences. What I would like to propose is that the length and complexity of sentences can be, although does not necessarily have to be, determined by the associative nature of the ideas comprising the completed sentence. In the case of two weakly associated ideas (or words), it would take a long, complex sentence to link the two. Often the sentence would even specify the relationship between the ideas. For example, one may wish to say, "The horse was sticky." Since "sticky" and "horse" are weakly associated ideas, one may wish to add a qualifying phrase to the statement such as, "the horse which had just been doused with glue was sticky." The latter sentence is longer and grammatically more complex. On the other hand, highly associated ideas could be used sensibly in shorter, less complex sentences. Without a doubt the associability of ideas does not account for a large proportion of the variance in length and structure. But it may be the case that in the present experiment increasing associability also reduces the need to make long sentences.

Solutions in Group TR present somewhat of a challenge to this interpretation. In this group the decrease in the use of word pairs and the increase in sentence length over the response sequence is as great as in Group TOP. But the increase in quality is not as great. It only approaches significance (P = .06). Perhaps this may be adequately explained as a ceiling effect. By not using as many word pairs as Group TOP (about 30% less), solutions rapidly increased in quality until they could get no better without an added boost which unusual meanings could give. But Group TR Ss used less unusual meanings than Groups NC10 and TIR. The data show that the use of word pairs decreases most in the first four solutions of Group TR, after which as many word pairs are used as in Groups NC10 and TIR. Solution quality increases significantly for the first five solutions, then levels off. It is possible, based on the correlations with quality, that quality could get no higher without the use of unusual meanings. With the last five solutions being of equal quality, analysis of variance would show no significant differences because of the disproportionate accumulation of within-position variance relative to the between-position variance.

As an experimental treatment learning sentences with old pairs has some disadvantages. Group TOP solutions, on the average, were different from those in any other group by being of lower quality and using more word pairs. They were also shorter and had less topic freedom. Apparently these differences were unique to _S_s who learned sentences with old pairs, even though the only differences between the learned sentences was the word pairs, not length. Had _S_s not learned sentences, perhaps these confounding results would have been eliminated. One way to do this would be to have _S_s learn paired-associate lists. The experimental words could be included with several control items. Voss (1968) and others have used this method effectively. Actually the real question is whether highly associated pairs are used more frequently and earlier than pairs with lower associative strength. This being the case one could simply compare high and low associates such as _red barn_ and _alone noise_. Because of the lack of adequate association norms, one could scale pairs by Kammann's (1968) associability rating method; i.e., ask _S_s to rate, 1 to 10, how likely it would be that a given pair of words would be used together in a sentence. Other _S_s would be asked to write ten solutions to the sentence problem, and different _S_s could use each set of four words. Since quality has already been shown to be determined, at least in part, by the use of word pairs over position. One could also use the beta weights obtained in the present experiment to estimate what quality ratings would have been had the present judges been used again. This would save paying the judges.

Unusual Meanings

It was hypothesized that learning unusual meanings (a) would increase the use of unusual meanings, but (b) that use would not increase over the response sequence. Thus, it was expected (c) that mean quality would be higher for that group, but (d) that quality would not increase over the response sequence. The data provide partial confirmation of these expectations.

The data follow the results with word pairs almost exactly. Learning unusual meanings increases their use in subsequent solutions and increases mean quality. The obtained unusual meanings, like word pairs, follow an order of emission which goes from conventional meanings to idiosyncratic meanings. The number of unusual meanings increases over the response sequence and solution quality increases as well. Thus, the interpretation of this data will closely follow that for word pairs, with the exception that unusual meanings serve to increase quality.

The term associative strength refers to the position of a particular response in a response series for an individual S, or the frequency of occurrence of a response in a group sample. The data for unusual meanings in Group TUM, like word pairs, shows that common meanings are given before unique meanings in the response sequence. Thus, one could say that the associative strength between a word and a conventional meaning is greater than between the word and a unique meaning. The quality rating is also a function of the commonness of the meaning. Conventional meanings are rated lowest, common unusual meanings are rated somewhat higher, but the highest rating comes from unique unusual meanings. As a rule conventional meanings are given first, then learned or common unusual meanings, and finally unique unusual meanings.

