A Multiple Choice Version of the Loyola Language Study: A Comparison with the Original Version

Richard C. Braun
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses

Part of the Psychology Commons

Recommended Citation
https://ecommons.luc.edu/luc_theses/1791

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 1963 Richard C. Braun
A MULTIPLE CHOICE VERSION OF THE LOYOLA LANGUAGE STUDY --

A COMPARISON WITH THE ORIGINAL VERSION

by

Richard C. Braun, S.J.

A Thesis Submitted to the Faculty of the Graduate School
of Loyola University in Partial Fulfillment of
the Requirements for the Degree of
Master of Arts

June
1963
LIFE

Richard C. Braun, S.J., was born in Buffalo, New York, July 9, 1927.

He was graduated from Canisius High School, Buffalo, New York, June, 1945 and entered the Society of Jesus in August of the same year. He was graduated from Woodstock College, Maryland, June, 1951 with the degree of Bachelor of Arts and June, 1953 with the degree of Master of Arts.

From 1952 to 1955 the author taught English and Classics at Fordham Preparatory School, Fordham University, New York, New York. From 1958 to 1959 he served as Chaplain at the Maryland Training School for Boys, Towson, Maryland. He began his graduate studies at Loyola University in September, 1960.

From September, 1961 to June, 1962 he was assistant to the Director of Guidance at Canisius College, Buffalo, New York and attended graduate courses in Psychology at the University of Buffalo.
PREFACE

The author wishes at this time to express his sincere gratitude to his director, the Reverend Vincent V. Herr, S.J. for his patient, continued guidance, and his thoughtful suggestions throughout this study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF THE RELATED LITERATURE</td>
<td>4</td>
</tr>
<tr>
<td>III. CONSTRUCTION OF THE MULTIPLE CHOICE TEST</td>
<td>14</td>
</tr>
<tr>
<td>IV. PROCEDURE AND RESULTS</td>
<td>29</td>
</tr>
<tr>
<td>V. CONCLUSIONS</td>
<td>51</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>57</td>
</tr>
<tr>
<td>APPENDIX I</td>
<td>59</td>
</tr>
<tr>
<td>APPENDIX II</td>
<td>60</td>
</tr>
<tr>
<td>APPENDIX III</td>
<td>61</td>
</tr>
<tr>
<td>APPENDIX IV</td>
<td>62</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table                                                                                                                     Page
I.  TYPICAL STIMULUS WORDS AND RESPONSE CHOICES FROM MALLER'S CONTROLLED ASSOCIATION TEST                                8
II. SAMPLE OF STIMULI AND RESPONSES ON FIRST DESIGN OF MULTIPLE CHOICE TEST                                              16
III. COMPARISON OF NATURE OF CHOICES GIVEN IN PRELIMINARY (PILOT STUDY) AND LATER MULTIPLE CHOICE DESIGNS:
     DESIGN X, DESIGN Y, DESIGN Z                                                                                           23
IV.  POSSIBLE Z-SCORES FOR RESPONSES TO 5 STIMULUS WORDS ON THE LOYOLA LANGUAGE STUDY (STANEK - CHICAGO MEN AND WOMEN) AND POSITION OF FOUR CHOICES OF DESIGN X ON Z-SCORE TABLES                          25
V.   POSSIBLE Z-SCORES FOR RESPONSES FOR 5 STIMULUS WORDS ON THE LOYOLA LANGUAGE STUDY (STANEK - CHICAGO MEN AND WOMEN) AND POSITION OF FIVE CHOICES OF DESIGN Y ON Z-SCORE TABLES                          26
VI.  POSSIBLE Z-SCORES FOR RESPONSES TO 5 STIMULUS WORDS ON THE LOYOLA LANGUAGE STUDY (STANEK - CHICAGO MEN AND WOMEN) AND POSITION OF SEVEN CHOICES OF DESIGN Z ON Z-SCORE TABLES                          27
VII. STIMULUS WORDS FOR CHICAGO MALES (M) AND FEMALES (F) WITH THE HIGHEST SCREENING EFFICIENCY BETWEEN CHICAGO NORMALS AND CHICAGO PATIENTS                                        33
VIII. SPEARMAN COEFFICIENT (rho), PEARSON COEFFICIENT (r_xy) AND STANDARD ERROR OF PEARSON COEFFICIENT (σ_r) BETWEEN SCORES OF GROUP X (N=61) ON THE LOYOLA LANGUAGE STUDY AND SCORES OF THE SAME GROUP X ON DESIGN X (4 CHOICES) ACCORDING TO EACH OF THREE METHODS OF SCORING                       36
Table of Contents

IX.  SPEARMAN COEFFICIENT (rho), PEARSON COEFFICIENT 
(rxy) AND STANDARD ERROR OF PEARSON COEFFICIENT  
(σr) BETWEEN SCORES OF GROUP Y (N:61) ON THE 
LOYOLA LANGUAGE STUDY AND SCORES OF THE SAME GROUP Y 
ON DESIGN Y (5 CHOICES) ACCORDING TO EACH OF THREE 
METHODS OF SCORING .............................. 36

X.  SPEARMAN COEFFICIENT (rho), PEARSON COEFFICIENT 
(rxy) AND STANDARD ERROR OF PEARSON COEFFICIENT  
(σr) BETWEEN SCORES OF GROUP Z (N:61) ON THE LOYOLA 
LANGUAGE STUDY AND SCORES OF THE SAME GROUP Z ON 
DESIGN Z (7 CHOICES) ACCORDING TO EACH OF THREE 
METHODS OF SCORING .............................. 37

XI.  COMPARISON OF 1, 2, and 3, I.E., 1. CORRELATION OF 
COEFFICIENTS BETWEEN SCORES OF GROUP X ON LOYOLA 
LANGUAGE STUDY AND ON DESIGN X (4 CHOICES); 2. COR- 
RELATION COEFFICIENTS BETWEEN SCORES OF GROUP Y ON 
LOYOLA LANGUAGE STUDY AND DESIGN Y (5 CHOICES); 
3. CORRELATION COEFFICIENTS BETWEEN SCORES OF 
GROUP Z ON LOYOLA LANGUAGE STUDY AND DESIGN Z 
(7 CHOICES) .............................. 41

XII. MEANS, STANDARD DEVIATIONS, STANDARD ERRORS OF MEANS, 
DIFFERENCES OF MEANS, STANDARD ERRORS OF DIFFERENCE 
BETWEEN MEANS AND CRITICAL RATIOS FOR DIFFERENCE 
BETWEEN MEANS FOR SCORES OF GROUP X (N:61) SCORED 
BY THREE METHODS (METHOD I - MEDIAN NORMS; METHOD II - 
25 CRITICAL ITEM NORMS; METHOD III - MINNESOTA NORMS) 
ON THE LOYOLA LANGUAGE STUDY AND ON MULTIPLE CHOICE DESIGN X (4 CHOICES) .............................. 43

XIII. MEANS, STANDARD DEVIATIONS, STANDARD ERRORS OF MEANS, 
DIFFERENCES OF MEANS, STANDARD ERRORS OF DIFFERENCE 
BETWEEN MEANS AND CRITICAL RATIOS FOR DIFFERENCE 
BETWEEN MEANS FOR SCORES OF GROUP Y (N:61) SCORED 
BY THREE METHODS (METHOD I - MEDIAN NORMS; METHOD II - 
25 CRITICAL ITEM NORMS; METHOD III - MINNESOTA NORMS) 
ON THE LOYOLA LANGUAGE STUDY AND ON MULTIPLE CHOICE DESIGN Y (5 CHOICES) .............................. 44

XIV. MEANS, STANDARD DEVIATIONS, STANDARD ERRORS OF MEANS, 
DIFFERENCES OF MEANS, STANDARD ERRORS OF DIFFERENCE 
BETWEEN MEANS AND CRITICAL RATIOS FOR DIFFERENCE 
BETWEEN MEANS FOR SCORES OF GROUP Z (N:61) SCORED 
BY THREE METHODS (METHOD I - MEDIAN NORMS; METHOD II - 
(Continued)
**Table**

**XIV.** 25 CRITICAL ITEM NORMS; METHOD III - MINNESOTA (Cont'd.) NORMS) ON THE LOYOLA LANGUAGE STUDY AND ON MULTIPLE CHOICE DESIGN Z (7 CHOICES) ........................................... 45

**XV.** SPEARMAN RANK DIFFERENCE CORRELATION COEFFICIENTS AND PEARSON PRODUCT MOMENT CORRELATION COEFFICIENTS BETWEEN SCORES OF GROUP Y SCORED BY THREE METHODS ON THE LOYOLA LANGUAGE STUDY AND ON MULTIPLE CHOICE DESIGN Y (5 CHOICES) ........................................... 48
CHAPTER I

INTRODUCTION

The Loyola Language Study, a semi-controlled word association test is a recent diagnostic instrument which attempts to provide a quantitative measure of some aspects of personality indicative of normalcy or psychopathology, specifically schizophrenia.

It is non-threatening in format, easily administered to individuals or groups. It is a departure from other association tests in its "set," the subject being asked to pause and think before responding to each of the eighty stimulus words, and then to give that response he thinks most people would give, confronted with the same stimulus.

The rationale is based on the assumption that marked deviation from communality of response and inability to recognize this deviation is an indicator of pathology. This test is an empirical technique for measuring such deviation.

The purpose of this present study is to begin to investigate whether, by adding the further control of limitation of the subject's responses by presenting a multiple choice of possible response words, the same deviation (or its absence) will still be able to be measured. Del Vecchio has stated that
the effectiveness of the Loyola Language Study should lie in its ability to tap communality of thought, but more important, it measures the individual's awareness of that communal element. It assumes that the presence of disease is directly related to the person's inability to recognize the deviation of his own thought from the majority of people. (1957)

Herr, in the same vein, has said:

The assumption was that psychotics in general and schizophrenics in particular would not be capable of complying with this type of instruction what most people would be likely to respond and finding common responses as the normals would. (1957)

If "awareness of the communal element" and "ability to recognize deviation of one's own thought" and the incapability of "complying with this type of instruction" are the essential elements that are tapped in the Loyola Language Study, it is hypothesized that a multiple choice version of the Loyola Language Study would measure the same elements as the original.

For, confronted with a complexity of definite printed response words ranging from those with the highest degree of communality to those with none at all -- as opposed to having to think of and rehearse these possibilities mentally -- it is assumed that the subject is confronted with essentially the same task, that of choosing the response that most people would choose.

Employing this assumption, an attempt will be made to construct a multiple choice version of the Loyola Language Study which will present the subject with approximately the same task, and with approximately the same degree of difficulty.
The attempt will be limited to a population of normal male subjects from the Chicago area. If degree of communality of thought (and conversely, absence of the same) can be measured by scores on the Loyola Language Study, it is hypothesized that a significant correlation between scores on the present version of the Loyola Language Study and a multiple choice version will indicate that the latter test will perform essentially the same task as the former -- at least for the limited segment of the population which will be tested.

If this limited, beginning attempt proves successful, perhaps further research in this area would be indicated. If all the many advantageous aspects of the Loyola Language Study could be combined with those of a multiple choice test, especially those of ease and simplicity of administration and scoring, and reduction of time spent by both the subject and administrator, it would seem that an advance would have been made in word association research.

In the following chapter a brief review of that research which is pertinent to the present study will be presented.
CHAPTER II

REVIEW OF RELATED LITERATURE

In the eighty-five years since Galton first published a report on free association (1879) hundreds of investigations have been reported in this area of research. This review will underline only the highpoints of this research and it will emphasize that family of investigation of which the Loyola Language Study is a member -- the empirical, psychometric approach to the word association as an instrument of diagnosis.

