PRODUCTION AND JUDGMENT PROCESSES IN A WORD PROBLEM

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
BRADLEY A. BREMER
1968

This is to certify that the

thesis entitled

Production and Judgment Processes

in a Word Problem

presented by

Bradley A. Bremer

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

Donald W. Johnson
Major professor

Date July 26, 1968

O-169

ABSTRACT

PRODUCTION AND JUDGMENT PROCESSES IN A WORD PROBLEM

by Bradley A. Bremer

Problem solving is frequently analyzed into production and judgment processes. In word problems, solution word frequency strongly influences the production process. Few such word problems involve the production of a number of potential solutions. The present study involves a problem which does and was designed to determine whether Ss would alter their production when instructed to emit high or low frequency solutions. The relative influence of production and judgment process on the final solution was also investigated.

Each S solved 16 problems consisting of two consonants. Ss were instructed to think of solutions consisting of four-letter words which began with the first letter and ended with the second. Production data were collected after the solution was given by means of a recognition test consisting of all potential solutions to that problem found in the Thorndike-Lorge (T-L) tables. Ss were instructed to check their earlier productions. Five Groups, of 20 Ss each,

were used. Group F was instructed to elicit the most common solution possible, Group I the most uncommon. Group KR received uncommon instructions and received knowledge of results after every problem. Group C-1 received uncommon instructions, but no recognition test. Finally C-2 did not receive any frequency instructions.

T-L frequencies were determined for each production. Frequencies above 50 were designated high frequency solutions, those of 5 or less as low frequency instructions. Median frequencies of final solutions were significantly higher for Group F than any of the remaining groups. Group F also elicited significantly more high frequency and significantly fewer low frequency solutions. The total number of productions, the number of low frequency productions and the number of high frequency productions did not vary among groups. The total number of productions did increase over blocks of trials and this could largely be accounted for by an increase in low frequency productions. Ss successfully evaluated their own productions and chose a "good" solution from those produced. Median frequencies of final solutions were much higher than median frequencies of all other productions in Group F and much lower in Groups I and KR.

It was concluded that Ss did not alter their production in a facilitory fashion. However, they produced

enough potential solutions of varying frequencies to allow for problem solution under either set of instructions. The judgment phase was critical to solution. These results can be accounted for within the framework of the spew hypothesis, but alternative explanations are possible. A more direct test would require data on production order.

PRODUCTION AND JUDGMENT PROCESSES IN A WORD PROBLEM

By

Bradley A. Bremer

A THESIS

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

DOCTOR OF PHILOSOPHY

Department of Psychology

1968

PRODUCTION AND JUDGMENT PROCESSES

IN A WORD PROBLEM

By

Bradley A. Bremer

A THESIS

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

DOCTOR OF PHILOSOPHY

Department of Psychology

1968

TO JOANNE

ACKNOWLEDGMENTS

The author wishes to express his gratitude to his major professor, Dr. Donald M. Johnson, for his support and guidance in the planning and execution of the research and the development of this manuscript. The author also extends thanks to the members of his guidance committee: Dr. Paul Bakan, Dr. James Phillips, and Dr. Gordon Wood for their advice and criticism.

TABLE OF CONTENTS

Page

DEDICATION ii

ACKNOWLEDGMENTS iii

LIST OF TABLES vi

LIST OF APPENDICES vii

Chapter

I. INTRODUCTION 1

Basic Analysis 1
Production 1
Judgment 7
Purpose 8

II. METHODS 12

Subjects 12
Problems 12
Recognition Test 13
Post-experimental Questionnaire 14
Conditions and Instructions 14
Procedure 16

III. RESULTS 20

Treatment of the Data 20
Analysis of Final Solutions 21
Analysis of Total Production 27
Within Groups Comparisons of Final Solutions and Intervening Productions . 33
Post-experimental Questionnaire 35

IV. DISCUSSION 37

Chapter Page

V. SUMMARY 45

BIBLIOGRAPHY 47

APPENDICES 50

LIST OF TABLES

Table		Page
1.	Summary of groups and procedures	16
2.	Means and standard deviations of the number of high and low frequency final solutions in four groups	23
3.	Summary of analysis of variance of number of high frequency final solutions	24
4.	Summary of analysis of variance of number of low frequency final solutions	26
5.	Means and standard deviations of the number of productions and the number of high and low frequency productions in four groups . .	29
6.	Summary of analysis of variance of number of high frequency responses	30
7.	Summary of analysis of variance of number of low frequency responses	31
8.	Summary of analysis of variance of total number of responses	32
9.	Medians of final solutions and intervening productions	34

LIST OF APPENDICES

Appendix — Page

A. INSTRUCTIONS 50

B. DISTRIBUTION OF RESPONSES TO QUESTIONS #1 AND #3 ON THE POST-EXPERIMENTAL QUESTIONNAIRE 54

INTRODUCTION

Basic Analysis

In many problem solving tasks it is possible to differentiate between production and judgment processes. Such formulations assume that the subject produces a series of possible solutions and then judges the quality of each in relationship to the problem requirements. Davis (1966) also differentiates between problems in which S must make an overt response in order to test the production and those in which the judgment is made covertly. The latter catagory includes a number of verbal problems characterized by one-word solutions. In solving anagrams, for example, the S presumably searches for, or produces, potential solutions until one is judged acceptable. In such problems, factors which influence production and judgment processes can be investigated independently or it is also possible to evaluate the relative importance of each phase in arriving at a solution.

Production

In the production phase of verbal problem solving, frequency of words in the language has proven to be a powerful variable. Ss are more likely to think of words with

which they have had more experience. Underwood and Schultz (1960) have labeled this tendency the "spew hypothesis." Specifically, this hypothesis states that high frequency words, i.e. those which S has encountered frequently, have a higher probability of being evoked and also tend to be evoked earlier. The spew hypothesis can be considered a specific case of response hierarchy theory (Duncan, 1966b; Maltzman, 1955). This more general formulation states that responses vary in habit strength and thus form a divergent hierarchy. In any situation, responses are elicited in order of their position on the hierarchy. Hierarchies may be inferred from association norms, bigram frequencies, etc. The spew hypothesis is a restricted case referring only to hierarchies based on word frequency norms.

Anagram studies provide the most extensive source of support for the spew hypothesis. Mayzner and Tresselt (1958) were the first to report an inverse relationship between solution word frequency and solving time. They explained that high frequency words "will probably be high on the Ss response repertoire and therefore possess a greater potential for evocation as an early implicit response . . . " The stability of the relationship has been demonstrated by several subsequent studies (Johnson, 1966).