"Meaning" as used in the present experiment has two meanings. The first refers to the conventional dictionary, or denotative, meaning. The degree of commonness varies from the usual horse as a "four-legged mammal" to the unusual "slang for heroin." A second form of meaning is the unusual use or function of a word. Examples would be to use happy as a person's name or to use a noun as a modifier in a name as in Lake Estates.

From the list of obtained unusual meanings in Group TUM it appears that there are more of the latter type of meaning. The high correlation between quality and unusual meanings indicates that the judges thought both types of meaning were clever and that both types of meaning were better than conventional meanings. It is at this point that the use of "associative hierarchy" and "associative strength" as used with word pairs seem somewhat inappropriate. Perhaps this is because the established language habits which result from word pair hierarchies are intuitively recognizable. With word pairs a hierarchy can be based on the frequency with which words appear contiguously in everyday speech. These concepts, however, should apply to potential associations as long as there is some probability that the response will occur. Furthermore, the data show similar response patterns in word pairs and in unusual meanings.

The unusual uses problem presents a similar situation. In this problem Ss are required to write several uses for a common object such as a brick. In general, uses which are familiar and can be readily recalled are rapidly produced. After the familiar uses are exhausted, however, more unusual uses are produced at a slower rate. Thus, obtained latency curves negatively accelerate and uncommonness increases over the response sequence. These results have been interpreted (e.g., Christensen, et al., 1957) to indicate that less common uses have a lower associative strength to the object, and it usually takes several responses before one works down the associative hierarchy to unique uses.

Turner's (1967) data show the progression from common to uncommon uses as a divergence from the uses and properties one usually associates with the given object. He was able to classify the uses of a brick, for example. The first responses were "engineered uses," or what a brick was designed to do, e.g., to use in construction. These are the uses with the greatest associative strength. Later the responses used the physical properties usually attributed to the object, "compositional uses," for example to use a brick as a paperweight. The most uncommon responses, and the ones to occur latest in the production sequence, were "shape-quality uses" where the use was not based on the concrete limitations of the original object. Commonly these would be transformations of the object, for example to crush a brick for a filter in a moonshine still. In other words, the solutions followed an orderly sequence of getting further and further away from the object as it is commonly used or perceived.

The unusual meanings obtained in the present experiment follow a similar pattern. The first meanings which are usually given are conventional ones. If <u>S</u>s learned sentences with unusual meanings, conventional meanings may be replaced with unusual meanings which were learned in the sentences. The unusual meanings which occurred later in the response sequence represented transformations of the learned meanings. For example, <u>Lake</u> <u>Park</u> was learned and reflected in <u>Horse</u> <u>Lake</u>, <u>Lake</u> <u>Happy</u>, <u>Lake</u> <u>Happy</u> <u>Horse</u>, which were progressively less common. The transformed meanings could have arisen, because <u>S</u>s could not correctly recall the learned example.

The infrequent use of most learned meanings seems to indicate this, rather than the knowledgeable transformation of learned meanings. As to why the use of unusual meanings increases over the response sequence, the best explanation is that they should appear as clever to Ss as they do to judges. As a result they should try to recall the ones they learned or to transform the ones they have already used. In any event, unusual meanings should be more reinforcing than conventional meanings.

The validity of describing the sequence of unusual meanings as an associative hierarchy may also be seen in the solutions of Group TR Ss. In this group Ss learned sentences with conventional meanings for the words to be used later in solving the sentence problem. As a result this group used less unusual meanings than Groups NC10 and TIR which had no interferring learning. Furthermore, the frequency of unusual meanings did not increase over the response sequence. It seems that learning these sentences limited the associative hierarchy to a few very dominant responses. As a result mean quality is slightly lower than other groups.

The connection between solution quality and the use of unusual meanings is rather straightforward. In every condition quality is highly correlated with the use of unusual meanings. The experimental treatment was able to increase the mean frequency of unusual meanings and to increase solution quality. Regression analyses consistently showed that unusual meanings predicted quality. Furthermore, as quality increased over the response sequence, the use of unusual meanings and their uniqueness increased.