One interested in a survey of the more qualitative approach in the Jungian tradition, with more of an emphasis on the formal mechanisms of abnormal mental processes will find an excellent one in Del Vecchio's work. (1957)

It is well to note that in this line of empirical investigation, beginning with Wundt (Humphrey, 1951) the term "free" association can only be used with differing degrees of inaccuracy. Galton allowed his subject to "freely" respond to a verbal stimulus with a single word or more than one word, or with any verbal description at all. But from the time of Wundt, who imposed the controls of a single word response and that of timing, experimenters have used some degree of control over the subject's response, the difference as shall be seen, being in the type or degree of control. And this introduction of controls made quantification of findings possible.
It was the work of Cattell and Bryant (1889) which further precisioned quantification in word association research. These investigators -- allowing only a single word response to each stimulus word -- were the first to record frequency of certain responses. Succeeding experiments have utilized their type of frequency tables as a basis for objective scoring systems. It seems also that Cattell and Bryant were the first to emphasize the notion of "communality of response", a construct which was empirically derived and which underlies much of the word association research, including the Loyola Language Study. Guppy emphasizes this point well when he states

"The fact that frequency tables can be constructed and used as norms for judging abnormal responses is evidence that the individuals in the norm group have had, to a significant degree, similar experiences. And, in spite of the individual differences ... there is a central core of associative matter which is common to all. Presumably, deviations from the norms are not so much a matter of not having the experience of others, but are due to the subjects' inability to make use of the experiences in an associative way." (1959)

Kent and Rosanoff (1910) were the first to attempt to establish empirically, on a large scale, what constituted a "normal" or "common" response. They used 1000 subjects from the ages of eight to eighty and from both sexes as a standardization sample. More than two hundred of these were professional people, about five hundred were civilian employees of a New York State Mental Hospital, the rest were boys and girls of high school age. (The reader may judge the "normalcy" of this group for himself.) To these subjects they presented a set of one hundred stimulus words to which the
subjects gave single word responses. The list of words excluded words likely to recall personal experiences, and words closely connected with each other were separated spatially. From the responses they constructed frequency lists. The lists, divided into responses that were "common" (91.7 percent), "doubtful" (1.5 percent), and "individual" (6.8 percent), provided a standard against which an unusual response could be judged. By a "common" response is meant a word given by at least one other individual. A "doubtful" word is "any reaction word which was not found in the table in its identical form, but which was a grammatical variant." An "individual" response is a response given by no other of the one thousand normal subjects.

Kent and Rosanoff then administered the same stimulus words to two hundred and forty-seven psychotic patients, and they found that with this population "common" responses decreased (70.7 percent) and "individual" responses increased (26.8 percent.) However, by their test they were not able to draw "a sharp distinction between mental health and mental disease (but rather) a gradual and not abrupt transition from the normal state to the pathological state." (Rosanoff, 1938)

However, their work did emphasize the validity of the concept of "communality of thought" among normals (and its relative absence in psychotics). For although by their definition, a "common" response could be one given by only two out of a thousand subjects, it was observed that the lowest percentage of agreement by normal subjects in responding to a stimulus word was 12 percent, i.e., 120 subjects gave the same response.
The top frequency of response was 67%, i.e., 670 out of the thousand normal subjects gave the same response to a stimulus word.

This finding was subsequently reaffirmed by Esper (1918) who further established that wide differences in age and education did not noticeably affect the great similarity to responses among normals. On the other hand, Noh and Guilford (1930) and Stanek (1956) have demonstrated that men have less communality of response than do women, although the responses of men extend over a greater range of frequency.

In 1934 a new dimension in control and "set" was introduced into the word association technique. Maller presented a test in which the subject was given a choice of two response words for each stimulus word presented to him. One of these Maller labeled "normal", the other "abnormal". The subject was asked to choose the word he thought was best associated with the stimulus word.

It should be noted that the response words presented as choices were aprioristically and not empirically derived, the "abnormal" responses being "loaded" with pathological content. In effect, the rationale (and transparency and possibility of malingering) of the test is very similar to that of the "Cornell Index." This can be clearly seen from examining the items in Table I.
TABLE I
TYPICAL STIMULUS WORDS AND RESPONSE CHOICES
FROM MALLER’S CONTROLLED ASSOCIATION TEST

<table>
<thead>
<tr>
<th>Stimulus Word</th>
<th>Response Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>lose</td>
<td>find</td>
</tr>
<tr>
<td>food</td>
<td>stomach</td>
</tr>
<tr>
<td>thoughts</td>
<td>ideas</td>
</tr>
<tr>
<td>can't</td>
<td>concentrate</td>
</tr>
<tr>
<td>feel</td>
<td>useless</td>
</tr>
<tr>
<td>unhappy</td>
<td>no</td>
</tr>
<tr>
<td>left</td>
<td>home</td>
</tr>
<tr>
<td>talked</td>
<td>spoke</td>
</tr>
<tr>
<td>raw</td>
<td>deal</td>
</tr>
</tbody>
</table>

Nevertheless Maller reported that in disordered Mental Hospital patients there was a higher than average percentage of "abnormal" responses and he reconfirmed the previously found "communality of responses" among normal subjects. However, he gave no real information on validity and reliability.

In 1946 Malamud supplied for this deficiency and presented the results of an extended validity study on Maller's test. He gave the test to 150 mental hospital patients and to one hundred and fifty normal subjects. He reported correct identification of sixty-three percent of the psychotics with "misses" of twenty-four percent for the normals. The split-half reliability of .91 was derived from the combined scores of both groups. Malamud felt that with a proper system of weights the discrimination was sufficient for individual clinical diagnosis and that further research
would further refine it as a diagnostic instrument.

A year later Crown (1947) using the fifty most discriminating stimuli from Maller's list of 200 stimuli constructed his "Word Connexion List". His purpose was to validate the test as a measure of neuroticism. He administered the test to two hundred neurotics and to two hundred normal subjects. Although the mean scores for the two groups of neurotics and normals were statistically different, he found that among the neurotics there were so many "misses" that he had to admit "the screening efficiency of the Word Connexion List is probably not good enough for it to have more than suggestive value in individual work." (1952) In the same report he mentioned that he found that his test did not discriminate psychotics from normals.

It will be noted that for the past fifty years with the exception of the Maller list and its modifications (a list which was not empirically derived) all research in the area of word association has used as its basis the list of stimuli of Kent-Rosanoff and their frequency tables. Recently there has been an awareness that because of the influence of local and cultural meanings on associative responses the normative data originally gathered for this list may no longer be applicable to subjects being investigated today. Consequently in 1954 Russell and Jenkins administered the one hundred stimulus words from the original Kent-Rosanoff test to 1008 normal college subjects and presented new frequency norms (based on the absolute frequency of response by all the subjects to each response) for each stimulus word. Although it is difficult to compare the diverse
populations of Kent-Rosanoff and Russell-Jenkins one finds exactly fifty percent of the top frequency words on the former list replaced by different words on the latter list. Also one notes that for the top frequency responses there is almost double the amount of communality of that found on the Kent-Rosanoff list. Unfortunately, the work of Russell and Jenkins has been limited to the college population, there being no recent well-standardized sample of the general population for "free" association responses.

The other major updating of the Kent-Rosanoff data is that connected with the Loyola Language Study. This research, using eighty of the stimulus words from the Kent-Rosanoff list continues to reconfirm the hypothesis of "communality of response" -- with a higher communality among women. However, as has been noted, the L.L.S. is significantly different from all other previous research in the free association area, in that the basic "set" is different. Instead of responding with the first response that comes to mind the subject is asked to consider what most people would be likely to respond to each of the eighty stimulus words and then give that response. The assumption, as has been mentioned above, is that normal subjects would be able to do this significantly better than psychotics (particularly schizophrenics) and that conversely the psychotic would not be able to conform himself to this "set" and will do significantly poorer in giving common responses.

The investigation began with Snider and Johnson (1954) who administered the test to a normal Boston population of four hundred males and four hundred
females stratified according to age and education in the general population. This constituted the normative group. Later a similar sampling was established for the Chicago area by Stanek (1956) and for the Seattle area by Guppy (1959). Then State Mental Hospital patients from the first two areas were given the same test (An abnormal group from the Seattle area has not yet been studied.) -- fifty male and fifty female patients in Boston and fifty-six male and fifty-three female patients in Chicago. These patients were matched on a person to person basis, according to age and education, with normals from the Boston area by Snider and Johnson (1954) and from the Chicago area by Del Vecchio (1957).

It was found that, using any of three methods of scoring, the Loyola Language Study significantly distinguishes between Schizophrenics and normals. (Herr, 1957)

It was further discovered that age affects scores, older people scoring poorer. However, this tendency is counteracted by education. (Stanek, 1956). It has also been demonstrated that intelligence, apart from education does not significantly affect scores (Smola, 1956; Stewart, 1956) nor does area of residence. (Guppy, 1959)

From the proceeding it is seen that the Loyola Language Study is the most recent and most thoroughly researched of the word association tests, in terms of reliability and validity and of a large and well standardized normative population. It has the further advantage of lack of transparency and of threat to the subject, and, of all the tests based on the hypothesis that deviation from communality of response is an indicator of pathology,
it shows the greatest degree of communality among normal subjects.

(Trainor, 1958)

So far in this review nothing has been said about multiple choice versions of word association tests. In 1946 Malamud had high hopes for such a test. At that time he stated:

A multiple choice word test bearing a multi-scoring character similar to that of the MMPI or the Strong Vocational Aptitude test might be devised for use both in the clinical and industrial fields. By means of item analyses of carefully defined clinical and control groups a variety of scoring keys might be derived for the various diagnostic syndromes. A single administration of such a test would yield a clinically useful profile without the necessity of gaining the subjects willingness to reveal himself directly, as is necessary in the MMPI, for example. (1946, 43)

Despite these hopes nothing has come of them. As has been seen, the Maller-Malamud-Crown two-choice tests cannot be strictly classed as multiple choice tests, in the sense that that word is usually employed.

A survey of the literature in the field indicates that, besides the above tests, there have only been two unrelated and independent studies in the area of multiple choice word association tests.

The first is that of Karwoski and Berthold. (1945) Their main interest was in the reliability of a "free" association "set" -- the first word that comes to mind -- as opposed to a multiple choice "set" in relation to classification of responses in five categories: essential similarity, general identification, contingent identification, essential opposition, and contiguity. In terms of classification in these categories
it was found that the "free" association method is more reliable.

The other study is that of Buchwald (1957), whose concern was with the difference between auditory and visual presentation of stimuli in both "free" associational and multiple choice versions. However, in the multiple choice version, the responses were presented before the stimuli. Further, although the population was fairly large, the author admits that results are inconclusive because of a great deal of overlapping of the tested subgroups.

The search of the literature uncovered no other study of the relation between a "free" association test and a multiple choice version of the same. No studies at all were found which investigated the relation between a semi-controlled test like the Loyola Language Study and a multiple choice version of the same.
The procedure employed in designing a multiple choice version of the Loyola Language Study was both logical and empirical. It was decided to first construct a multiple choice test, give it to a small sample who had already been given the Loyola Language Study and to compare and analyze the results, using this analysis and comparison as a basis for construction of other versions.

The first test which was constructed -- on the basis of assumptions stated in the Introduction -- was a five choice test, containing two categories of response words among its five choices.

The first category of response words were chosen from the list of responses and response frequencies for each of the eighty stimulus words of the Loyola Language Study recorded by Stanek (1956) for a normal population of four hundred men and 400 women from the Chicago area.

From this list were selected those responses to each of the eighty stimuli whose summed frequency for both the male and female population was at least fifty percent, i.e., at least fifty percent of the men and at least fifty percent of the women gave this response.

In some cases a single response had such a frequency, e.g., for the stimulus "HUNGRY" the response "food" was given by fifty percent of the men
and fifty percent of the women. In other cases, two or three responses had such a summed frequency, e.g., for the stimulus "TABLE" the two responses "food" and "chair" have a summed frequency of fifty percent of all responses given; for the stimulus "CHEESE" the three responses "food", "crackers", "mouse" have a summed frequency of fifty percent of all responses.