Additional support for the spew hypothesis is derived from a study of ambiguous anagrams by Johnson and Van Monfrans (1965). When Ss were asked to give more than one

solution to an anagram they found that the solution with the highest T-L frequency tended to be evoked first.

Duncan (1966a) devised a test of the spew hypothesis which utilized a task designed to "minimize the effects of letter order and to maximize the frequency of the word as a whole." Each problem consisted of two consonants. In the first experiment subjects were instructed to think of a five-letter word which began and ended with the letters specified. The frequency of all potential solutions was determined by the T-L general count (Thorndike and Lorge, 1944). As hypothesized, significantly more high frequency than low frequency words were given. In a second experiment letter combinations were used which had only two possible solutions. Duncan correctly predicted that the higher frequency member of the pair would be given more often. A further restriction was employed in a third experiment. Only the first letter of the solution was given, but the category of the solution was specified, e.g. trees. With two potential solutions, the higher frequency word was given more often. When the subjects were given the higher frequency word, as an example, the number of low frequency responses did increase, but the number of high frequency responses did not increase when the low frequency word was used as an example.

The data cited above provide evidence that solution word frequency is an important variable in solving word problems and supports the spew hypothesis. When Ss are

searching for solutions they are more likely to produce high than low frequency responses. However, this general tendency may not be appropriate for many problems. Many problems are problems because their solutions are uncommon. Therefore, it would often be to the Ss advantage to reverse this general tendency and emit more uncommon responses and to do so earlier in the production process.

Most experimental attempts to increase the production of uncommon responses have gone under the heading of "originality training." These investigations employ various techniques to evoke a greater number of original responses from Ss. While the originality training studies have not dealt directly with word frequency, they have used closely related measures. One aspect of originality, uniqueness, is operationalized in terms of statistical frequency. Uniqueness is usually determined by the number of times the response occurred within the experimental group itself or has occurred in prior groups. This type of criterion is used in Guilfords Unusual Uses Test and Quick Response Test (Wilson, et al., 1953).

"Brainstorming" is one example of a method designed to increase the production of uncommon responses (Parnes and Meadows, 1960). The essential feature of this method is a relaxation of critical evaluation during problem solving. Another method used to facilitate unique responses is Maltzman's originality training (Maltzman, et al., 1958;

1960). The main feature is a training session consisting of repeated free associations to the same stimulus words. Some success in increasing the number of infrequent productions has been achieved through both of these methods.

The elicitation of low frequency responses has also been facilitated by the use of instructions to be original. Maltzman, et al. (1958) increased the number of uncommon associates to a stimulus word by presenting instructions for originality just prior to the test session. The effect also transferred to the Unusual Uses Test. Also working within the framework of Maltzman's technique, Rosenbaum, et al. (1964) presented the originality instructions before the training task. The number of uncommon associates on the subsequent test task was increased, but the effect did not transfer to the Unusual Uses Test. Gerlach, et al. (1964) hypothesized that creativity would be improved when the instructions included cues "concerning criteria for scoring the responses." Therefore, they devised "criteria-cued directions" which contained such cues. The effect of these instructions was compared to 5 other kinds of instructions, including brainstorming. The criteria-cued directions produced more responses rated unique than any other method. Apparently instructions to be original are effective, especially when they specify the characteristics of the desired response.

It seems appropriate to generalize from the uniqueness studies and ask whether Ss can place restrictions on

the "spew" when solving word problems. Under certain conditions Ss may be able to limit their production of implicit responses to low frequency words or at least increase the relative proportion of such words produced. The literature appears devoid of studies relating directly to this question. However, Underwood (1966, p. 590) has suggested that such a restriction of "spew" might occur in anagram solving if a series of anagrams with low frequency solutions were presented to the same Ss. In this situation " . . . it seems reasonable to expect that spew would become more and more restricted as each anagram is solved. . . . " Underwood's hypothesis is based on a study of bigram frequency by Dominowski and Duncan (1964). They found an interaction between anagram bigram frequency and solution word bigram frequency. Performance was best when both types of frequency were high or both were low. In the case of low frequency bigrams Underwood suggested that the low frequency anagram may restrict S's production to words composed of low frequency bigrams. Underwood generalized from bigram frequency to word frequency and predicted that Ss may be able to restrict the "spew" until low frequency anagrams were solved as quickly as high frequency anagrams. This reasoning may be expanded to include other word problems. The central question is: Can Ss restrict the "spew" and limit production to low frequency responses or is the spew resistant to such alterations?

Judgment

The role of the judgment process in most word problems is to determine whether or not the solution which the S has produced is correct. In anagram studies for example, the S has to determine whether his production is a word. The same is true for Duncan's (1966a) task. While the judgmental task is probably relatively simple in such problems, Johnson, Lynch and Ramsey (1967) have shown that word frequency influences the judgment process as well as the production process. In one experiment Ss were able to determine whether a letter arrangement following a word was the same as the word or different more easily when the words were common than when they were uncommon. In a second experiment, Ss were able to evaluate the correctness of three letter starts to five letter anagrams better when the solutions were common words.

In the uniqueness studies judgment plays a somewhat different role. There is no response which is obviously best or correct. The task is more "open ended." The S may be able to produce a variety of potential solutions. The quality of his final solution depends heavily on his ability to choose the best one. This type of task has not been used in studies where word frequency is a variable. It could be introduced in a problem in which there were a series of possible one-word solutions that vary in frequency and the S was induced to evoke a low frequency solution. In such a

task, the judgment phase is essential to achieving a good solution.

There is evidence that Ss are capable of making judgments of frequency. Atteneave (1953) demonstrated that Ss can judge the frequency of use of individual letters. Underwood and Schultz (1966) have shown that Ss can also estimate the frequency of letter bigrams. Both of these findings were supported by Mayzner and Tresselt (1962). Howes (1954) asked Harvard students to rank words "in order of their frequency among Harvard undergraduates." Rank order correlations with the Lorge Magazine Count were .71 for 25 relatively rare words and .87 for 15 words covering a wider range of frequencies.

Purpose

The current study was designed to investigate the process by which Ss arrive at a low frequency response to a word problem. The task was chosen with three considerations in mind. 1) The problem had to have several potential solutions of varying frequency. 2) No single solution was obviously correct or best. 3) The S had to attempt to produce the most uncommon solution he could. A modified version of Duncan's (1966a) task was chosen. Ss were instructed to produce the most uncommon four-letter word which began and ended with specified consonants.