In Group TOP the relationship between quality and word pairs was somewhat confounded by concomitant increases in sentence length and complexity. In Group TUM there are also increases in sentence length, topic freedom and decreases in word pairs. In the latter group, however, the use of unusual meanings contributes to solution quality so disproportionately that it must be concluded that the increase in quality is almost strictly due to changes in the use of unusual meanings.

As an experimental treatment learning sentences with unusual meanings was able to produce rather large differences due to the importance of unusual meanings in determining solution quality. On other measures Group TUM used as many word pairs as Group TR, and Group TUM solutions had more topic freedom than other groups. Neither of these variables confounded the results. If this treatment were to be redone without learning sentences, the methods mentioned for Group TOP could be suggested. Paired-associate lists would be unable to convey a meaning, because they lack the context of a sentence. On the other hand, words could be rated for meaningfulness by Kammann's (1968) method. Ss would be asked how many ways a word could be used in a sentence. Using groups of high and low rated words, groups of Ss could produce ten solutions to the sentence problem. The data would be simply how many unusual meanings were used for each group and whether the use increased over the response sequence.

Instructional Effects

It was initially hoped that instructing Ss to produce five or ten solutions might illustrate some differences in solution processes. Under standard instructions there were no differences observed between Groups NC10 (ten solutions) and Group NC5 (five solutions). Mean values based on all solutions were identical for all response measures. Comparing just the first five solutions of each group produced no significant differences either. In fact, for quality the regression lines were almost identical. In all respects it appears that under standard instructions the first five solutions are alike whether produced as the whole task or as half of the task.

Using simpler problems other researchers have found that higher productivity is associated with more high quality solutions, but also lower mean quality (e.g., Gerlach, et al., 1964). Johnson, Parrott and Stratton (1968) found, on the average, that producing one solution was better than producing many and that mean quality correlated negatively with the number of solutions produced. In a more controlled study Johnson (1968) found that this is, in part, dependent upon how many good solutions are possible. With the conclusions problem there are a few perfect solutions (about 10). If Ss are restricted to produce 1, 2, 4, or 6 solutions, mean quality decreases as a logarithmic function of the number of solutions produced. What appears to be happening with this problem is that Ss persist with a single way of looking at the data. By approaching the data from several different perspectives better solutions could be obtained. In the sentence problem the biggest decrease in quality was between Ss producing one solution and Ss producing more than one. With this problem there are many more good solutions possible.

The present results with the sentence problem extend these data to the five and ten solution case. Although five solutions have a slightly higher mean quality, the difference is not statistically significant for standard instructions.

Criteria-cued instructions produce a significant decrease in mean quality from five to ten solutions. These solutions, however, do not differ on any other response measure. Ss producing five solutions were given ten minutes, and Ss producing ten solutions were given fifteen minutes to finish. It is possible that with more time per solution five-solution Ss were better able to use the criteria-cued instructions. Another possibility is that both ten-solution groups show no increase in quality after the first few solutions. This could be due to fatigue or lack of time, as

Ss were informed of the time after the session was half over. Or Ss with criteria-cued instructions could just be more conscientious for the first few solutions.

When the first five solutions are compared, there are no differences between producing five or ten solutions. For quality the first three solutions of Group CC10 lie exactly on the regression line for Group CC5. All curves match each other embarrassingly well.

Comparing these groups on response measures over the response sequence gives some insight into why quality increases over the response sequence. Whereas Group CC10 does not increase in quality, Group CC5 does. The increase in quality for Group CC5 is uniquely accompanied by significant increases in unusual meanings and decreases in the use of word pairs. The importance of these variables was verified by Groups TOP and TUM.

The comparison of mean quality between groups with and without criteria-cued instructions shows criteria-cued instructions did not increase mean quality. Furthermore, quantity instructions did not affect quality. The four groups did not differ on any other response measure either.

Every study reviewed and the Johnson, et al., (1968) studies have shown that criteria-cued instructions are a most influential variable. Why did they have no facilitating effect in the present experiment? One possibility would be that the criteria-cued instructions did not accurately reflect the judges' criteria. The data show that the present judges did weigh response measures differently than the original judges. In fact, the original judges used the criteria-cued instructions as a basis for the first Rating Guide. This could easily be tested with different instructions.