For fifty-three out of the eighty stimuli, either one, two, or three responses equalled or slightly exceeded the summed frequency of fifty percent of all responses given by the normative population. For these stimuli, either one, two, or three responses were presented from this category. With the other twenty-seven stimuli it was not possible to obtain anywhere near the frequency of fifty percent with these three responses. For example, for the stimulus "TROUBLE", nine responses constitute the summed frequency of at least fifty percent of all responses for the female population of four hundred subjects; and fifteen responses were needed to constitute at least fifty percent of all responses for the male population of four hundred normal subjects. With these twenty-seven stimuli the three responses having the highest frequencies were used as choices.

The first category then, was made up of one, two, or three top frequency responses. The second category was composed of those responses having the lowest frequency, as recorded for the Stanek population, i.e., "singleton" responses. Since Stanek did not record the actual singleton responses given by the normative population, but only the number of such
responses given to each stimulus word, the Russell-Jenkins' list described above (1954) was used as a source of possible "singleton responses." Those possible responses to be used as choices, i.e., "singleton" responses on the Russell-Jenkins' lists, were then checked against the Stanek list in order to eliminate those responses which had a higher than singleton frequency on the Stanek list. After this elimination, singleton responses were chosen at random for each of the stimulus words, and used as the remaining choices on this first multiple choice test. A sample of this first design is given in Table II.

**TABLE II**

SAMPLE OF STIMULI AND RESPONSES ON FIRST DESIGN OF MULTIPLE CHOICE TEST

<table>
<thead>
<tr>
<th>BED:</th>
<th>sex sleep sick warmth twin</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWEET:</td>
<td>sacharrine dreams innocent sugar candy</td>
</tr>
<tr>
<td>COLD:</td>
<td>ice hot winter bitter angry</td>
</tr>
</tbody>
</table>

: top frequency responses

"singleton responses

This test was given after an eight week interval to fifteen college freshmen women and to fifteen college freshmen men, who had been given the Loyola Language Study. The two sets of tests were scored according to
three methods (which will be described in detail below) and comparisons and correlations were made between the scores achieved on the Loyola Language Study and the scores obtained on the multiple choice version. It was found that there was no significant correlation between the two tests. Further while the Loyola Language Study scores were normally distributed, the scores on the multiple choice test were extremely bunched at the upper end of its range, i.e., with an extreme negatively skewed distribution.

An item analysis showed why. Confronted with the two categories of choices of which this test was composed, i.e., top frequency responses and "singleton" responses, the choice was too easy and the subjects had no difficulty spotting the response most people would give, when confronted with the eighty stimuli.

In only one respect did this pilot study agree with previous findings, i.e., that the men showed less communality of response than did the women. This agrees with the finding of Stanek (1956), and that of Noh and Guilford (1930).

Utilizing the data provided by this small pilot study, it was decided to construct three new multiple choice versions of the Loyola Language Study, each with an increasing amount of difficulty for the subject -- all of them more difficult than the version constructed for the pilot study. Although these versions would be constructed for both men and women, their administration would be limited to college men. The three designs, which will be called "Design X", "Design Y", and "Design Z" were constructed in
the following manner.

In all three designs each of the eighty stimulus words of the Loyola Language Study are presented. To the right of these stimulus words are given multiple choices, one of which is to be circled. E.g., for the stimulus word "King" on Design X:

**KING**: royalty queen majestic reign

In Design X, four choices are given; in Design Y, five choices are given, and in Design Z, seven choices are given. The choices presented are chosen from the list of responses and response frequencies for each of the eighty stimuli of the Loyola Language Study recorded by Stanek for the normative Chicago population of eight hundred recorded by Stanek. (1956)

In Design X (four choices) three categories of choices are presented for each of the eighty stimulus words.

**Category 1**: One of the four choices presented was selected from the top responses, i.e., the response having the single highest frequency to each of the eighty stimuli as recorded for the Stanek population. For sixty-six out of the eighty stimuli the top response is identical for both the four hundred male and four hundred female population. For the remaining fourteen stimuli, the top two responses are identical for both the male and female normative population, but inversely. E.g., to the stimulus "BUTTERFLY" the top male response is "insect" with the response "flower" having the second highest frequency; to the same stimulus "BUTTERFLY" the top female response is "flower" with "insect" having the second highest frequency. For these fourteen stimuli the one choice presented is therefore
selected from the response having the first or second highest frequency as recorded by Stanek for the Chicago normative population. Seven male top responses are used as choices -- hence seven female second highest frequency responses; and seven female top responses are used as choices -- hence seven male second highest frequency responses.

Category 2: The second of the four choices is a "below the median" response, i.e., that response given by both the men and women of Stanek's Chicago normals which is below and nearest to the median.

Category 3: The two remaining choices of the four were selected from those responses having the lowest frequency to each of the eighty stimuli, as given by Stanek's normative population of eight hundred, i.e., singleton responses. As noted above, Stanek did not record "singleton" responses on his lists of responses for each of the eighty stimuli. However, instead of selecting singleton responses from the Russell-Jenkin's lists and cross-checking them with the greater than singleton responses, as was done for the pilot study, it was decided to obtain the original Stanek test booklets in order to be able to use singleton responses actually used by the normative population.

This was done for both the male and female booklets. It was felt that a sample of approximately one third of the booklets (one hundred and thirty-five male booklets and one hundred and thirty-five female booklets -- a total of 21,600 responses) would provide a sufficiently representative sample of the "singletons" to be used as choices. The sample was proportioned to Stanek's twelve categories of age and education based on the
1950 census for the Chicago Metropolitan area. After these singleton responses had been recorded for each of the eighty stimulus words, each of the male singleton responses was cross-checked against Stanek's lists of female responses. Those male responses which had a frequency of more than one for the females was eliminated. E.g., to the stimulus word "THIEF" a male singleton response is "money"; but to the same stimulus word "THIEF", eleven females responded "money". Likewise for the stimulus word "HEALTH" a male singleton response was "vitamins", whereas six females responded with "vitamins". Thus the responses "money" and "vitamins" were eliminated as possible choices. The same procedure was followed for the female singleton responses.

Next, those responses which are singletons for both men and women were eliminated. E.g., if only one man and only one woman gave the response "pan" to the stimulus word "BED", then the response "pan" was eliminated as a possible choice. This was to ensure, as far as possible, that a singleton response could be defined as "that response to a stimulus word which was given only once by Stanek's normative population of eight hundred men and women."

Finally, those singletons were eliminated, which were very similar in form to high frequency responses. E.g., for the stimulus word "WHITE", "blackness" was a singleton response, but because of its close similarity in form to the response "black" which has a response frequency of 84 for men and 61 for women, it was eliminated as a possible choice.

From the list of singleton responses obtained in the above manner,
two were chosen at random and used as the final two choices in Design X. This distribution is illustrated in Table III.

In Design Y (five choices) the same three categories as were used in Design X constitute the five choices given for each of the eighty stimulus words.

Category 1: Is identical with Category 1 of the previous Design X, i.e., it consists of one top frequency response.

Category 2: Is the same as Category 2 of the previous Design X, in that it contains a "below the median" response, i.e., that response given by both men and women of the Stanek normative population which is below and nearest to the median. In addition it includes another response, i.e., that response given by both men and women of Stanek's normative population which is below and next nearest to the median.

Category 3: Is the same as Category 3 of the previous Design X, i.e., the two choices presented are singleton responses, selected in the manner described above.

In Design Z (seven choices) the same three categories as were used in Design X and Design Y were employed to constitute the seven choices given for each of the eighty stimulus words.

Category 1: Was the same as Category 1 of Design X and of Design Y, i.e., a top frequency response.

Category 2: Was the same as Category 2 of Design Y with the addition of a third "below the median" response. Therefore, to compare: in Category 2 of Design X there was one "below the median" response, i.e.,
that response given by both the men and women of Stanek's Chicago population of eight hundred normals, which is below and nearest to the median. In Category 2 of Design Y a second "below the median" response was added, i.e., that response given by both men and women of Stanek's normal population which is below and next nearest to the median. In Category 2 of Design Z (this design) a third "below the median" response will be added, i.e., that response given by both the men and women of Stanek's normative population which is below the median and next nearest to the above two responses.

Category 3: Was the same as Category 3 of Design X and Design Y with the addition of another singleton response, i.e., in this design three singleton responses will be presented as choices.

As can be seen from the above, all three designs were constructed so as to be more difficult than the preliminary design constructed for the pilot study, a design which proved too easy. In that early design only two categories of responses were presented, i.e., top frequency responses and singleton responses. Further, the subject was given a choice of one, two, or three top frequency responses. In the later designs the subject is presented with only one top frequency response. Besides limiting his choice, it was made progressively more difficult by the addition of "below the median" responses, one in Design X, two in Design Y, and three in Design Z. Further progression of difficult was ensured by the progressive addition of singleton responses. In Table III a sample of stimulus words and responses choices for each Design are given, including a sample of the preliminary design used in the pilot study, so that comparisons can be made.
### TABLE III

**COMPARISON OF NATURE OF CHOICES GIVEN IN PRELIMINARY (PILOT STUDY) AND LATER MULTIPLE CHOICE DESIGNS:**

**DESIGN X, DESIGN Y, DESIGN Z**

<table>
<thead>
<tr>
<th>Sample of Preliminary Design:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B</strong>ED: sex sleep sick warmth twin</td>
<td></td>
</tr>
<tr>
<td><strong>SWEET:</strong> sacharrine dreams innocent sugar candy</td>
<td></td>
</tr>
<tr>
<td><strong>COLD:</strong> ice hot winter bitter angry</td>
<td></td>
</tr>
<tr>
<td><strong>Sample of Design X:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>K</strong>ING: royalty queen majestic reign</td>
<td></td>
</tr>
<tr>
<td><strong>OCEAN:</strong> swim shore ship water</td>
<td></td>
</tr>
<tr>
<td><strong>SICKNESS:</strong> nurse distress doctor operation</td>
<td></td>
</tr>
<tr>
<td><strong>Sample of Design Y:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>B</strong>ED: sleep rest slumber enjoyment sheets</td>
<td></td>
</tr>
<tr>
<td><strong>SWEET:</strong> taste heart sour innocent sugar</td>
<td></td>
</tr>
<tr>
<td><strong>COLD:</strong> snow winter metal weather numb</td>
<td></td>
</tr>
<tr>
<td><strong>Sample of Design Z:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>SALT:</strong> season element pepper peanuts taste necessary spice</td>
<td></td>
</tr>
<tr>
<td><strong>CARPET:</strong> thick plush home soft house rug beater</td>
<td></td>
</tr>
<tr>
<td><strong>HEAVY:</strong> cumbersome lead lift big weight sorrow bricks</td>
<td></td>
</tr>
</tbody>
</table>

---

- **Category 1:** Top frequency responses
- **Category 2:** "Below the median" responses
- **Category 3:** Singleton responses.
From the Chicago normative population of four hundred men and four hundred women, standard z-scores were established, with an arbitrary mean standard score set at 20 with a standard deviation of 10. In Tables IV, V, and VI are given z-scores for both men and women based on the Chicago population of Stanek for five of the eighty stimulus words used on the Loyola Language Study and in the three multiple choice versions, Design X, Design Y, and Design Z. The possible z-scores for all responses given for these five stimulus words are presented.

Below these stimulus words and the possible z-scores are presented the response words which have been used as choices in the three multiple choice designs, i.e., Design X (four choices), Design Y (five choices) and Design Z (seven choices).