The first major purpose of the study was to determine whether Ss could voluntarily restrict the range of their production on the frequency dimension, i.e. whether they could increase the rate of production of high or low frequency words at will. When the S has been instructed to think of a very uncommon solution, several production models may be considered. First, the S may begin by producing common responses and then continue to go down through the frequency hierarchy at a normal rate until he produces a response which he judges to be sufficiently rare. In this case the S has virtually no control of the production process. He proceeds mechanically through his repertoire until he gets to uncommon responses. A strict response hierarchy explanation seems to coincide with this model.

A second possibility is that the S begins at the top of the frequency hierarchy, by producing common responses, and then accelerates the rate at which he moves through the hierarchy to the rarer responses. This alternative can be considered as a case of partial control over the production phase.

Finally, it may be possible for the S to "jump into" the response hierarchy at any point, including the low frequency end. If so, he can eliminate production of high frequency responses and restrict his production to uncommon responses when desired. In this alternative, the S has rather complete control of the production phase.

It is rather difficult to differentiate between the latter models. Both predict a reduction of high frequency responses. It should be possible to distinguish between them and the first model however, since it predicts no production change dependent on low frequency instructions. The present study is designed to determine whether the S can alter his production when instructed to do so.

The second major purpose of the study was to evaluate the relative importance of production and judgment process in this type of problem. As indicated above, a successful or adaptive production change implies that the S can restrict his production to low frequency responses, or at least increase the proportion of such responses evoked. The judgment phase is considered successful if the S can choose a "good" response from those produced. The S who can choose successfully from a number of productions could give a low frequency solution without altering his production. Of course, it is possible for both processes to be operating in the same problem. The S could restrict his production and then choose the best response from this limited group. The experiment was designed to determine whether one or both of these processes influence the final solution to the problem.

In addition to the two major purposes stated above, there were two secondary purposes. First, practice effects were investigated. Several investigators have looked for practice effects within a series of anagrams. In a recent

review, however, Johnson (1966) reports, "It is clear that no general practice effect has been demonstrated. . . . " The present study was designed to determine if such practice effects occurred in another word problem. Finally, knowledge of results was provided to one experimental group to determine whether this powerful variable would have an effect on a series of problems.

METHODS

Subjects

The subjects were 100 introductory psychology students who participated as part of a class requirement. Subjects were assigned to one of five groups in the order of their appearance with the restriction that the number of males and females be equal across groups. Each of the resulting groups was composed of seven males and 13 females.

Problems

Each of the 16 problems used in the experiment consisted of two consonants. These consonant combinations were the initial and final letters of a subset of four-letter words from the T-L (Thorndike and Lorge, 1944) general word count. The 16 letter combinations were: B-D (for which the relevant words were bead, bend, bard, etc.), B-K, B-T, C-P, C-T, D-T, L-D, L-K, M-T, P-K, R-T, S-D, S-N, S-P and S-R. The subsets of T-L words which began and ended with these letter pairs ranged in number from 10 to 16 words. In each subset, between 2 and 6 words had a T-L frequency of 50 or higher and an equal number had a T-L frequency of 5 or less. There were a total of 207 words for all 16 problems and a mean of 12.94 words per problem. The median T-L frequencies

of the subsets ranged from 3 to over 50. The median of the medians was 15.5 and the median of all 207 words was 15.

A separate 3 by 5 stimulus card was prepared for each combination. The two relevant consonants were printed on the card with a black felt pen and separated by two blank spaces (e.g. D--K).

All 16 problems were presented to a pilot group in order to determine approximately how many solutions would be produced for each problem. Ss were run in a group and the problems were presented verbally. They were instructed to give as many solutions as they could which began with the first letter and ended with the second. With a 90 second solving period, the mean number of solutions per problem ranged from 3.42 to 5.96 (N=24). The level of difficulty of the problems appeared suitable for the purpose of the investigation.

Recognition Test

An individual recognition test was devised for each problem for the purpose of determining which potential solutions Ss had produced during the solving period. It consisted of a small mimeographed form listing all words from the T-L tables which were potential solutions for that problem. The words were listed in a single column in alphabetical order. Each word was followed by a blank space. The

tests were administered after each problem to determine which words S had thought of during the solving period.

Post-experimental Questionnaire

A typewritten questionnaire consisting of four questions was prepared and administered to all Ss. The questions were designed to provide "self report" information about the S's approach to the problem and S's confidence in the technique employed to collect data on implicit behavior. The questions follow: 1) After each problem you were asked to check the words you had thought of earlier. How confident are you that you were able to make an accurate identification of those words you thought of while you were solving the problem? very confident__, fairly confident__, (?)__, fairly unconfident__, very unconfident__. 2) How much emphasis did you place on the check list while you were solving the problem? 3) Did you use any particular method or technique to solve the problems? 4) Do you have any other comments about the task?

Conditions and Instructions

Five groups, three experimental and two control, were used in the experiment. The same problems were given to Ss in all groups, but instructions concerning the type of solution varied. A summary of the groups is found below. Complete instructions for each group can be found in Appendix A.

The three experimental groups were instructed to think of four-letter words which began and ended with the consonants composing the problem. Foreign words and proper nouns were disallowed. Further instructions specified the frequency in the language of the solution word to be produced. Examples of high and low frequency solutions to the T-K problem were given. In Group F (frequent solution instructions) the Ss were instructed to think of the most frequently used word they could which met the other requirements. In Group I (infrequent solution instructions) opposite instructions were given. Ss were told to think of the most infrequently used word they could within the limitations of the task. Ss in Group KR (knowledge of results) received the same instructions as those in Group I. In addition, after each solution had been emitted, the S was told how common or uncommon his response was according to the T-L tables. In all three experimental groups the appropriate recognition test was administered subsequent to each problem.

Group C-1 (control one) received infrequent solution instructions, but was not administered the recognition tests. Its purpose was to determine whether the recognition tests had any influence on solution frequency of subsequent problems. Finally Group C-2 (control two) was simply told to think of as many solutions as possible to the problems. At the end of the solving time, the subjects were instructed to check the items on the recognition test that they had

thought of while solving. The main purpose of this group was to provide a baseline for the three experimental groups on the recognition task. A summary of the groups and procedures is presented in Table 1.

Table 1. Summary of groups and procedures.

Groups	Type of frequency instructions	Final solution elicited	Recognition test administered	Knowledge of results obtained
F	frequent	yes	yes	no
I	infrequent	yes	yes	no
KR	infrequent	yes	yes	yes
C-1	infrequent	yes	no	no
C-2	none	no	yes	no

Procedure

The problems were administered to each subject individually with S facing E across an office desk. A plywood shield on the desk prevented S from seeing E's materials, but did not obstruct the S's view of E or vice versa.