By specifying the number of solutions to be produced, the effects of criteria-cued instructions could have been reduced. Other studies have allowed Ss to produce as many, or as few, solutions as they desired, and under criteria-cued instructions usually fewer solutions are produced. Specifying the number of solutions to be produced could emphasize quantity to the extent that quality will be sacrificed and the criteria-cued instructions ignored. It could be that fifteen minutes was not sufficient time to produce ten solutions that met the criteria (at least in Ss' mind this may be true). The result of emphasizing quantity and providing inadequate criteria would be to reduce the average quality level to that of the standard instructions. This could be tested by intentionally varying the degree of accuracy, or adequacy, of criteria-cued instructions under fixed or undefined quantity instructions. In this way the quality-quantity tradeoff could be better specified.

CHAPTER V

SUMMARY

In an effort to investigate processes involved in productive thinking, this experiment tested the response hierarchy theory of problem solving with a problem requiring subjects to write a series of sentences. This theory predicts that under free responding instructions, as more solutions are produced, the solutions will be less like initial solutions and will be judged to be of higher quality, i.e. more clever, etc. The quality dimension will be based on criteria which reflect prior learning. Since the sentence problem requires subjects to write many sentences which include four given words, there could be associative hierarchies involving pairs of the given words and word meanings. The theory would predict that strong adjective-noun associations and popular word meanings would appear early in a series of solutions. Remote associations and unusual meanings would appear later and should be correlated with high quality. Under instructions which restrict responding by specifying criteria for high quality solutions, early solutions should be the same as later solutions in all respects, i.e. there should be no response hierarchy.

The present experiment used two naive judges to evaluate quality on a scale from one (low) to seven (high). Interjudge reliability was .86. With criteria-cued instructions and noncriteria-cued instructions regression analyses indicated that quality could be predicated by the number of word pairs (negatively correlated), unusual meanings, and sentence length. Quality increased over the response sequence for noncriteria-cued instructions when subjects wrote five or ten solutions, but not with criteria-cued instructions and ten solutions. Average quality was not significantly affected by quantity or criteria-cued instructions.

The experimental treatments involved subjects learning six sentences prior to producing solutions. These sentences included word pairs, unusual meanings, relevant words without pairs or unusual meanings, or irrelevant words. Results for subjects learning irrelevant sentences or none at all were identical. Learning word pairs increased the associative strength between the learned adjective-noun pairs such that (a) more word pairs were used, (b) learned pairs shifted upwards in the associative hierarchy such that they were given more frequently and earlier in the response sequence, (c) sentence length decreased, and (d) mean solution quality was lower. Learning sentences with relevant, but unpaired, words increased the use of word pairs, but relative position of pairs within the associative hierarchy did not change from that shown by control conditions. Also both conditions increased the associative strength between the words and their most common meanings, so that less unusual meanings were used. In both conditions quality increased as more remote associations were used. Learning sentences with unusual meanings also increased the number of word pairs used, but the nouns acquired more new and unusual meanings. As more solutions were produced, unusual meanings became more unique, hence quality increased over the response sequence. Mean quality for this condition was higher than for other conditions, because unusual meanings were more highly weighed in judging quality than other variables.

These data were interpreted to be consistent with the response hierarchy theory and the general view that creativity involves a process of breaking up old associations and forming new ones, i.e. a divergent process. To the extent that externally or internally produced instructions provide cues to high quality solutions, the response hierarchy may be by-passed or may remain covert. Then no low quality solutions will be recorded, and no response hierarchy will be apparent in the solutions.

BIBLIOGRAPHY

BIBLIOGRAPHY

Armstrong, J. S. and Soelberg, P. On the interpretation of factor analysis. Phychological Bulletin, 1968, 70, pp. 361-364.

Bain, A. The senses and the intellect. 1855.

Binet, A. The psychology of reasoning. Chicago: Open Court, 1899.

Boneau, C. A. The effects of violations of assumptions underlying the t test. Psychological Bulletin, 1960, 57, pp. 49-64.

Bruder, G. A. Shifts in associative hierarchies over time as a function of paired-associate training. Journal of Verbal Learning and Verbal Behavior, 1968, 7, pp. 413-416.