The scores that would be given to these responses which were used as choices, if they were scored by the standard z-scoring method, are circled, so that one can see graphically the position of these responses, in terms of responses given by the normative sample. One again can see the progressive difficulty of choice from Design X, through Designs Y and Z.
### TABLE IV

POSSIBLE Z-SCORES FOR RESPONSES TO 5 STIMULUS WORDS ON THE LOYOLA LANGUAGE STUDY (STANEK - CHICAGO MEN AND WOMEN) AND POSITION OF FOUR CHOICES OF DESIGN X ON Z-SCORE TABLES

<table>
<thead>
<tr>
<th>HEAVY</th>
<th>ANGER</th>
<th>TABLE</th>
<th>SALT</th>
<th>CARPET</th>
</tr>
</thead>
<tbody>
<tr>
<td>F 8</td>
<td>M 8</td>
<td>F 10</td>
<td>M 12</td>
<td>F 12</td>
</tr>
<tr>
<td>13 17</td>
<td>11 23</td>
<td>23 20</td>
<td>22 18</td>
<td>14 17</td>
</tr>
<tr>
<td>14 20</td>
<td>23 24</td>
<td>25 22</td>
<td>24 19</td>
<td>29 25</td>
</tr>
<tr>
<td>25 23</td>
<td>25 24</td>
<td>27 25</td>
<td>25 20</td>
<td>31 29</td>
</tr>
<tr>
<td>27 25</td>
<td>26 26</td>
<td>30 28</td>
<td>29 27</td>
<td>32 30</td>
</tr>
<tr>
<td>28 27</td>
<td>27 27</td>
<td>33 30</td>
<td>30 27</td>
<td>33 32</td>
</tr>
<tr>
<td>29 29</td>
<td>28 28</td>
<td>34 32</td>
<td>31 28</td>
<td>34 33</td>
</tr>
<tr>
<td>31 30</td>
<td>29 29</td>
<td>35 33</td>
<td>32 29</td>
<td>35 34</td>
</tr>
<tr>
<td>32 31</td>
<td>30 30</td>
<td>37 34</td>
<td>33 30</td>
<td>38 35</td>
</tr>
<tr>
<td>33 32</td>
<td>34 31</td>
<td>37 34</td>
<td>36 31</td>
<td>38 35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>weight</th>
<th>mad</th>
<th>chair</th>
<th>pepper</th>
<th>rug</th>
</tr>
</thead>
<tbody>
<tr>
<td>lead</td>
<td>fight</td>
<td>furniture</td>
<td>taste</td>
<td>soft</td>
</tr>
<tr>
<td>cumbersome</td>
<td>control</td>
<td>hungry</td>
<td>necessary</td>
<td>thick</td>
</tr>
<tr>
<td>sorrow</td>
<td>bitter</td>
<td>round</td>
<td>peanuts</td>
<td>plush</td>
</tr>
</tbody>
</table>

---

**Category 1:** Top frequency response

**Category 2:** "Below the median" response

**Category 3:** "Singleton" responses
<table>
<thead>
<tr>
<th>HEAVY</th>
<th>ANGER</th>
<th>TABLE</th>
<th>SALT</th>
<th>CARPET</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>8</td>
<td>F</td>
<td>10</td>
<td>F</td>
</tr>
<tr>
<td>M</td>
<td>6</td>
<td>M</td>
<td>7</td>
<td>M</td>
</tr>
<tr>
<td>13</td>
<td>17</td>
<td>23</td>
<td>22</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>20</td>
<td>25</td>
<td>24</td>
<td>29</td>
</tr>
<tr>
<td>25</td>
<td>23</td>
<td>27</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td>27</td>
<td>25</td>
<td>30</td>
<td>28</td>
<td>32</td>
</tr>
<tr>
<td>28</td>
<td>27</td>
<td>33</td>
<td>30</td>
<td>33</td>
</tr>
<tr>
<td>29</td>
<td>29</td>
<td>34</td>
<td>32</td>
<td>34</td>
</tr>
<tr>
<td>31</td>
<td>30</td>
<td>35</td>
<td>33</td>
<td>35</td>
</tr>
<tr>
<td>32</td>
<td>31</td>
<td>36</td>
<td>34</td>
<td>38</td>
</tr>
<tr>
<td>33</td>
<td>32</td>
<td>37</td>
<td>35</td>
<td>38</td>
</tr>
<tr>
<td>37</td>
<td>33</td>
<td>36</td>
<td>37</td>
<td>37</td>
</tr>
</tbody>
</table>

- **Category 1:** Top frequency response
- **Category 2:** "Below the median" responses
- **Category 3:** "Singleton" responses
TABLE VI
POSSIBLE Z-SCORES FOR RESPONSES TO 5 STIMULUS WORDS ON THE LOYOLA LANGUAGE STUDY (STANEK - CHICAGO MEN AND WOMEN) AND POSITION OF SEVEN CHOICES OF DESIGN Z ON Z-SCORE TABLES

<table>
<thead>
<tr>
<th></th>
<th>HEAVY</th>
<th>ANGER</th>
<th>TABLE</th>
<th>SALT</th>
<th>CARPET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td>weight</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>lead</td>
<td>13</td>
<td>17</td>
<td>23</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>big</td>
<td>25</td>
<td>23</td>
<td>25</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>lift</td>
<td>27</td>
<td>25</td>
<td>26</td>
<td>24</td>
<td>29</td>
</tr>
<tr>
<td>sorrow</td>
<td>28</td>
<td>27</td>
<td>27</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>bricks</td>
<td>29</td>
<td>29</td>
<td>28</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>30</td>
<td>29</td>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>31</td>
<td>30</td>
<td>30</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>32</td>
<td>34</td>
<td>31</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>33</td>
<td>30</td>
<td>30</td>
<td>36</td>
</tr>
</tbody>
</table>

- **Category 1**: Top frequency response
- **Category 2**: "Below the median" responses
- **Category 3**: "Singleton" responses
As can be seen from a comparison of the page of instructions preceding the Loyola Language Study and those instructions preceding Designs X, Y, and Z, all of which are included in the Appendices, the directions for the multiple choice designs are verbatim the same as those of the Loyola Language Study with the minimum necessary addition of multiple choice directions, e.g., with regard to circling one's choice. The pilot study indicated that there was no difficulty by the subjects (N:30) in comprehending the multiple choice directions. All subjects underlined one choice -- and only one choice -- for each of the eighty stimulus words.
CHAPTER IV

PROCEDURE AND RESULTS

The present form of the Loyola Language Study was administered to one hundred and ninety-five male college freshmen from the Chicago area. All of the subjects were volunteers, who were to receive credit in their General Psychology courses for participating in this experiment. All of the prospective volunteers were told beforehand that they would have to stay at least thirty-five minutes for the experiment. (When they arrived they were told this again and all of the subjects remained at least this time.) This was done to eliminate rushing through the test and doing it superficially. The subjects were divided into five groups, three groups on one day, two groups the next day: Group 1, N: 44; Group 2, N: 31; Group 3, N: 42; Group 4, N: 40, Group 5, N: 38. A large lecture hall was used, so that the subjects could be physically separated. The subjects were asked to keep silence during the testing and they did so. The directions on the Loyola Language Study were read and re-read to the group. In addition, their attention was called to a sign on the blackboard: "Please write clearly. Please respond to all the stimulus words." At the end of the test a random sampling of twenty-percent were asked in writing about the instructions. All correctly gave the right set, i.e., the one word the greatest number of people would be most likely to think of when they see or hear the word.
Only six out of the one hundred and ninety-five failed to respond to all of the stimuli.

So that a comparison of the amount of time taken by subjects on the Loyola Language Study could be made with the amount of time taken on the multiple choice versions, the subjects were told: "At the end of the test, put down the time it took you. Note, however, this is the least important instruction. Take all the time you need; you have all afternoon." The second sentence was added so that the subjects would not stress the time element. Unfortunately, however, because of this lack of stress on the time element, a great number of the subjects did not mark down the time taken. Among those who did there is strong reason to believe that the times given are often inaccurate. This is indicated by the fact that many put down widely varying starting times, although all subjects started at the same time. It was further noted that many subjects did not have watches and guessed at the time. So, no experimentally valid statement about the time taken for the Loyola Language Study can be made. The experimenter's observation is that each of the groups took between twenty-five and thirty-five minutes. This is verified by Trainor (1958) who gives a mean time of thirty minutes for a similar group. Profiting from this experience, a more accurate record of time was kept for the multiple choice tests. But any comparison between time taken on the Loyola Language Study and on the multiple choice versions will be weakened by the above inaccuracy and will have to be based -- on the side of the Loyola Language Study -- on the above observation of between twenty-five and thirty-five minutes, and Trainor's confirmatory data.
All Loyola Language Tests were then corrected by the Herr-Rimoldi median method (Herr, 1957), i.e., any response falling within the upper fifty percent of the responses given by the Chicago male normative sample of Stanek was credited with a score of 1. Any response not falling within the upper fifty percent was given a score of 0. The correlation between this method of scoring the Loyola Language Study and the standard z-scoring method is .93. And with regard to the abnormal population both methods yield significant differences between patients and normals which are significant beyond the .001 level of confidence. (Herr, 1957)

The booklets -- hence the subjects -- were then divided into three matching groups on the basis of these median scores. These will be called Group X, Group Y, and Group Z. Each group had a mean score of 39.9, a standard deviation of 8.94 and a range of 21 to 57.

After an interval of 11 weeks the subjects were reassembled. Only two of the original subjects failed to report for a retest. To those who composed Group X (N:61) Multiple Choice Design X (4 choices) was given. To those who composed Group Y (N:61) Multiple Choice Design Y (5 choices) was given. To those who composed Group Z (N:61) Multiple Choice Design Z (7 choices) was administered. The conditions of administration were identical to the original testing situation. The one exception was the written instructions on time which can be seen in Appendices II, III, and IV.

1Whereas a better method of matching would be by means of some norm not biased in either direction by reason of the currently used Loyola Language instructions and method of testing used at Loyola. This, however, will have to await further research.
The Multiple Choice Design X (four choices) tests of Group X, the Multiple Choice Design Y (5 choices) tests of Group Y, and the Multiple Choice Design Z (7 choices) tests of Group Z were then scored according to the median method described above.

Next, all three sets -- Group X on the Loyola Language Study and on Multiple Choice Design X; Group Y on the Loyola Language Study and on Multiple Choice Design Y; and Group Z on the Loyola Language Study and Multiple Choice Design Z -- were scored according to two additional methods. The first of these other methods was that method of scoring which was found to be most discriminating between normals and abnormals. If a high correlation were found between the Loyola Language Study and one of the multiple choice designs when scored by the median method described above, would a similarly high correlation also be found between the Loyola Language Study and one of the multiple choice designs, when scored by this other method which is highly discriminating. The second of these additional methods, was a method using independent norms for scoring, so that correlations would not be merely in terms of scores derived from the normative population from which the above two methods of scoring were developed. These two additional methods of scoring are described herein.

The first of these methods was to score all three sets, using the standard z-scores, on the twenty-five most discriminating items. These items are those which it was found contributed most to the difference in scores between the Chicago normals and the Chicago patients (Herr, 1957) and which were found to have a higher screening efficiency than all eighty
items scored by the standard z-scoring method. It will be noted that only 11 of these items are common for both men and women from Chicago. These items are presented in Table VII.