After each S had received instructions appropriate to his group and had indicated that he understood the instructions, the first problem was presented. Both verbal and visual presentations were used. The E held up the stimulus card for approximately five seconds. During this

time E also read the problem letters aloud, (e.g. "the problem is D-K"). The S was then allowed 90 seconds to solve the problem. Any S who attempted to give his solution before the 90 seconds were up was instructed to take the full amount of time for solving. At the end of that period the S was asked to give his solution verbally. This response was then recorded as the S's final solution (FS) to that problem.

Immediately after the FS had been elicited, E handed S the recognition test appropriate to that problem. S was instructed to check only those items which he had thought of while working on the problem. If S indicated that he was uncertain of his ability to recognize his earlier productions, he was instructed to "do the best you can." If the question arose, S was told to ignore the words that he had thought of which did not appear on the list. The incidence of the latter two occurances was rather low.

The items checked on the recognition test included the FS plus other words S had implicitly produced during the solving period. The latter were designated as intervening productions (IPs). Total production (TP) for the problem then consisted of the FS plus the IPs.

For Groups F, I and C-2 the problem was completed when S returned the recognition test to E. In Group KR one step was added. Ss in this group were informed of the frequency of their responses. This information was determined

by the T-L tables. Since no recognition test was used in Group C-1, the problem was considered complete when the S verbalized his solution. In all groups, the next problem was presented without delay upon the completion of the former.

A total of 18 problems were presented to each S. The first two problems were practice trials and were not included in the analysis. The same two problems, D-K and S-M, were used for all Ss. The Ss were not aware that these were practice problems, since the 16 experimental problems followed immediately and the sequence was uninterrupted. The S was not told how many problems were to be solved. Upon completion of the final problem the post-experimental questionnaire was presented to each S. The time required for the entire session was approximately 35 to 50 minutes.

Twenty different problem orders were used in the experiment. Ss in each group were numbered in the order of their appearance. One problem sequence was used for the first S in all groups, another for the second S in all groups, etc. The sequences were determined in the following manner. The 16 problems were randomly divided into two equal groups. One group consisting of problems B-D, B-T, C-T, L-D, P-K, R-T, S-N and S-R was presented first to odd-numbered Ss and last to even-numbered Ss. The other group consisting of problems B-K, D-T, C-P, L-K, L-T, M-T, S-D and S-P was presented first to even-numbered Ss and last to

odd-numbered Ss. One additional restriction was placed on the problem orders. For the first 16 Ss in each group, each problem occurred once in the initial position and once in the final position. Within each sequence, the other 7 problems in each half were randomly ordered. Thus the problem orders were balanced to allow for first and second half comparisons.

RESULTS

Treatment of the Data

The T-L frequency was determined for every response used in the following analysis. Frequencies were easily assigned to all IPs because Ss were limited to the words on the check list. However, some Ss elicited FSs which were not included in the T-L Tables. These exceptions fell into three catagories. 1) Unlisted Words. These FSs were words according to Websters New Collegiate Dictionary, but were not found in the T-L Tables. It was assumed that these words were too uncommon to appear in the tables. Therefore, they were attributed a frequency of less than one per million and were included in the analysis. 2) Non-words. These FSs were neither T-L nor dictionary entries. They were considered invalid responses and were not included in the analysis. 3) Blanks. In a few cases, Ss were unable to produce any solution to the problem.

The number of blanks and non-words was relatively small and fairly evenly divided among groups. The 17 non-words and 6 blanks were distributed as follows: 5 in Group F, 8 in Group I, 6 in Group KR, and 4 in Group C-1. Since each of the 20 Ss within each group attempted to solve 16 problems, there were 320 potential FSs per group. That

meant the number of FSs for which a frequency could be determined was 315, 312, 314, and 316 respectively in Groups F, I, KR, and C-1.

For purposes of analysis, all usable responses were divided into three categories on the basis of their T-L frequencies. The first category, high frequency productions, included all responses with a T-L rating of A or AA (50 or higher). Responses with a T-L rating of five or less were designated as low frequency productions. The third category consisted of the remaining productions in the middle frequency range.

The 16 experimental problems were divided into two equal parts. The data from the first eight problems administered to a S were combined and are designated as Block I. The corresponding data from the final eight problems are referred to as Block II. Problem orders had been counterbalanced to allow for comparisons between the two blocks.

Analysis of Final Solutions

Ss in 4 groups, F, I, KR and C-1, elicited FSs to the problems. The T-L frequency of the FSs reflected the Ss ability to solve the problems. This ability was an assumption essential to the remaining analyses. The ability would be reflected by a tendency to produce high frequency FSs in Group F and an opposite tendency to produce low frequency solutions in Group I, KR, and C-1.

Since the T-L tables do not specify frequencies greater than 50, and many FSs fell into this category, the mean frequencies of the solutions could not be calculated. It was possible, however, to determine the median T-L frequency of each S's FSs. Medians of these medians indicate a sizable difference between Group F and the other 3 groups, all of which received low frequency instructions. The median of the medians exceeded 100 in Group F, while in Groups I, KR, and C-1 the median medians were 8, 4.5 and 10.25 respectively. There was no overlap between the groups. The lowest median in Group F was 43.5, while the highest in the other three groups was 29.5. All comparisons between groups were evaluated by means of the Composite-Rank Test (Guilford, 1965). For each difference, the smaller sum of ranks was compared to the significant R value at the .05 level on the table devised by Wilcoxon for samples of equal size. The analysis showed that Group F differed significantly from all of the other groups, but none of the differences among these groups was significant.

The number of high frequency FSs was analyzed to further evaluate the Ss ability to solve the problem. Means and standard deviations of the number of high frequency responses are presented in Section A of Table 2. Separate means are given for Block I, Block II and the entire session. All standard deviations are shown in parentheses. Hartley's F-max Test of Heterogeneity of Variance (Walker and Lev,

Table 2. Means and standard deviations of the number of high and low frequency final solutions in four groups.

A. Number of high frequency final solutions.

	Groups			
	F	I	KR	C-1
Block I	5.8 (1.29)	1.60 (1.11)	.95 (1.02)	1.20 (1.08)
Block II	5.8 (1.21)	1.00 (.95)	.95 (1.03)	1.45 (1.25)
Total	11.60 (2.11)	2.60 (1.71)	1.90 (1.76)	2.65 (1.85)

B. Number of low frequency final solutions.

	Groups			
	F	I	KR	C-1
Block I	.35 (.66)	3.40 (1.56)	3.60 (1.83)	3.40 (1.39)
Block II	.60 (.66)	3.80 (1.36)	4.05 (1.83)	2.65 (1.15)
Total	.95 (1.07)	7.20 (2.42)	7.65 (3.34)	6.05 (1.83)

1953) was performed on the data and the assumption of homogeneity appears tenable (F_{max}=1.51, 19 df).