Cattell, R. B. Factor analysis: an introduction to essentials: I. The purpose and underlying models. Biometrics, 1965, 21, pp. 190-215, 405-435.

Christensen, P. R., Guilford, J. P., and Wilson, R. C. Relations of creative responses to working time and instructions. Journal of Experimental Psychology, 1957, 52, pp. 82-88.

Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, pp. 161-182.

Dominowski, R. L. Problem difficulty as a function of relative frequency of correct responses. Psychonomic Science, 1965, 3, pp. 4128-4129.

Dominowski, R. L. Anagram solving as a function of bigram rank and word frequency. Journal of Experimental Psychology, 1967, 75, pp. 299-306.

Duncan, C. P. Problem solving within a verbal response hierarchy. Psychonomic Science, 1966a, 4, pp. 147-148.

Duncan, C. P. Response hierarchies in problem solving. Presidential Address to Midwestern Psychological Associations, May, 1966, Chicago. In Duncan, C. P. (Ed.), Thinking: current experimental studies. NYC: Lippincott, 1967, pp. 18-44.

Epstein, W., Rock, I. and Zuckerman, c. B. Meaning and familiarity in associative learning. Psychological Monograph, 1960, 74, No. 4., (Whole No. 491).

Freud, S. Wit and its relations to the unconscious. In, The Basic Writings of Sigmund Freud. New York: Modern Library, 1938.

Gerlach, V. S., Schutz, R. E., Baker, R. L. and Mazer, G. E. Effects of variations in test directions on originality test responses. Journal of Educational Psychology, 1964, 55, pp. 79-83.

Golann, S. E. Psychological study of creativity. Psychological Bulletin, 1963, 60, pp. 548-565.

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

Hollingworth, H. L. Psychology: its facts and principles. New York: Appleton, 1928.

Humphrey, G. Thinking: an introduction to its experimental psychology. New York: Wiley, 1951.

Jensen, A. R. and Rohwer, W. D. Syntactical mediation of serial and paired-associate learning as a function of age. Child Development, 1965, 36, pp. 601-608.

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

Johnson, D. M. Improvement of problem solving processes. Final Report to USOE, Project No. 5-0705., Dec., 1968.

Johnson, D. M., Parrott, G. L. and Stratton, R. P. Production and judgment of solutions to five problems. Journal of Educational Psychology, 1968, 59, Monograph Supplement.

Johnson, P. E. and Collier, R. O. A note on methods of indexing associative relatedness. Behavior Research Methods and Instrumentation, 1969, 1, pp. 140-142.

Johnson, T. J. and Van Mondfrans, A. P. Order of solutions in ambiguous anagrams as a function of word frequency of the solution words. Psychonomic Science, 1965, 3, pp. 565-566.

Kammann, R. Associability: a study of the properties of associative ratings and the role of association in word-word learning. Journal of Experimental Psychology Monograph, 1968, 78, No. 4.

Locke, J. Essay concerning the human understanding. 1969.

Maier, N. R. F., Julius, M., and Thurber, J. A. Studies in creativity; individual differences in the storing and utilization of information. American Journal of Psychology, 1967, 80, pp. 492-519.

Maltzman, I. Thinking: from a behavioristic point of view. Psychological Review, 1955, 62, pp. 275-286.

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

Maltzman, I. Motivation and direction in thinking, Psychological Bulletin, 1962, 59, pp. 457-467.

Manske, M. E. and Davis, G. A. Effects of simple instructional biases upon performance in the unusual uses test. Journal of General Psychology, 1968, 79, pp. 25-33.

Mayzner, M. S. and Tresselt, M. E. Anagram solution times: a function of letter order and word frequency. Journal of Experimental Psychology, 1958, 56, pp. 376-379.

Mednick, M. T. Research creativity in psychology graduate students. Journal of Consulting Psychology, 1963, 27, pp. 265-276.

Mednick, Mt., Mednick, S. A., and Jung, C. C. Continual association as a function of level of creativity and type of verbal stimulus. Journal of Abnormal and Social Psychology, 1964, 69, pp. 511-515.