**TABLE VII**

**STIMULUS WORDS FOR CHICAGO MALES (M) AND FEMALES (F) WITH HIGHEST SCREENING EFFICIENCY BETWEEN CHICAGO NORMALS AND CHICAGO PATIENTS**

<table>
<thead>
<tr>
<th>soldier (F)</th>
<th>sweet (M)</th>
<th>moon (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>head (M, F)</td>
<td>stomach (M, F)</td>
<td>bread (M, F)</td>
</tr>
<tr>
<td>dark (M)</td>
<td>soft (M)</td>
<td>whistle (M, F)</td>
</tr>
<tr>
<td>loud (F)</td>
<td>cold (M)</td>
<td>needle (M)</td>
</tr>
<tr>
<td>joy (M)</td>
<td>window (M, F)</td>
<td>hand (M)</td>
</tr>
<tr>
<td>rough (F)</td>
<td>scissors (M)</td>
<td>thief (M)</td>
</tr>
<tr>
<td>high (F)</td>
<td>foot (M, F)</td>
<td>dream (F)</td>
</tr>
<tr>
<td>king (M)</td>
<td>doctor (M, F)</td>
<td>trouble (F)</td>
</tr>
<tr>
<td>sleep (F)</td>
<td>wish (M)</td>
<td>religion (F)</td>
</tr>
<tr>
<td>black (M)</td>
<td>house (F)</td>
<td>bed (F)</td>
</tr>
<tr>
<td>hammer (F)</td>
<td>sickness (M)</td>
<td>tobacco (M, F)</td>
</tr>
<tr>
<td>table (M, F)</td>
<td>mountain (M, F)</td>
<td>citizen (F)</td>
</tr>
<tr>
<td>thirsty (M, F)</td>
<td>stove (F)</td>
<td>butter (M)</td>
</tr>
</tbody>
</table>
The second of these methods employed independent norms for scoring so that correlations would not be merely in terms of scores derived from the normative population from which the above two methods of scoring were developed. Since Russell and Jenkins also used the Kent-Rosanoff list of stimuli for developing their normative sample -- a sample of 1008 college freshmen -- the absolute recorded frequencies listed by them were used as absolute scores in the scoring of all three sets. To avoid astronomical scores, the Russell-Jenkins' frequencies were divided by ten. Thus, for example, for the stimulus word "WISH" the top response, "want" has a listed frequency of 124. This would be scored as 12 points. All three sets were scored by this method.

Henceforward these three methods of scoring, for convenience sake, will be referred to as Method I (the median norms); Method II (the twenty-five most discriminating items), and Method III (the Minnesota norms).

After all 90,000 items had been scored, a Pearson Product Moment Correlation was run to determine the correlation between the scores on the Loyola Language Study and the Multiple Choice Design given to the same group. A correlation was worked out for each group, using the scores from the three methods of scoring. Thus, for Group X, correlations were computed between the groups' scores on the Loyola Language Study and the same group's scores on Multiple Choice Design X -- both having been scored by Method I, Method II, and Method III. Thus between the Loyola Language Study, given to Group X and Design X, which was later given to the same Group X, three Product Moment Correlation Coefficients were computed, one
for each method of scoring. In a like manner, three correlation coefficients were obtained -- one for each method of scoring -- between the scores that Group Y earned on the Loyola Language Study and the scores that the same group later on earned on Multiple Choice Design Y. The same was done for Group Z.

Lest someone might object that the scores obtained by the three methods, described above, represent variables of a qualitative rather than a quantitative type, and hence non-parametric treatment should be given to them, Spearman's rho ($r_s$) was also computed for each of the three groups, for each of the three scoring methods.

These correlations -- the Pearson ($r_{xy}$) and the Spearman rho are listed in Tables VIII, IX, and X. As will be noted from the Tables, except for two instances, there was never more than two points difference between the Pearson and Spearman correlations.

The standard error of coefficient was obtained for each of the Pearson correlations, according to the formula given by Guilford (1956):

$$s_r = \frac{1 - r^2}{\sqrt{N - 1}}$$

To determine whether the obtained coefficients of correlation were significantly different from zero, a $t$ Test was used, according to the formula (Guilford, 1956) of Fisher:

$$t = r \sqrt{\frac{N - 2}{1 - r^2}}$$
TABLE VIII

<table>
<thead>
<tr>
<th>Method I Median N</th>
<th>Method II 25 Items</th>
<th>Method III Minnesota N</th>
</tr>
</thead>
<tbody>
<tr>
<td>rho</td>
<td>.45</td>
<td>.25**</td>
</tr>
<tr>
<td>$r_{xy}$</td>
<td>.46</td>
<td>.27**</td>
</tr>
<tr>
<td>$\sigma_r$</td>
<td>±.10</td>
<td>±.12</td>
</tr>
</tbody>
</table>

TABLE IX

<table>
<thead>
<tr>
<th>Method I Median N</th>
<th>Method II 25 Items</th>
<th>Method III Minnesota N</th>
</tr>
</thead>
<tbody>
<tr>
<td>rho</td>
<td>.57</td>
<td>.38</td>
</tr>
<tr>
<td>$r_{xy}$</td>
<td>.58</td>
<td>.46</td>
</tr>
<tr>
<td>$\sigma_r$</td>
<td>±.08</td>
<td>±.10</td>
</tr>
</tbody>
</table>

* Except for the coefficients marked **, which are significant at the .05 level of confidence, all other coefficients are significant at the .01 level of confidence. For all rho's the standard error is ±.13.
**TABLE X**

SPEARMAN COEFFICIENT (rho), PEARSON COEFFICIENT (r<sub>xy</sub>) AND STANDARD ERROR OF PEARSON COEFFICIENT (σ<sub>r</sub>) BETWEEN SCORES OF GROUP Z
(N: 61) ON THE L.L.S. AND SCORES OF THE SAME GROUP Z ON DESIGN Z (7 CHOICES) ACCORDING TO EACH OF THREE METHODS OF SCORING

<table>
<thead>
<tr>
<th>Method I</th>
<th>Method II</th>
<th>Method III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>25 Items</td>
<td>Minnesota</td>
</tr>
<tr>
<td>Norms</td>
<td></td>
<td>Norms</td>
</tr>
<tr>
<td>rho</td>
<td>.51</td>
<td>.32***</td>
</tr>
<tr>
<td>r&lt;sub&gt;xy&lt;/sub&gt;</td>
<td>.49</td>
<td>.30***</td>
</tr>
<tr>
<td>σ&lt;sub&gt;r&lt;/sub&gt;</td>
<td>±.10</td>
<td>±.11</td>
</tr>
</tbody>
</table>

* Except for the coefficients marked ***, which are significant at the .05 level of confidence, all other coefficients are significant at the .01 level of confidence. For all rho's the standard error is ±.13.

No standard error of coefficient was obtained for the Spearman correlations. Guilford states that "there is no generally accepted formula for estimating the standard error of rho". (1956) McNemar (1959) and Siegel (1956) agree with this. Guilford suggests the possibility of using a formula for the standard error of coefficient correlation when the population value is hypothesized to be zero, i.e.,

\[ \sigma_r = \frac{1}{\sqrt{N - 1}} \]
For a (any) population of 61, the standard error of rho is computed to be .1290, and so any correlation obtained for Groups X, Y, or Z -- or any group of 61 subjects -- will be rho ±.13.

However, since the obtained Spearman rho's so closely approximate the Pearson's r's, it seems that the standard error of coefficients computed for the Pearson's r's are adequate and sufficient. These standard errors are presented in Tables VIII, IX, AND X.

Before an analysis of the results of these correlations is presented, it should be noted that correlation by itself is no criterion of interchangeability of these two techniques of testing, i.e., the Loyola Language Study "open end" technique and the more controlled multiple choice technique. If there would be a relatively high correlation between the scores obtained by the same group tested by these two techniques, while the means would turn out to be significantly different, a reliable transformation of scores would not be indicated. So data on correlation alone will only be of relative value in determining the interchangeability of the Loyola Language Study technique with the multiple choice technique. Further judgment will have to be withheld until a comparison of means is presented.

With this reservation in mind an analysis of the Pearson and Spearman correlations obtained between the Loyola Language Study scores and the scores obtained on Designs X, Y, and Z indicates the following facts:

First on the basis of all three methods of scoring the highest correlations are those between the scores obtained by Group Y on the Loyola Language Study and the scores obtained by the same Group Y on Design Y.
(five choices). For Method I (Median norms): $r_{xy} = .58$, $\rho = .57$; for Method II (25 critical items): $r_{xy} = .46$, $\rho = .38$; for Method III (Minnesota norms): $r_{xy} = .69$, $\rho = .68$. All are significant at the .01 level of confidence.

The next highest correlations are those between the scores obtained by Group Z on the Loyola Language Study and the scores obtained by the same Group Z on Design Z (7 choices). For Method I (Median scores): $r_{xy} = .49$, $\rho = .51$; for Method II (25 critical items): $r_{xy} = .30$, $\rho = .32$; for Method III (Minnesota norms): $r_{xy} = .65$, $\rho = .65$.

The lowest correlations are those obtained between the scores obtained by Group X on the Loyola Language Study and the scores obtained by the same Group X on Design X (4 choices). For Method I (Median Scores): $r_{xy} = .46$, $\rho = .45$; for Method II (25 critical items): $r_{xy} = .27$, $\rho = .25$; and for Method III (Minnesota norms): $r_{xy} = .58$, $\rho = .51$.

It should be noted that with the exception of the correlations between Method II scores (25 items) on the Loyola Language Study and Design X, and the correlations between Method II scores on the Loyola Language Study and Design Z, which are significant at the .05 level of confidence, all other correlations are significant at the .01 level of confidence.

So, if the discussion is limited to correlations, it can be said that Design Y (5 choices) most closely correlates with the Loyola Language Study, Design Z is the next closely correlated with the Loyola Language Study, and Design X is the least closely correlated with the Loyola Language Study. However, this comparison between correlations cannot be
very meaningful owing to the fact that none of the differences approximates significance. This fact can be seen from Tables VIII, IX, and X, by noticing that the greatest differences are still small in comparison to the standard error of the various correlations. (For all rho coefficients, it will be remembered that the standard error is $\pm 1.13$).

One also observes from Table XI that the correlation between each of the Multiple Choice Designs ($X$, $Y$, $Z$) and the Loyola Language Study is consistently highest for scoring Method III (Minnesota norms), lower for scoring Method I (Median norms), and lowest for scoring Method II (25 critical items.)
TABLE XI

COMPARISON OF 1, 2, AND 3, i.e.,
1. CORRELATION COEFFICIENTS BETWEEN SCORES OF GROUP X ON LOYOLA LANGUAGE STUDY AND ON DESIGN X (4 CHOICES)
2. CORRELATION COEFFICIENTS BETWEEN SCORES OF GROUP Y ON LOYOLA LANGUAGE STUDY AND ON DESIGN Y (5 CHOICES)
3. CORRELATION COEFFICIENTS BETWEEN SCORES OF GROUP Z ON LOYOLA LANGUAGE STUDY AND ON DESIGN Z (7 CHOICES)

<table>
<thead>
<tr>
<th>Method I</th>
<th>Method II</th>
<th>Method III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Norms</td>
<td>25 Items</td>
<td>Minn. Norms</td>
</tr>
<tr>
<td>Loyola Language Study</td>
<td>Loyola Language Study</td>
<td>Loyola Language Study</td>
</tr>
<tr>
<td>$r_{xy}$</td>
<td>$r_{xy}$</td>
<td>$r_{xy}$</td>
</tr>
<tr>
<td>Design X</td>
<td>.46 (.45)</td>
<td>.27 (.25)</td>
</tr>
<tr>
<td>Design Y</td>
<td>.58 (.57)</td>
<td>.46 (.38)</td>
</tr>
<tr>
<td>Design Z</td>
<td>.49 (.51)</td>
<td>.30 (.32)</td>
</tr>
</tbody>
</table>

Also computed for each group (X, Y, and Z) according to the three methods of scoring were the mean scores, to test the null hypothesis, i.e., that there was no significant difference between the means, for each group, on the Loyola Language Study and on each of the Multiple Choice Designs. E.g., was there a significant difference between the mean scores of Group X obtained on the Loyola Language Study as opposed to the mean scores obtained on Multiple Choice Design X -- for each of the three methods of scoring? These comparisons are presented in Tables XII, XIII, and XIV.
The standard errors of means were derived according to the formula given by Guilford (1956):

$$\sigma_M = \frac{\sigma}{\sqrt{N - 1}}$$

Using these data the standard error of difference between the means was obtained by the formula (McNemar, 1959):

$$\sigma_{dM} = \sqrt{\sigma^2_{M_1} + \sigma^2_{M_2} - 2r_{12}\sigma_{M_1}\sigma_{M_2}}$$

Having derived this and obtaining the differences between means, a critical ratio could be computed by the formula

$$t = \frac{DM}{\sigma_{dM}}$$

The results of the t tests are indicated at the bottom of Tables XII, XIII, and XIV. In the case of Group X, the null hypothesis is proven incorrect in every instance, i.e., by every method of scoring the mean scores obtained by Group X on the Loyola Language Study are significantly different from the mean scores obtained by the same Group X on Multiple Choice Design X (5 choices.)