A 2X4 factorial analysis of variance was performed on the data to evaluate the effects of conditions, practice and their interaction. The analysis was "mixed" since the conditions effect involved an inter-subject comparison and the comparison between blocks of trials was intra-subject. The analysis corresponds to Lindquist's Type I design (Lindquist, 1953; p. 267). The results of this analysis are summarized in Table 3. The F ratio for conditions was significant at the .01 level. Both practice and interaction F ratios failed to reach statistical significance.

Table 3. Summary of analysis of variance of number of high frequency final solutions.

Source	df	Mean Square	F	P
Between Subjects				
Conditions	3	211.56	116.26	.01
Error between Ss	76	1.82		
Within Subjects				
Blocks	1	.10	.08	n.s.
B x C	3	1.70	1.42	n.s.
Error within Ss	76	1.19		
Total	159			

The significant F ratio for conditions indicates that there were significant differences in the number of high frequency FSs produced by the different groups. Duncans range test was used to determine which of the specific differences were significant. The analysis indicated that Group F produced a significantly larger number of high frequency solutions than any of the other three groups. All differences among the remaining three groups were not significant. Inspection of Table 2 indicates that the magnitude of the differences is rather large. The mean number of high frequency FSs is more than 4 times greater in Group F than in the next group.

The number of low frequency FSs was analyzed in the same manner as the number of high frequency FSs. The relevant means and standard deviations are found in Table 2 Section B. The F_{max} Test indicated that the variances were heterogeneous (F_{max} = 9.68, df 19). In spite of the fact that the assumption of homogeneity was not satisfied, an analysis of variance was performed to assess the effects of conditions and practice. Edwards, (1960) p. 132 states, " . . . since the F test is very unsensative to nonnormality and since with equal n's it is also insensative to variance inequalities, it would be best to accept the fact that it can be used safely under most conditions."

The same 2X4 factorial analysis of variance was repeated on the low frequency data. Table 4 summarizes this

analysis. Again, the practice and interaction effects were insignificant while the F ratio for conditions was significant at the .01 level. Duncans Range test was used to determine which of the separate group means differed significantly. The results were the opposite of those found for high frequency FSs. The analysis indicated significantly fewer low frequency FS had been evoked in Group F than in any of the other groups. Only one other difference was significant. Group KR produced more low frequency FSs than Group C-1.

Table 4. Summary of analysis of variance of number of low frequency final solutions.

Source	df	Mean Square	F	p
Between Subjects				
Conditions	3	90.09	32.35	.01
Error between Ss	76	2.79		
Within Subjects				
Blocks	1	1.06	.80	n.s.
B x C	3	2.59	1.95	n.s.
Error within Ss	76	1.32		
Total	159			

The difference between the 3 groups which received low frequency instructions and Group F is rather large. The smallest of the means for these 3 groups is over 6 times as great as the mean for Group F. This difference, together with the differences in means on the high frequency FSs and the differences in median frequencies, provides substantial evidence that the Ss were able to solve the basic problem and emit the kind of solution requested by the instructions.

Analysis of Total Production

Data on the Ss TP during the problem solving period was obtained from Groups F, I, KR and C-2 by means of the recognition test. Differences in the T-L frequencies in the TP were taken as evidence of an alteration in the Ss production process. Group C-2 provided a production baseline since its production was not influenced by frequency instructions. Expected production changes in Group F would take the form of relatively high frequency responses while Ss in Groups I and KR would produce relatively fewer frequent responses and more uncommon responses.

The median T-L frequency was derived for every S's TP. Then, for each group, a median of these medians was determined. These medians of medians follow: Group F, 31; Group I, 24; Group KR, 24.75; and Group C-2, 32. The Kruskal-Wallis H-Test was applied to the data (H = 5.83, 3 df). The null hypothesis of differences among groups

could not be rejected. Thus these data do not provide any evidence of higher or lower frequency production in any of the 4 groups.

To further evaluate the extent of the production changes, the mean number of high frequency productions and the mean number of low frequency productions was calculated for each group. These means and corresponding standard deviations are presented in Sections B and C of Table 5.

The data on the number of high frequency productions was subjected to the F_{max} Test to evaluate the hypothesis of heterogeneity of variance. The resulting ratio indicates that homogeneity may be assumed (F_{max} = 1.87, 19 df). The same 2X4 "mixed" factorial analysis of variance used on the FSs was applied to the TP data. A summary of this analysis for high frequency TP is presented in Table 6. None of the F ratios in the analysis were significant. There was no evidence that any group produced any more high frequency responses than any other group. Also, there was no increase or reduction in the number of high frequency productions from Block I to Block II.

The same analysis was applied to the number of low frequency responses. Again, the F_{max} ratio indicated that the assumption of homogeneity was not violated (F_{max} = 2.44, 19 df). The analysis of variance is summarized in Table 7. The conditions effect was not significant indicating there

Table 5. Means and standard deviations of the number of productions and the number of high and low frequency productions in four groups.

A. Total number of productions.

	Groups			
	F	I	KR	C-2
Block I	35.40 (12.18)	39.45 (11.90)	39.05 (10.59)	37.40 (12.87)
Block II	37.65 (9.65)	40.90 (11.90)	40.75 (11.13)	39.85 (14.75)
Total	73.05 (21.33)	80.35 (26.41)	79.80 (21.13)	77.25 (26.82)

B. Number of high frequency productions.

	Groups			
	F	I	KR	C-2
Block I	15.30 (5.18)	16.45 (5.11)	15.30 (3.72)	15.70 (5.06)
Block II	15.70 (4.12)	15.75 (4.56)	15.30 (3.67)	16.45 (5.76)
Total	31.00 (8.62)	32.20 (9.03)	30.60 (6.99)	32.15 (9.96)

C. Number of low frequency productions.

	Groups			
	F	I	KR	C-2
Block I	6.65 (3.47)	9.00 (3.29)	9.20 (5.11)	7.90 (4.11)
Block II	7.45 (2.73)	10.35 (3.15)	10.30 (4.33)	9.30 (4.57)
Total	14.10 (5.58)	19.35 (5.55)	19.50 (8.67)	17.20 (8.16)

Table 6. Summary of analysis of variance of number of high frequency responses.