Mednick, S. A. The associative basis of the creative process. Psychological Review, 1962, 69, pp. 220-232.

McConkie, G. W. Response hierarchy changes resulting from learning additional associations. Journal of Experimental Psychology, 1969, 79, pp. 495-503.

Osgood, C. E. Method and theory in experimental psychology. Oxford: Oxford University Press, 1953.

Parnes, S. J. Effects of extended effort in creative problem solving. Journal of Educational Psychology, 1961, 52, pp. 117-122.

Ridley, D. R. and Birney, R. C. Effects of training procedures on creativity test scores. Journal of Educational Psychology, 1967, 58, pp. 158-164.

Rohwer, W. D. Constraint, syntax and meaning in paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1966, 5, pp. 541-547.

Turner, T. B. The creative process. Journal of Creative Behavior, 1967, 2, pp. 63-70.

Voss, J. F. Shifts in associative hierarchies as a function of associative training. Journal of Verbal Learning and Verbal Behavior, 1968, 7, pp. 417-420.

Yngve, V. H. A model and an hypothesis for language structure. Proceedings: American Philosophical Society, 1960, 104, pp. 444-466.

APPENDICES

APPENDIX A

APPENDIX A

A QUANTITATIVE DESCRIPTION OF RESPONSES TO THE SENTENCE PROBLEM

The purpose of this study was to determine the relative contribution of a wide variety of descriptive variables to the quality of a solution to the sentence problem. Based on the introductory discussion and upon the Rating Guide, 16 variables were derived which appeared to quantitatively describe how one sentence could differ from another. This number was reduced to the nine used in this study after correlations showed some variables to be repetitious or not at all correlated to solution quality.

The solution scores on these measures were factor analyzed to indicate the underlying relationships between the measures. By this method several independent measures were found, and their relative contribution to solution quality was ascertained. Furthermore, the relative contribution of position in the response sequence to quality and to other quality-related variables was ascertained.

Method

Subjects

The S were 40 students of Introductory Psychology at Michigan State University. They were tested during the Fall term of 1965 as part of a larger experiment on productive thinking.

Materials

From the several problems completed by each S the sentence problem was selected for analysis in the present study, because it was most amenable to quantitative description. The instructions for this problem were:

> Below are four words. Your task is to write sentences using all four of these words. Write as many sentences using all four words as you can in the time allotted.

The four words were: happy, expensive, horse and lake.

Procedure

As Ss entered the lecture hall for class the booklets containing all problems were distributed, and Ss were instructed that this would be an experiment and not to open the booklets. On signal everyone worked for seven minutes on each problem, and on completion the the booklets were collected. For each problem Ss wrote as many, or as few, solutions as they desired.

Results

The obtained 215 complete solutions were typed, coded, and independently judged for quality by two judges on a scale from 1 (low) to 7 (high). Interjudge agreement was .89, which is acceptable. The basic criteria were "the sentence should contain all four words, be well-constructed, and the given words do not appear obtrusive.

The following response measures were used in this analysis and defined in the previous study: quality, word pairs, unusual meanings, topic freedom, complexity and length. The following response measures were only used in the present study. Word moves--number of moves from the word order as given in the problem (range of possible scores = 0-3). Number of responses--number of solutions written by S in seven minutes (2-20). Position--ordinal position of that solution in the whole response sequence divided by the quantity, number produced + 1 (.15 - .99). Each solution was scored on each variable.

The solutions were initially factorized using the Factor A program for orthogonal factors and secondly by the Factor C program for oblique factors. Both programs were furnished by the Michigan State University Computer Center. Guttman communalities and Kiel-Wrigley criterion of "three" were used. Quartimax and varimax rotations were computed and were identical for the two-factor solution. Table A1 presents the correlation matrix which was factor analyzed. Note the correlations between quality and the other variables.

Table A2 presents the factor loadings for the three-factor varimax solution. Factor loading below .20 are not given.

TABLE A1*.--Intercorrelation Matrix for Quality and Descriptive Variables.