For Group Y, for two methods of scoring -- Method I (Median norms) and Method III (Minnesota norms), there is no significant difference between the mean scores obtained by this same group on the Loyola Language Study and on Design Y (5 choices.) However, when scoring Method II (25 critical items) was used, there was a significant difference between the
TABLE XII

MEANS, STANDARD DEVIATIONS, STANDARD ERRORS OF MEANS, DIFFERENCES OF MEANS, STANDARD ERRORS OF DIFFERENCE BETWEEN MEANS AND CRITICAL RATIOS FOR DIFFERENCE BETWEEN MEANS FOR SCORES OF GROUP X (N:61) SCORED BY THREE METHODS (METHOD I - MEDIAN NORMS; METHOD II - 25 CRITICAL ITEM NORMS; METHOD III - MINNESOTA NORMS) ON THE LOYOLA LANGUAGE STUDY AND ON MULTIPLE CHOICE DESIGN X (4 CHOICES)

<table>
<thead>
<tr>
<th>Method I: Median Norms</th>
<th>Method II: 25 Critical Items</th>
<th>Method III: Minnesota Norms</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>39.9</td>
<td>509</td>
</tr>
<tr>
<td>σ</td>
<td>8.94</td>
<td>7.68</td>
</tr>
<tr>
<td>σ_M</td>
<td>1.153</td>
<td>0.990</td>
</tr>
<tr>
<td>D_M</td>
<td>81</td>
<td>76</td>
</tr>
<tr>
<td>σ_dM</td>
<td>1.127</td>
<td>11.96</td>
</tr>
<tr>
<td>C.R.</td>
<td>7.1872</td>
<td>6.3545</td>
</tr>
<tr>
<td>Significant Difference</td>
<td>P: .001</td>
<td>Significant Difference</td>
</tr>
</tbody>
</table>
TABLE XIII

MEANS, STANDARD DEVIATIONS, STANDARD ERRORS OF MEANS, DIFFERENCES OF MEANS, STANDARD ERRORS OF DIFFERENCES OF MEANS, STANDARD ERRORS OF DIFFERENCE BETWEEN MEANS AND CRITICAL RATIOS FOR DIFFERENCE BETWEEN MEANS FOR SCORES OF GROUP Y (N=61) SCORED BY THREE METHODS (METHOD I - MEDIAN NORMS; METHOD II - 25 CRITICAL ITEM NORMS; METHOD III - MINNESOTA NORMS) ON THE LOYOLA LANGUAGE STUDY AND ON MULTIPLE CHOICE DESIGN Y (5 CHOICES)

<table>
<thead>
<tr>
<th>Method I: Median Norms</th>
<th>Method II: 25 Critical Items</th>
<th>Method III: Minnesota Norms</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.L.S. Design Y</td>
<td>L.L.S. Design Y</td>
<td>L.L.S. Design Y</td>
</tr>
<tr>
<td>N</td>
<td>39.9</td>
<td>515</td>
</tr>
<tr>
<td>σ</td>
<td>8.94</td>
<td>7.68</td>
</tr>
<tr>
<td>σ_M</td>
<td>1.153</td>
<td>0.9909</td>
</tr>
<tr>
<td>D_M</td>
<td>1.5</td>
<td>30</td>
</tr>
<tr>
<td>σ_dM</td>
<td>.9950</td>
<td>9.899</td>
</tr>
<tr>
<td>C.R.</td>
<td>1.5075</td>
<td>3.0306</td>
</tr>
<tr>
<td>No Significant Difference</td>
<td>Significant Difference</td>
<td>No Significant Difference</td>
</tr>
<tr>
<td>P: .13</td>
<td>P: .001</td>
<td>P: .58</td>
</tr>
</tbody>
</table>
### TABLE XIV

Means, standard deviations, standard errors of means, differences of means, standard errors of difference between means, and critical ratios for difference between means for scores of Group Z (N: 61) scored by three methods (Method I - Median Norms; Method II - 25 Critical Item Norms; Method III - Minnesota Norms) on the Loyola Language Study and on multiple choice design Z (seven choices)

<table>
<thead>
<tr>
<th>Method I: Median Norms</th>
<th>Method II: 25 Critical Items</th>
<th>Method III: Minnesota Norms</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.L.S.</td>
<td>L.L.S.</td>
<td>L.L.S.</td>
</tr>
<tr>
<td>Design Z</td>
<td>Design Z</td>
<td>Design Z</td>
</tr>
<tr>
<td>M</td>
<td>39.9</td>
<td>518</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>511</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>8.94</td>
<td>8.60</td>
</tr>
<tr>
<td></td>
<td>84</td>
<td>69</td>
</tr>
<tr>
<td>( \sigma_M )</td>
<td>1.153</td>
<td>1.109</td>
</tr>
<tr>
<td></td>
<td>11.35</td>
<td>8.90</td>
</tr>
<tr>
<td>( D_M )</td>
<td>6.9</td>
<td>7</td>
</tr>
<tr>
<td>( \sigma_{d_M} )</td>
<td>1.140</td>
<td>11.96</td>
</tr>
<tr>
<td>C.R.</td>
<td>6.052</td>
<td>.58528</td>
</tr>
<tr>
<td>Significant Difference</td>
<td>No Significant Difference</td>
<td>Significant Difference</td>
</tr>
<tr>
<td>( P )</td>
<td>.001</td>
<td>.54</td>
</tr>
</tbody>
</table>
mean scores of Group Y obtained on the Loyola Language Study and the mean scores of the same Group Y obtained on Design Y.

For Group Z, scored according to Method I (Median norms) and Method III (Minnesota norms) there are significant differences between the means scores obtained by this same group on the Loyola Language Study and the mean scores obtained on Design Z (7 choices). But for Group Z, scored according to Method III (25 critical items) there is no significant difference between the mean scores obtained by the same group on the Loyola Language Study and the mean scores obtained on Design Z (seven choices.)

In summary, therefore, a comparison of mean scores indicates that there is no significant difference in mean scores between the Loyola Language Study and Design Y (5 choices), when scoring Method I (Median norms) and scoring Method III (Minnesota norms) are used; there is no significant difference in mean scores between Loyola Language Study and Design Z (7 choices) when scoring Method II (25 critical items) is employed.

Combining this data, obtained through a comparison of means, with that above on the correlation coefficients between the Loyola Language Study and Designs X, Y, and Z, it is clear that in terms of both amount of correlation and comparison of means, Design Y (5 choices) most closely approximates the Loyola Language Study, when scored according to Method I (Median norms) and Method III (Minnesota norms.) However, when scored according to Method II (25 critical items) there is no significant difference between the mean scores of Group Z obtained on the Loyola Language Study and those obtained on Design Z (seven choices); whereas there is a significant
difference between the scores obtained by Group Y on the Loyola Language Study and those obtained on Design Y (5 choices.) On the other hand, it was seen, that, using scoring Method II, the correlation coefficients between the scores of Group Y ($r_{xy} .46; r_s .48$) on the Loyola Language Study and on Design Y (5 choices) were considerably higher than those between the scores of Group Z ($r_{xy} .30; r_s .32$) on the Loyola Language Study and on Design Z (seven choices.) Should this higher correlation offset the lack of goodness of fit with regard to mean scores? It seems it should, when one notes that the significant difference between mean scores is attributable to the fact that the mean score on Design Y (5 choices) is 30 points better than the mean score on the Loyola Language Study. (For Method II $/25$ critical items, the lower the score the greater the degree of communality). If this were not combined with a substantial degree of correlation, it would mean nothing. But with the high degree of correlation, it may later allow for a greater differentiation between normal and abnormal subjects. This would have to be tested empirically. If the abnormal subjects are better differentiated there will be a gain; if they, too, obtain a mean score thirty points higher, nothing is lost.

However, it should be noted that the Multiple Choice Designs were constructed primarily with Method I (Median norms) in mind. It is these norms that give a Multiple Choice Design one of its primary advantages, i.e., ease and simplicity of scoring. And using this method of scoring, on all counts, Design Y (5 choices) most closely approximates the Loyola Language Study.
How do the correlation coefficients between the Loyola Language Study scores and the scores by the same group on Design Y (5 choices), as seen in Table XV, compare with the test-retest reliability for the Loyola Language Study?

**TABLE XV**

SPEARMAN RANK DIFFERENCE CORRELATION COEFFICIENTS AND PEARSON PRODUCT MOMENT CORRELATION COEFFICIENTS BETWEEN SCORES OF GROUP Y SCORED BY THREE METHODS ON THE LOYOLA LANGUAGE STUDY AND ON MULTIPLE CHOICE DESIGN Y (5 CHOICES)

<table>
<thead>
<tr>
<th>Method I</th>
<th>Method II</th>
<th>Method III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Scores</td>
<td>25 Items Scores</td>
<td>Minnesota Scores</td>
</tr>
<tr>
<td>$r_s$</td>
<td>.57</td>
<td>.48</td>
</tr>
<tr>
<td>$r_{xy}$</td>
<td>.58</td>
<td>.46</td>
</tr>
</tbody>
</table>

Unfortunately, data on test-retest reliability on the Loyola Language Study are limited. Stewart (1963) has presented the results of a long range (4 year) study. The Loyola Language Study was given to 96 college men and 104 college women when they were freshmen. When they were seniors 17 men and 23 women from this original group were given a retest. The scoring method used was Method II described above, i.e., using z-scores the 25 most discriminating items were scored. For the 17 men the correlation coefficient was .82 and for the women .62. For both together .68.
The other pertinent test-retest data on the Loyola Language Study is an unpublished report\(^2\) on a short-term, test-retest reliability study. In this study a group of 31 college sophomore men were given the Loyola Language Study on a test-retest basis with a three month interval. The scoring method employed was Method II, described above, i.e., using z-scores to score the 25 most discriminating items. The test-retest correlation was .67. In the same report data is given for 24 college sophomore and junior men who were given a retest on the Loyola Language Study after the same three month interval. The same scoring method, i.e., the 25 most discriminating items, was used. For this group of college men the test-retest correlation was .72.

Given this type of correlation between the test and itself, it would seem that the correlations, ranging from .58 -- .46 -- .69 between the Loyola Language Study and Design Y (5 choices), as well as those ranging from .49 -- .30 -- .65 for Design Z can be considered substantial, and perhaps the possibility of using either of these designs as alternate forms of the Loyola Language Study could be further investigated.

With regard to the time of administration it was found that the mean time taken by the 61 subjects to complete Design X was 12.8 minutes, for Design Y - 13.1 minutes and for Design Z - 15.9 minutes, --- as opposed to 30 minutes for the Loyola Language Study.

\(^2\)From the files of V. V. Herr on the Loyola Language Study, located at the Lake Shore Campus of Loyola University, Chicago.
However, it must be remembered that this figure of 30 minutes for the Loyola Language Study suffers because of the inaccuracy reported above (p. 31) with regard to the measurement of the time taken by each group on the Loyola Language Study. The time of 30 minutes for the Loyola Language Study, unlike the times for the Multiple Choice Designs, is an observed, not measured figure, although it does correspond with reports for similar groups. (Trainor, 1958)
CHAPTER V

CONCLUSIONS

It was empirically attempted to construct a multiple choice version of the Loyola Language Study, which would measure what the Loyola Language Study measures, i.e., presence of, or deviation from communality of response. It was hypothesized that a significant correlation between scores on the present version of the Loyola Language Study and a multiple choice version would indicate that the latter test would perform approximately the same task as the former -- at least for the limited segment of the population tested.