Source	df	Mean Square	F	p
Between Subjects				
Conditions	3	7.32	.18	n.s.
Error between Ss	76	41.26		
Within Subjects				
Blocks	1	.01	.00	n.s.
B x C	3	3.54	.43	n.s.
Error within Ss	76	8.24		
Total	159			

were no differences in the mean number of low frequency productions among groups. The practice effect was significant, however. The number of low frequency productions increased from Block I to Block II. The interaction effect was not significant.

The analysis of high and low frequency productions supports the findings based on medians. The high and low frequency instructions and knowledge of results did not significantly affect the frequency of the responses produced in the problem situation.

Table 7. Summary of analysis of variance of number of low frequency responses.

Source	df	Mean Square	F	p
Between Subjects				
Conditions	3	57.64	2.11	n.s.
Error between Ss	76	27.38		
Within Subjects				
Blocks	1	66.31	11.57	.01
B x C	3	1.17	.20	n.s.
Error within Ss	76	5.73		
Total	159			

Means and standard deviations of the total number of productions (disregarding frequencies) can be found in Table 5, Section A. The F_{max} Test was applied to this data and the resulting ratio indicated that the hypothesis of heterogeneity of variance is unsupported (F_{max} = 1.61, 19 df). The analysis of variance used in the prior sections was repeated on this data. Table 8 summarizes the analysis. The conditions effect was not significant indicating that the TP did not vary across groups. The Conditions X Blocks interaction was also non-significant. The F ratio for the practice effect indicates that significantly more productions

were elicited in Block II than Block I. This difference corresponds to the increase in low frequency productions in Block II.

Table 8. Summary of analysis of variance of total number of responses.

Source	df	Mean Square	F	p
Between Subjects				
Conditions	3	111.49	.396	n.s.
Error between Ss	76	281.33		
Within Subjects				
Blocks	1	152.10	9.681	.01
B x C	3	2.27	.144	n.s.
Error within Ss	76	15.71		
Total	159			

Product-moment correlations were obtained between the number of "good" productions and the total number of other productions. The latter statistic was used instead of TP to eliminate the spuriousness which would have resulted from common occurance of high or low frequency productions in both of the correlated measures. Groups I and KR were combined and yielded a coefficient of .65 between the number

of low frequency productions ("good" solutions by definition) and productions with a T-L frequency of 6 or higher. For Group F the correlation coefficient between number of high frequency productions and productions with T-L frequencies below 50 was .76. These coefficients are within the range found by Parnes and Meadows (1959) and Gerlach et al. (1964) in uniqueness studies.

The coefficient of correlation between the number of low frequency FSs and remaining productions was .34 in Groups I and KR. The related coefficient between the number of high frequency responses and other responses for Group F was .68.

Rank order correlation coefficients were obtained between the total number of productions and the median frequency of TP. The resulting coefficients were -.12 in Groups I and KR and -.25 in Group F.

Within Groups Comparisons of Final Solutions and Intervening Productions

Ss in only 3 of the 5 groups (the 3 experimental groups) both emitted a FS and responded to the recognition test. Within each of these groups a comparison was made between the T-L frequencies of the FSs and the IPs.

Neither individual nor group means can be reported because many of the responses have a T-L frequency above 50 and interval data on these frequencies were not available. Therefore, the median FS and median IP was determined for

each of the 60 Ss. Group medians (actually medians of the individual S's medians) were then determined for each of the 3 groups. These group medians, for FSs and IPs, can be found in Table 9.

Table 9. Medians of final solutions and intervening productions.

	Groups		
	F	I	KR
Median FS	>100.0	8.0	4.5
Median TP	21.5	35.5	34.0

In Group F, the median T-L frequency of the FSs is considerably higher than the median frequency of the IPs. In fact, when each S was considered individually, the median frequency of the FSs was greater than the median frequency of the IPs for every one of the 20 Ss. A sign test was performed on the data and the difference is highly significant.

In Groups I and KR the results were in the opposite direction. In both groups, the group medians are distinctively higher for IPs than FSs. Again an inspection of individual S medians revealed that there were no reversals. Every Ss median IP is greater than his median FS. The sign

test indicates the difference is highly significant in both cases.

Post-experimental Questionnaire

The post-experimental questionnaire was given to all Ss. However, Questions 1 and 2 were omitted in Group C-1 since those questions pertained to the recognition test which was not administered to those Ss. Similarly Question 2 was omitted for Group C-2. Questions 2 and 4 did not yield any data worth reporting. Responses were too widely varied for Question 4 and Ss apparently did not understand Question 2.

On Question 1, 36 Ss replied that they were very confident and 39 considered themselves fairly confident that they could identify the words on the checklist that they had thought of while working on the problem. Only 3 Ss responded by checking the (?) alternative and the fairly unconfident and very unconfident responses were each checked once. In all then, 75 out of 80 Ss responded that they felt some degree of confidence. Analysis of the responses according to groups is presented in Appendix B.

On Question 3, only 11 out of 100 Ss reported that they did not use any special method or technique to solve the problems. Out of the 89 who answered yes, 65 reported using some type of letter substitution system, i.e. going

through the alphabet or vowel substitution. Other methods included rhyming, thinking of objects, etc.

DISCUSSION

The major purpose of the study was to determine the importance of the production and judgment processes in reaching a "good" solution to a relatively simple word problem. In order to study these intervening processes, it was essential that a good solution be reached. A good solution in this study consisted of a high or low frequency word, depending on the experimental instructions. It would have been impossible to evaluate the extent to which production and judgment processes affected the solution if there had been no differences between the frequencies of the FSs elicited.

The results leave little doubt that the instructions were effective and the basic problems were solved. There were substantial differences in the FS frequencies of Group F and Groups I, KR, and C-1, all of which received low frequency instructions. The separation of median T-L frequencies for FSs illustrates the extend of the differences. The lowest median T-L frequency in Group F was considerably higher than the highest median T-L frequency in the other groups. The data on the number of high and low frequency FSs substantiate the conclusion that the Ss were able to

reach good solutions. With this fundamental assumption satisfied, the major issue can be approached.

If changes in the production phase instrumental to solving took place, there should have been differences in the T-L frequencies of the potential solutions. However, there were no significant differences in the median T-L frequency of the T-P, the mean number of high frequency productions, or the mean number of low frequency productions. Thus the data indicates that the Ss produced potential solutions of comparable frequencies whether they were instructed to reach high or low frequency solutions. There is no way to determine whether or not they were capable of making facilitory production changes, but they obviously did not do so in this situation.

Three models of solution production were described in the introductory section. The results are more consistent with the first model than with the latter two. According to the former model, Ss have little control over the production process and proceed mechanically through the frequency hierarchy. The latter models are inappropriate since both predict a production change.