Response Measure	1.	2.	3.	4.	5.	6.	7.	8.
1. Quality Rating	-							
2. Word Pairs	-43	-						
3. Unusual Meanings	38	03	-					
4. Sentence Complexity	39	-39	-04	-				
5. No. R's Produced	-32	30	-14	-38	-			
6. Sentence Length	46	-42	-01	64	-43	-		
7. No. of Word Moves	26	-48	09	23	-19	22	-	
8. Topic Freedom	44	-19	36	00	-13	-13	06	-
9. Ordinal Position / No. Prod. + 1	08	-01	10	14	-06	20	17	-01

*With n = 200 correlations above .138 are statistically significant at the .05 level, and those above .181 at the .01 level.

TABLE A2.--Factor Matrix for Varimax Rotations with Intact Sample.

Response Measure	Factor 1	Factor 2	Factor 3
Quality	.42	.60	.34
Complexity	.73		.29
Length	.79		.27
Number of Responses	-.51		-.21
Word Pairs	.28		
Word Moves			.66
Unusual Meanings		.69	
Topic Freedom		.67	
Variance Accounted For (Total 49.1%)	19.4%	14.7%	15.0%

The sample of 215 solutions was randomly split in half and factorized using the same options as above. Table A3 presents the last varimax factor matrix obtained for both samples. Only factor loadings above .20 are given. Consideration of the deviation of these factor patterns from that obtained with the whole sample must be tempered by the fact that samples of 100 will inherently be less stable than samples of 200 or larger. Thus, the difference may be attributable to the unreliability of the measures and to the instability in the analysis of small samples.

As suggested by Cattell (1965) and Armstrong and Soelberg (1968) the factor loadings on every variable (regardless of size) were correlated between split halves. The product-moment correlation was .75 and the Spearman rho coeffecient for ranks was .71. Thus in spite of the inherent instability in the small samples the factor patterns obtained were reliable, at least in the test-retest sense.

Quality loading on all factors suggested that the factors all referred to the same concept and would be correlated. An oblique factor analysis was executed using the three-factor varimax matrix as input. The resulting factor matrix is given in Table A4 for factor loadings above .20. Factor 1 was correlated with Factor 3 at .233 which was the highest correlation obtained between factors.

TABLE A3.--Factor Matrices for Varimax Rotations with Split Sample.

Response Measure	Factor 1		Factor 2		Factor 3		Factor 4
	a.	b.	a.	b.	a.	b.	b.
Quality	.50	.46	.60	.63	.31	.18	-.20
Complexity	.70	.74			.37		
Length	.75	.81			.30		
No. of Responses	-.55	-.61		-.29			
Word Pairs	-.67	-.34	-.27			-.70	
Position				.72			.65
Word Moves	.29			.46		.72	
Unusual Meanings			.63	.73	.26		
Topic Freedom	.30		.61	.74			
Variance Accounted For a.	24.9%		13.9%		12.7%		6.4%
b.		21.3%		17.9%		12.6%	

Total Variance Accounted for a. 51.5%, b. 51.9%.

TABLE A4.--Factor Matrix for Oblique Rotation with Intact Sample.

Response Measure	Factor 1	Factor 2	Factor 3
Quality	.36	.55	.26
Complexity	.70		
Length	.76		
No. of Responses	-.48		
Word Pairs			-.71
Position	.28		
Word Moves			.63
Unusual Meanings		.69	
Topic Freedom		.66	
Variance Accounted For (Total 43.2%)	17.1%	14.0%	12.1%

Removing the restrictions imposed by orthogonality did not change the factor patterns, except to reduce all the loadings. With the same factor space as the orthogonal rotations the oblique rotation was able to account for somewhat less variance in a three-factor solution which suggests that more factors would have been drawn off had this solution not been limited to the input communalities based on the three factor varimax solution. It would be recommended that the principal axis factor matrix be the input in further oblique factor analyses of this data.

Discussion

It is apparent that the factor patterns presented by the orthongonal and oblique analyses are reliable in the test-retest sense. Because of the stability of the analyses, it is possible to interpret factors in terms of the questions initially posed. The factor patterns indicates that each variable relates to the quality of solutions to this productive thinking problem. Thus, the variables are interpreted to contribute to quality in three unique ways.