After a preliminary pilot study, three multiple choice designs were constructed -- Design X (four choices), Design Y (five choices) and Design Z (seven choices). Each of these multiple choice designs was administered to a matched group of male college students (each group, N:60) who had taken the Loyola Language Study, eleven weeks before. Correlations were then made between the scores of each group on the Loyola Language Study and the scores of the same groups on the multiple choice design that had been given to them. The results of these correlations are as follows:

All three designs have a significant correlation with the Loyola Language Study, using any of three methods of scoring. These correlations are significant at the .01 level of confidence, except for Design X and
Design Z, when scored for the 25 most discriminating items. In this instance the significance is at the .05 level of confidence.

There is no significant difference in the size of the coefficients of correlation between the scores on the Loyola Language Study and the scores on each of the multiple choice designs.

However, although its correlation is not significantly higher than for the other two designs, the correlation between Design Y and the Loyola Language Study is greater than that between Design X and the Loyola Language Study or between Design Z and the Loyola Language Study. This is true for all three methods of scoring.

For Scoring Method I (median scores) the correlation coefficient between Design Y and the Loyola Language Study is .58. This is .12 higher than the coefficient between Design X and the Loyola Language Study and .09 higher than the coefficient between Design Z and the Loyola Language Study.

For Scoring Method II (25 most critical items) the correlation coefficient between Design Y and the Loyola Language Study is .46. This is .19 higher than the coefficient between Design X and the Loyola Language Study and .16 higher than the coefficient between Design Z and the Loyola Language Study.

For Scoring Method III (independent Minnesota norms) the correlation coefficient between Design Y and the Loyola Language Study is .69. This is .09 higher than the coefficient between Design X and the Loyola Language Study and .03 higher than the coefficient between Design Z and the Loyola Language Study.
The above correlations are Pearson correlation coefficients. Spearman correlations were also computed. For each correlation they are one or two points different, but the above proportion is maintained.

In terms of correlations, then, it is seen that **Design Y** more closely approximates the Loyola Language Study than does **Designs X or Z**. However, the correlation between **Design Y** and the Loyola Language Study is not significantly greater than that between **Design X** and the Loyola Language Study and that between **Design Z** and the Loyola Language Study.

A comparison of mean scores obtained on the Loyola Language Study versus mean scores obtained on each of the multiple choice designs was also made. The results are as follows:

In the case of **Design X**, by all three methods of scoring the null hypothesis is proven incorrect, i.e., the mean scores of the group which was given the Loyola Language Study were significantly different from the mean scores obtained by the same group on **Design X**, no matter which scoring method was used.

There was no significant difference in mean scores between the Loyola Language Study and **Design Y**, when scoring **Method I** (Median scores) and scoring **Method III** (independent Minnesota norms) were used. There was a significant difference in mean scores when scoring **Method II** (25 most discriminating items) was used.

When Scoring **Method I** (Median norms) and Scoring **Method III** (independent Minnesota norms) were employed, there was a significant difference in means between the Loyola Language Study and **Design Z**. When Scoring
Method II (25 most discriminating items) was utilized there was no significant difference in mean scores between Loyola Language Study and Design Z.

So in terms of relationship between means, there is a significant relation between Design Y and the Loyola Language Study, when Scoring Method I (Median norms) and Method III (independent Minnesota norms) are used; on the other hand, when Scoring Method III (25 most discriminating items) is used there is a significant relation between Design Z and the Loyola Language Study. All other relations between means are not significant.

In light of the above data, i.e., in terms of correlation coefficients and relationships of means, it seems that Design Y (5 choices) best (but not significantly "best") approximates the Loyola Language Study, when Scoring Method I (median norms) and Scoring Method III (independent Minnesota norms) are employed.

When Scoring Method II (the 25 most discriminating items) is used, the decision is not so clear. In terms of mean scores there is a significant difference between Design Y and the Loyola Language Study; there is no significant difference between Design Z and the Loyola Language Study. In terms of correlation coefficients, the Pearson correlation coefficient of .46 between Design Y and the Loyola Language Study is .16 higher than that between Design Z and the Loyola Language Study. Should this higher correlation offset the lack of goodness of fit with regard to means? One is inclined to respond in the affirmative since the difference in means between Design Y and the Loyola Language Study is due to the fact that the
mean score on Design Y is 30 points better than the mean score on the Loyola Language Study. This may allow for better discrimination between normals and abnormals; if not, i.e., if the abnormals also would score 30 points better, nothing is lost.

However, it is in terms of Scoring Method I (Median Scores) that Design Y should be judged. It is these norms that give a multiple choice design one of its primary advantages, i.e., ease and simplicity of scoring, although this method is not perfectly applicable to the case in which extremely poor scores are obtained, as for example with schizophrenics. And using this method of scoring, on all counts, Design Y, although not significantly different in size of correlation from the other two designs, most closely approximates the Loyola Language Study — with a correlation coefficient of .58 and with no significant difference in mean scores.

It would be worthwhile to investigate if there is a correlation between Design Y and the Loyola Language Study, when scored by this same method and given to abnormals, and if it will yield similar differences between patients and normals without producing scores which are too extreme to be used with schizophrenics.

On the other hand, when Design Y is scored according to that method which has been found to discriminate normal and abnormal subjects best (standard z-scores for the 25 most discriminating items) the mean score differs significantly from the mean score obtained on the Loyola Language Study and the correlation of .45 is fairly poor. Design Z's mean score is not significantly different, but the correlation of .30 is even poorer. Neither Design Y nor Design Z -- when scored by this method -- seem very
promising.
BIBLIOGRAPHY


Galton, F. Psychometric experiments. Brain, 1879, 2, 149-162.


APPENDIX I

LOYOLA LANGUAGE STUDY
REVISED
LOYOLA LANGUAGE STUDY

Instructions

When people see or hear a word, they often think of another word. If you say the word *stem*, most people would think of *flower*. Some, but not the greatest number, might think of *pipe*, *grass*, *stop*, and so forth.

This study wants to find out what word you think the *greatest number of people* would be most likely to think of when they see or hear each of the words on the next two pages.

Please write next to each of the words the *one word* which you think the *greatest number of people* would be most likely to think of when they see or hear the word in the list. Take as much time as you need to think about the word which seems to you to “go along” with each printed word. Then choose the *one word* which you think the *greatest number of people* would be most likely to think of when they see or hear the given word. Write the *one word* which you choose beside the printed word. Do not skip any word.

Remember, you are not asked to write down just any word that comes to your mind. You should write down the *one word* which you think the *greatest number of people* would be most likely to think of.

**Important:** please fill out the information blank on page 4.

Copyright 1954, by LOYOLA UNIVERSITY, CHICAGO
Beside each of the words printed below write the one word which you think the greatest number of people would be most likely to think of when they see or hear that word.

soldier  soldier
hungry  hungry
butterfly  butterfly
long  long
head  head
anger  anger
afraid  afraid
fruit  fruit
dark  dark
red  red
loud  loud
bath  bath
eating  eating
joy  joy
rough  rough
heavy  heavy
high  high
white  white
command  command

whiskey  whiskey
yellow  yellow
window  window
scissors  scissors
foot  foot
doctor  doctor
wish  wish
house  house
justice  justice
river  river
sickness  sickness
mountain  mountain
stove  stove
girl  girl
salt  salt
man  man
cheese  cheese
baby  baby
moon  moon
spider  spider
bread  bread

whistle  whistle
carpet  carpet
needle  needle
hand  hand
thief  thief
dream  dream
trouble  trouble
religion  religion
street  street
health  health
ocean  ocean
bed  bed
child  child
tobacco  tobacco
woman  woman
cabbage  cabbage
citizen  citizen
earth  earth
lion  lion
butter  butter
music  music
The following information is essential for research purposes. Without it, no good can come from the trouble you have taken to fill out the two previous pages.

RESIDENCE (city and state)..................................................................................................................

BIRTHPLACE (city and state)..................................................................................................................

MONTH AND YEAR OF BIRTH..............................................................................................................

SEX (male or female).................................................................................................................................

Highest year of school completed (circle one):

<table>
<thead>
<tr>
<th>HIGH SCHOOL</th>
<th>COLLEGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>10</td>
<td>18</td>
</tr>
</tbody>
</table>

From what countries did your parents' people come?

Father's people........................................................................................................................................

Mother's people........................................................................................................................................

YOUR OCCUPATION....................................................................................................................................

If you are a student or housewife, what is your father's or husband's occupation?

If you wish, give your name and address

NAME..........................................................................................................................................................

STREET....................................................................................................................................................

CITY...........................................................................................................................................................

Return to:
LOYOLA LANGUAGE STUDY
820 North Michigan Avenue
Chicago 11, Illinois
APPENDIX II

LOYOLA LANGUAGE STUDY

(Multiple Choice - Design X)
LOYOLA LANGUAGE STUDY
(Multiple Choice- Design X)

Instructions

When people see or hear a word, they often think of another word. If you say the word "STEM" most people would think of "flower." Some, but not the greatest number, might think of "pipe," "grass," "stop," and so forth.

This study wants to find out what word you think the greatest number of people would be most likely to think of when they see or hear each of the Capitalized words on the next pages.

For each Capitalized word you will be given four choices. Take the word "STEM" as an illustration:

STEM: flower grass pipe stop

You are asked to circle the one word which you think the greatest number of people would be most likely to think of when they see or hear the word in the list.

Therefore, for example, if you think that, when they see or hear the word "STEM", the greatest number of people would think of "flower," you would circle the word "flower." Like this:

STEM: (flower) grass pipe stop

Take as much time as you need to think about the word which seems to you to "go along" with each Capitalized word. Then choose the one word which you think the greatest number of people would be most likely to think of when they see or hear the given word. Circle the one word which you choose. Do not skip any word.

Remember, you are not asked to circle just any of the four possible choices that come to your mind. You should circle the one word which you think the greatest number of people would be most likely to think of.

Important: At the end of the study you will be asked to put down how long it took you to complete the study. So, as you begin, notice what time it is, so that when you finish you will know how long it took.

But note, the time that it takes is not at all important. Take all the time you want. What is important is that you circle the one word which you think the greatest number of people would be most likely to think of when they see or hear the Capitalized word.
SOLDIER: man troops weary army

HUNGRY: craving food sandwich eat

BUTTERFLY: meadow insect species bug

LONG: short speech big hours

HEAD: brain heart lettuce hair

ANGER: fight mad bitter control

AFRAID: dark scared cringe dread

FRUIT: eat cocktail ripe apple

DARK: fear tunnel night bright

RED: color gaudy blush rose

LOUD: silent noise soft cheap

BATH: refreshed clean bubble wash

EATING: waistline calories food hunger

JOY: laughter mirth grief happiness

ROUGH: smooth rugged chapped callous

HEAVY: weight cumbersome sorrow lead

HIGH: score low towering mountain

WHITE: lies light shiny black

COMMAND: boss order make stern
SOUR: puss dill pickle sweet
KING: majestic royalty reign queen
DEEP: sea water ditch crevice
SLEEP: tired rest snore pillow
BLACK: color midnight white coffee
HAMMER: anvil pound nails workman
TABLE: eat chair round hungry
THIRSTY: parched dry beverage water
QUIET: meek still shy noisy
HARD: difficult granite soft labor
BLUE: Monday ribbon color sky
SWEET: candy innocent heart sour
STOMACH: hungry food paunch weight
WORKING: conditions support earning labor
COMFORT: solace ease pillow easy
SOFT: rough flabby feather hard
SHORT: brief small sleeves snappy
BEAUTIFUL: pretty description weather scenery
COLD: numb weather metal winter
WHISKEY: fifth smell drink drunkard