The practice effects in the production phase are also congruent with the spew hypothesis. T-P increased significantly from Block I to Block II. The increase in production, moreover, occurred at one end of the frequency continuum. The increase in T-P apparently resulted from an

increase of low frequency responses, since a significant practice effect was found for low frequency productions but not for high frequency productions. This latter finding can be explained in terms of the spew hypothesis. In the second block of problems Ss presumably went through the upper part of the hierarchy in the same manner as they did in Block I, but were able to proceed further down the hierarchy. The additional productions, therefore, came from the lower end of the frequency hierarchy i.e. were uncommon words. It is significant that the increase in low frequency productions occurred in all four groups, including Group F. This indicates that spew, or some similar mechanism, operated in the face of instructions to produce high frequency words.

While the production data can be accounted for by the spew hypothesis, the evidence is necessarily indirect. The spew hypothesis makes specific predictions about the sequence of production, but the method used did not reveal the order in which the potential solutions were produced. Therefore, it is not possible to determine exactly what happened during the production phase. If a major production change had occurred, i.e. if Group F had produced a preponderance of high frequency words, while Groups I, KR and C-2 had produced a preponderance of low frequency responses it would have been possible to conclude that spew had been restricted in the manner suggested by Underwood (1966). It can only be concluded that there is no evidence that that happened. The

distribution of frequencies was approximately the same for all groups. This distribution might have occurred in a number of ways. It is possible that spew was operating in all groups in the manner described above. It is also possible that the order of production was haphazard. Or, it is even possible that Ss actually produced the relevant responses first and then went on to produce the other solutions. The S's responses to Question 3 on the post-experimental questionnaire does provide some suggestions about the production process. Most Ss indicated that they used some strategy to generate solution words. Many Ss mentioned some type of letter substitution technique, e.g. placing vowels in the second letter position and then filling in position three with a series of consonants. This type of method would not necessarily lead to a production order congruent with spew.

While the spew hypothesis does account for the data other explanations are equally tenable. Future research may determine which alternative is most reasonable. In the present investigation the important fact remains that the Ss were unable to alter or restrict their production in a manner that would facilitate problem solving.

The failure to obtain production differences based on instructions contradicts the findings from the uniqueness studies, particularly Gerlach et al. (1964). Their criteria-cued directions increased the number of unique responses on

the Unusual Uses Test. In the present study, the criterion was clearly specified. Ss were instructed to give "infrequent" and "uncommon" solutions and examples were given. More stringent restrictions are placed on production in the present study than in the Unusual Uses Test. Perhaps criteria-oriented instructions are more effective when the potential responses are less limited than in the present study.

The data indicate that the judgment process was critical to solving. Results of the comparison between median T-L frequency of FSs and IPs are clear. In Group F, the median FS frequency is higher than the median IP frequency for every S. In the three groups given low frequency instructions, the differences are equally obvious, but in the opposite direction. Every S's median FS frequency was lower than his median IP frequency. An inspection of the individual data revealed that the differences were substantial in almost every case. Ss did choose good solutions, high or low frequency as the instructions required, from among those they produced.

Some general conclusions about the problem solving process, based on the results discussed, can be made at this point. Ss had little trouble generating a number of word solutions. They did not succeed in restricting their productions to solutions of relevant frequency, for the distribution of potential solution frequencies was similar for all groups. However, the array of words generated apparently

included a sufficient number of both high and low frequency words to allow for problem solution under either set of frequency instructions. The key to the solution was S's judgment ability. Ss successfully chose a "good" solution from the number produced. The probability of a good solution being available and subsequently being chosen increased if many potential solutions were produced. This is reflected in the correlations between T-P and the number of "good" productions and number of "good" FSs.

The current investigation utilized a unique method for collecting production data. "Thinking out loud" techniques have been used for similar purposes in anagram studies (Mayzner, Tresselt and Helbrook, 1964). In the present study, however, the data was collected after the solution had been elicited, to avoid interference with the solving process. The nature of the problem itself afforded an opportunity to employ a recognition test for this purpose, since the major portion of potential productions could be specified in advance. There were no indications that the administration of the recognition test had any substantial effects on the remaining problems in the series. The FSs of Group C-1, which did not receive the recognition test, did not differ significantly from Group I on any of the dependent variables. The data from the post-experimental questionnaire indicated that a large majority of the Ss felt at least fairly confident that they were able to recognize their

earlier productions. The recognition test is apparently a workable method and suitable for further research.

The recognition test has at least one important limitation. It does not yield information about the order of production. Order information could be obtained by asking Ss to verbalize productions as they occur. With the present problem, Ss may be able to specify order on a post-test if the number of productions were reduced. This could be done by using five-letter words, which Duncan's (1966a) Ss found quite difficult, or by using problems with fewer potential solutions.

A revision of the current design in which solving time is varied should provide a more direct test of the spew hypothesis. If spew is operating additional solving time should facilitate problem solving under conditions of low frequency instructions, but not high frequency instructions. If, as spew predicts, high frequency solutions are produced first, all or at least most of the relevant productions for high frequency instructions should occur during the early portion of the solving period and later productions should be relatively useless. With low frequency instructions, however, the spew hypothesis predicts that the longer the S has to solve the further he will proceed down the hierarcy, the more uncommon his responses will be and therefore the more relevant. If spew is operating there should be a significant

interaction between solving time and high or low frequency instructions in the production of "good" solutions.

SUMMARY

Problem solving is frequently analyzed into production and judgment processes. In word problems, solution word frequency strongly influences the production process. Few such word problems involve the production of a number of potential solutions. The present study involves a problem which does and was designed to determine whether Ss would alter their production when instructed to emit high or low frequency solutions. The relative influence of production and judgment process on the final solution was also investigated.

Each S solved 16 problems consisting of two consonants. Ss were instructed to think of solutions consisting of four-letter words which began with the first letter and ended with the second. Production data were collected after the solution was given by means of a recognition test consisting of all potential solutions to that problem found in the Thorndike-Lorge (T-L) tables. Ss were instructed to check their earlier productions. Five Groups, of 20 Ss each, were used. Group F was instructed to elicit the most common solution possible, Group I the most uncommon. Group KR received uncommon instructions and received knowledge of results after every problem. Group C-1 received uncommon

instructions, but no recognition test. Finally C-2 did not receive any frequency instructions.