Factor 1 is composed of variables which represent quantitative and structural properties of the solution: position in the response sequence, number of responses produced, and sentence length and complexity. This may be called "verbal structure creativity." The negative loading of number of responses indicates that fewer solutions were conducive to higher quality. One way, then, to produce good sentences would be to concentrate on a few solutions making them long and complex. Sentences may get longer as more are produced through a building from one sentence to another.

Factor 2 has the highest loading on quality and is related to only two response measures, unusual meanings and topic freedom. If one of the given nouns is the subject of a sentence, it could be said that the noun "dominates" the sentence. To use another noun as the subject or direct object would turn the topic of the sentence away from horses and lakes, and the given nouns would become buried in the sentence structure and be less obtrusive to the reader. Similarly, to use a given word with an unusual meaning is also to diverge from the structure inherent in the problem as stated. Because of the high loading of quality on this factor, I would guess that these are the solutions which appear to be out of the ordinary--they are the obviously creative responses. A solution exemplifying this factor would be, "When Happy, our horse, won the Concord Lake Derby, my father no longer grumbled about how expensive it was to care for him." This was rated "6." A reasonable label for this factor would be "ideational creativity."

Factor 3 is drawn from the first factor as illustrated in the difference between the varimax and quartimax solutions and is conceptually related to the same variables. The highest loadings on this factor are the number of word moves and number of old pairs. Quality as a result of these variables would be due to the random scrambling of the given words to achieve some unique sounding sentence. It is a superficial form of creativity, and quality is easily reduced by the appearance of old pairs. Getting at quality in this way could be called "superficial creativity."

APPENDIX B

APPENDIX B

MEMORY TASKS FOR GROUPS TOP, TUM, TR, AND TIR

The format and instructions were the same for all groups which had memory tasks, only the sentences to be learned were changed.

Instructions for the memory tasks:

This is a short memory experiment. You will proceed in this fashion: (1) Read one sentence from this page. (2) Try to remember this sentence long enough to write it down on the next page. (3) Do this from memory--do not look back at this page when you are writing! Follow this procedure for each subsequent sentence.

Instructions on the following page:

Record the sentences you have memorized on this page. Do not look back at the other sentences until you have finished the one you are working on.

At the bottom of this page were the instructions, "Stop! Do not go on until you are told to do so."

Memory task for group TOP--learn old pairs:

1. Because it was a hot day, the happy horse enjoyed the cool water of the expensive lake.

2. The happy horse lives in a beautiful red barn right next to the expensive lake.

3. The elf made the happy horse into an expensive lake.

4. If you lived on an expensive lake, would you like a happy horse playing in the water?

5. The happy horse trotted full speed along the side of the expensive lake with a rider on his back.

6. The happy horse jumped into the expensive lake.

Memory task for Group TUM--learn unusual meanings:

1. I had a happy summer horseback riding at our expensive lake-front cottage.

2. At Lake Park today a horse named "Happy" won an expensive purse.

3. Looking at the expensive lake-front cottage my not-so-happy father gave a horse laugh.

4. At Happy Horse Dude Ranch on the lake my sister spent an expensive summer just horsing around.

5. Heroin, sometimes called "Horse," is expensive to purchase at Happy Hours Resort on the lakeshore.

6. Horse, my little brother, was happy to just sit in the lake making gentle waves with his expensive hat.

Memory task for Group TR--learn relevant sentences control:

1. The horse was happy by the expensive house near the lake.

2. The horse was so happy as it pranced by the lake in its expensive attire.

3. Watching the horse standing by the lake was not an expensive way of being happy.

4. The horse swam the lake to get to the expensive fodder and then was happy.

5. The happy jockey of the horse that won the race bought an expensive house on a lake.

6. Although it was an expensive ordeal, I was happy about buying the horse and the lake.

Memory task for Group TIR--learn irrelevant sentence control:

1. Money alone can buy big things full of noise.

2. Leave my money alone unless you want by big dog to make a noise at you.

3. I'd give all my money to get away from all the noise of this big party and be all alone.

4. Alone, the man stole the money without a noise as big as a squeak.

5. The big wad of money sat alone on the table in the noise-filled room.

6. Alone in a big city with no money, I was afraid of each noise.