T-L frequencies were determined for each production. Frequencies above 50 were designated high frequency solutions, those of 5 or less as low frequency instructions. Median frequencies of final solutions were significantly higher for Group F than any of the remaining groups. Group F also elicited significantly more high frequency and significantly fewer low frequency solutions. The total number of productions, the number of low frequency productions and the number of high frequency productions did not vary among groups. The total number of productions did increase over blocks of trials and this could largely be accounted for by an increase in low frequency productions. Ss successfully evaluated their own productions and chose a "good" solution from those produced. Median frequencies of final solutions were much higher than median frequencies of all other productions in Group F and much lower in Groups I and KR.

It was concluded that Ss did not alter their production in a facilitory fashion. However, they produced enough potential solutions of varying frequencies to allow for problem solution under either set of instructions. The judgment phase was critical to solution. These results can be accounted for within the framework of the spew hypothesis, but alternative explanations are possible. A more direct test would require data on production order.

BIBLIOGRAPHY

Attneave, F. Psychological probability as a function of experienced frequency. Journal of Experimental Psychology, 1953, 46, 81-86.

Davis, G.A. Current status of research and theory in human problem solving. Psychological Bulletin, 1966, 66, 35-54.

Duncan, C.P. Effect of word frequency on thinking of a word. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 434-440. (a)

Duncan, C.P. Response hierarchies in problem solving. Presidential Address to Midwestern Psychological Association, Chicago, 1966. (Reprinted in, Duncan, C.P. Thinking: Current experimental studies. New York: J.B. Lippencott Company, 1967) (b).

Dominowski, R.L. and Duncan, C.P. Anagram solving as a function of bigram frequency. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 321-325.

Edwards, A.L. Experimental design in psychological research. New York: Holt, Rinehart & Winston, 1960.

Gerlach, V., Schultz, R., Baker, R., and Mazer, G. Effects of variations in test instruction on originality test response. Journal of Educational Psychology, 1964, 55, 79-83.

Guilford, J.P. Fundamental statistics in psychology and education. New York: McGraw-Hill Book Company, 1965.

Howes, D. On the interpretation of word frequency as a variable affecting speed of recognition. Journal of Experimental Psychology, 1954, 48, 106-112.

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

Johnson, D.M., Lynch, D.O. and Ramsey, J.G. Word frequency and verbal comparisons. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 403-407.

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

Lindquist, E.F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin Company, 1953.

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

Maltzman, I., Brooks, L.O., Bogartz, W. and Summers, S. The facilitation of problem solving by prior exposure to uncommon responses. Journal of Experimental Psychology, 1958, 56, 399-406.

Maltzman, I., Simon, S., Raskin, D., and Licht, L. Experimental studies in the training of originality. Psychological Monographs, 1960, 74, 493.

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

Mayzner, M.S. and Tresselt, M.E. The ranking of letter pairs and single letters to match digram and single-letter frequency counts. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 203-207.

Mayzner, M.S., Tresselt, M.E. and Helbock, H. An exploratory study of mediational responses in anagram problem solving. Journal of Psychology, 1964, 57, 263-274.

Parnes, S.J. and Meadow, A. Evaluation of persistance of effects produced by a creative problem solving course. Psychological Reports, 1960, 7, 357-361.

Rosenbaum, M.E., Arensen, S.J. and Panman, R.A. Training and instructions in the facilitating of originality. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 50-56.

Thorndike, E.L. and Lorge, I. The teacher's word book of 30,000 words. New York: Columbia University Press, 1944.

Underwood, B.J. Experimental psychology. New York: Appleton-Century-Crofts, 1966.

Underwood, B.J. and Schultz, R.W. Meaningfulness and verbal learning. Chicago: Lippencott, 1960.

Walker, H.M. and Lev, J. Statistical inference. New York: Holt, Rinehart & Winston, 1953.

Wilson, R.C., Guilford, J.P. and Christensen, P.R. The measurement of individual differences in originality. Psychological Bulletin, 1953, 50, 362-370.

APPENDIX A

INSTRUCTIONS

INSTRUCTIONS

Complete instructions for Group I are presented below. Instructions for the remaining groups involve various modifications which are described subsequently.

Group I:

(Paragraph 1) "I have a series of short problems I want you to solve. Each problem consists of two letters. You are to produce a four letter word that begins with the first letter and ends with the second. You may not use foreign language words or proper nouns. For example, if I give you the letters T-K words like TASK, TOOK or TUSK meet all the requirements."

(Paragraph 2) "Words vary in the extent to which they are used in the language. Some words like FROM or DOOR are very common, while others like FURL or GARB are much less common. I want you to think of the most infrequently used word you can which meets the other requirements. As an example, in the T-K problem TUSK is a better solution than TOOK because it is much less common."

(Paragraph 3) "You will have 90 sec. to work on each problem. Do not give me your answer until I tell you the time is up. I will check to see what other words you have produced after

the problem is solved, but don't be too concerned with this while you are solving the problem."

(Paragraph 4) "Here again is a list of the solution requirements. (1) It must be a word. (2) It must begin and end with the letters specified. (3) It must have exactly four letters. (4) It can be neither a proper noun or foreign word. (5) Finally, it should be the most uncommon word you can think of, within the other limitations."

Group F:

The word "frequently" was substituted for "infrequently" in the second sentence of paragraph 2 and the words "TOOK" and "TUSK" were interchanged in the third sentence. The word "common" was substituted for "uncommon" in paragraph 4, item 5.

Group KR:

The following sentence was added to paragraph 2. "After each problem, I will tell you how infrequent your answer is so you know how well you are doing."

Group C-1:

The second sentence was deleted from paragraph 3.

Group C-2:

Paragraph 2 was deleted. The second sentence of paragraph 3 was changed to read, "I will check to see what words you thought of after the time is up." Finally, item 5 in paragraph 4 was deleted.

Instructions for the Recognition Test: (Groups I,F,KR and C-1)

"Here is a list of words related to the problem you have just completed. Check those words and only those words which you thought of while you were working on the problem. Some of the words you produced may not appear on the list, but don't be concerned about that.

APPENDIX B

DISTRIBUTION OF RESPONSES TO QUESTION #1 & #3 ON THE POST-EXPERIMENTAL QUESTIONNAIRE

Frequency Distribution of Responses to Question #1.--"After each problem you were asked to check the words you had thought of earlier. How confident are you that you were able to make an accurate identification of those words you thought of while you were solving the problem."

Response	Group			
	F	I	KR	C-2
very confident	13	8	7	8
fairly confident	5	11	12	11
(?)	2	1		
fairly unconfident			1	
very unconfident				1

Frequency Distribution of Responses to Question #3.--"Did you use any particular method or technique to solve the problems?"

Response	Group			
	F	I	KR	C-2
Yes	2	3	2	1
No	18	17	18	19