YELLOW: color canary streak yoke

WINDOW: glass house ledge drape

SCISSORS: cutting pair clippers cut

FOOT: sore shoe toe inches

DOCTOR: examine scalpel sick nurse

WISH: happen money hope unattainable

HOUSE: privacy building home warmth

JUSTICE: law democracy virtue judge

RIVER: water current muddy stream

SICKNESS: hospital doctor distress operation

MOUNTAIN: climb high rugged grandeur

STOVE: baking heat hot burner

GIRL: scout dress boy figure

SALT: pepper taste necessary peanuts

MAN: labor young boy woman

CHEESE: milk apple cheddar food

BABY: child diaper born offspring

MOON: lovers orbit night romance

SPIDER: creature web dread bug

PREACH: to hilde, post, dither
WHISKEY: fifth smell drink drunkard

YELLOW: color canary streak yoke

WINDOW: glass house ledge drape

SCISSORS: cutting pair clippers cut

FOOT: sore shoe toe inches

DOCTOR: examine scalpel sick nurse

WISH: happen money hope unattainable

HOUSE: privacy building home warmth

JUSTICE: law democracy virtue judge

RIVER: water current muddy stream

SICKNESS: hospital doctor distress operation

MOUNTAIN: climb high rugged grandeur

STOVE: baking heat hot burner

GIRL: scout dress boy figure

SALT: pepper taste necessary peanuts

MAN: labor young boy woman

CHEESE: milk apple cheddar food

BABY: child diaper born offspring

MOON: lovers orbit night romance

SPIDER: creature web dread bug

BREAD: eat daily yeast butter
WHISTLE: train flirt melody sound

CARPET: thick plush soft rug

NEEDLE: darn thread sewing injection

HAND: finger clasp work grip

THIEF: stealing valuables night steal

DREAM: nightmare indigestion sleep unconscious

TROUBLE: heart sickness adversity worry

RELIGION: truth Christianity church faith

STREET: residence road walk paving

HEALTH: happiness feeling appetite energy

OCEAN: water ship shore swim

BED: slumber enjoyment sleep rest

CHILD: growing blessing baby school

TOBACCO: smoking farmer field smoke

WOMAN: wife marriage femininity man

CABBAGE: vegetable kings red head

CITIZEN: nation American average man

EARTH: ground dwell planet life

LION: man-eater run roar animal

BUTTER: jelly bread table food

MUSIC: tempo notes song scale

(TIME TAKEN: __________ minutes.)
APPENDIX III

LOYOLA LANGUAGE STUDY

(Multiple Choice - Design Y)
Instructions:

When people see or hear a word, they often think of another word. If you say the word "STEM" most people would think of "flower." Some, but not the greatest number, might think of "pipe," "grass," "stop," and so forth.

This study wants to find out what word you think the greatest number of people would be most likely to think of when they see or hear each of the Capitalized words on the next pages.

For each Capitalized word you will be given five choices. Take the word "STEM" as an illustration:

STEM: grass flower pipe stop twig

You are asked to circle the one word which you think the greatest number of people would be most likely to think of when they see or hear the word in the list.

Therefore, for example, if you think that, when they see or hear the word "STEM," the greatest number of people would think of "flower," you would circle the word "flower." Like this:

STEM: grass (flower) pipe stop twig

Take as much time as you need to think about the word which seems to you to "go along" with each Capitalized word. Then choose the one word which you think the greatest number of people would be most likely to think of when they see or hear the given word. Circle the one word which you choose. Do not skip any word.

Remember, you are not asked to circle just any of the five possible choices that come to your mind. You should circle the one word which you think the greatest number of people would be most likely to think of.

Important. At the end of the study you will be asked to put down how long it took you to complete the study. So, as you begin, notice what time it is, so that when you finish you will know how long it took.

But note, the time that it takes is not at all important. Take all the time you want. What is important is that you circle the one word which you think the greatest number of people would be most likely to think of when they see or hear the Capitalized word.
SOLDIER: man troops weary army uniform

HUNGRY: sandwich starved eat food craving

BUTTERFLY: meadow bug insect species beauty

LONG: hours short big road speech

HEAD: hat brain hair lettuce heart

ANGER: mad bitter rage control fight

AFRAID: dark dread cringe danger scared

FRUIT: eat cocktail ripe tree apple

DARK: fear bright danger tunnel night

RED: blue rose color gaudy blush

LOUD: silent cheap quiet noise soft

BATH: refreshed wash cleanliness bubble clean

EATING: calories enjoyment food waistline happiness

JOY: happiness mirth grief laughter fun

ROUGH: rugged mean callous smooth chapped

HEAVY: lead sorrow big cumbersome weight

HIGH: score towering mountain building low

WHITE: black light shiny lies sheet

COMMAND: army make order boss stern
SOUR: dill sweet pickle spoiled puss
KING: royalty England majestic queen reign
DEEP: depth crevice water ditch sea
SLEEP: tired night snore pillow rest
BLACK: coffee white midnight color coal
HAMMER: nails pound carpenter workman anvil
TABLE: round eat furniture chair hungry
THIRSTY: beverage parched drink water dry
QUIET: sleep still meek shy noisy
HARD: work soft difficult granite labor
BLUE: color sad sky ribbon Monday
SWEET: innocent taste candy sour heart
STOMACH: paunch food hunger weight hungry
WORKING: laboring conditions earning support labor
COMFORT: easy solace ease relaxed pillow
SOFT: flabby rough fur feather hard
SHORT: small midget sleeves brief snappy
BEAUTIFUL: flower scenery pretty weather description
COLD: numb weather metal winter snow
WHISKEY: drunkard tavern fifth drink smell

YELLOW: butter canary color streak yoke

WINDOW: ledge glass drape house view

SCISSORS: cut pair clippers sewing cutting

FOOT: leg inches shoe sore toe

DOCTOR: nurse sick examine scalpel physician

WISH: hope thought money unattainable happen

HOUSE: warmth privacy home building comfort

JUSTICE: judge virtue democracy law fairness

RIVER: current water stream boat muddy

SICKNESS: disease hospital operation distress doctor

MOUNTAIN: high climb snow grandeur rugged

STOVE: fire hot burner heat baking

GIRL: dress child s-cout figure boy

SALT: taste necessary spice pepper peanuts

MAN: father young labor woman boy

CHEESE: apple rat food cheddar holes

BABY: love born offspring child diaper

MOON: orbit night love romance lovers

SPIDER: dread creature web bug fly

BREAD: flour daily eat yeast butter
WHISTLE: melody sound blow train flirt

CARPET: plush thick rug soft home

NEEDLE: pin darn sewing thread injection

HAND: work finger clasp shake grip

THIEF: rob night steal valuables stealing

DREAM: sleep indigestion nightmare unconscious sleeping

TROUBLE: worry heart sadness adversity sickness

RELIGION: Christianity church truth faith belief

STREET: residence home paving walk road

HEALTH: appetite energy sick feeling happiness

OCEAN: swim water boat ship shore

BED: sheet rest enjoyment slumber sleep

CHILD: infant school baby blessing growing

TOBACCO: field pipe smoking farmer smoke

WOMAN: man girl wife marriage femininity

CABBAGE: vegetable head red slaw kings

CITIZEN: alien man American nation average

EARTH: planet ground life dwell land

LION: roar run beast man-eater animal

BUTTER: bread jelly table yellow food

MUSIC: tempo rhythm scale song notes

(Time taken ________ minutes.)
APPENDIX IV

LOYOLA LANGUAGE STUDY

(Multiple Choice - Design Z)
LOYOLA LANGUAGE STUDY
(Multiple Choice- Design Z)

Instructions

When people see or hear a word, they often think of another word. If you say the word "STEM" most people would think of "flower." Some, but not the greatest number, might think of "pipe," "grass," "stop," and so forth.

This study wants to find out what word you think the greatest number of people would be most likely to think of when they see or hear each of the capitalized words on the next pages.

For each Capitalized word you will be given seven choices. Take the word "STEM" as an illustration:

STEM: grass flower pipe stop twig petal bud

You are asked to circle the one word which you think the greatest number of people would be most likely to think of when they see or hear the word in the list.

Therefore, for example, if you think that, when they see or hear the word "STEM," the greatest number of people would think of "flower," you would circle the word "flower." Like this:

STEM: grass flower pipe stop twig petal bud

Take as much time as you need to think about the word which seems to you to "go along" with each Capitalized word. Then choose the one word which you think the greatest number of people would be most likely to think of when they see or hear the given word. Circle the one word which you choose. Do not skip any word.

Remember, you are not asked to circle just any of the seven possible choices that come to your mind. You should circle the one word which you think the greatest number of people would be most likely to think of.

IMPORTANT: At the end of the study you will be asked to put down how long it took you to complete the study. So, as you begin, notice what time it is, so that when you finish you will know how long it took.

But note, the time that it takes is not at all important. Take all the time you want. What is important is that you circle the one word which you think the greatest number of people would be most likely to think of when they see or hear the Capitalized word.
SOLDIER: army uniform general man troops weary gun

HUNGRY: craving eat poor food sandwich starved pain

BUTTERFLY: meadow species beauty insect bug bird free

LONG: speech big road tall hours short boring

HEAD: hat brain person ability lettuce hair heart

ANGER: fight ire mad bitter rage curse control

AFRAID: dark coward helpless cringe scared dread danger

FRUIT: eat tree vegetable cocktail apple seeds ripe

DARK: room colored fear bright tunnel color night

RED: blush color paint gaudy white rose blue

LOUD: silent quiet soft noise boisterous commotion cheap

BATH: refreshed clean wash cleanliness soap relax bubble

EATING: necessity hunger calories waistline meal food enjoyment

JOY: happiness elation mirth sorrow laughter fun grief

ROUGH: mean ready chapped smooth unmannered callous rugged

HEAVY: lift bricks sorrow weight cumbersome big lead

HIGH: altitude low score building kite towering mountain

WHITE: bleach lies clear light sheet shiny black

COMMAND: army order boss authority forward stern make
SOUR: cream sweet pickle puss spoiled dill cherries

KING: majestic crown queen royalty England subjects reign

DEEP: water sea depth ditch bottom crevice low

SLEEP: tired snore night pillow lazy awake rest

BLACK: dirty white color coal pitch midnight coffee

HAMMER: nails pound workman carpenter anvil chisel hit

TABLE: round chair plates eat furniture eating hungry

THIRSTY: drink beverage dry hot water parched liquid

QUIET: noisy still nightfall sleep restful meek shy

HARD: severe work rough granite labor difficult soft

BLUE: ribbon color mood sky Monday red sad

SWEET: candy demure taste good heart innocent sour

STOMACH: food hunger mouth eating hungry weight paunch

WORKING: labor task support conditions playing earning laboring

COMFORT: solace pillow console easy relaxed ease convenience

SOFT: skin flabby rough cushion feather fur hard

SHORT: midget brief sudden stubby sleeves snappy small

BEAUTIFUL: scenery flower pretty weather attractive sky description

COLD: weather hard warm metal numb winter snow
WHISKEY: smell drunkard drank warmth tavern fifth sour

YELLOW: stained gold color butter streak yoke canary

WINDOW: glass washing view ledge see drape house

SCISSORS: cut sewing pair clippers cutting haircut sharp

FOOT: hand shoe leg inches sore tall ball

DOCTOR: nurse sick life physician examine hospital scalpel

WISH: request thought unattainable bone happen hope money

HOUSE: rambling warmth building comfort people privacy home

JUSTICE: judge fairness virtue law criminal fair democracy

RIVER: boat current fishing muddy water stream rushing

SICKNESS: distress hospital disease medicine operation doctor weakness

MOUNTAIN: rugged peak grandeur high climb majestic snow

STOVE: hot heat warmth fire burner baking cozy

GIRL: figure scout charm lady child boy dress

SALT: pepper season spice necessary element peanuts taste

MAN: young labor masculinity person father woman boy

CHEESE: sandwich rat apple cheddar holes milk food

BABY: crib helpless offspring love child born diaper

MOON: lovers love night rocket evening orbit romance

SPIDER: chills creature fly web dread bug poison

BREAD: yeast flour butter daily eat sandwich homemade
WHISTLE: sound shrill blow melody rude train flirt

CARPET: beater thick home rug covering soft plush

NEEDLE: stick sewing point injection darn pin thread

HAND: shake finger strength grip help clasp work

THIEF: prowler valuables police night stealing rob steal

DREAM: sleeping indigestion sleep nightmare unconscious thoughts envision

TROUBLE: adversity worry sickness heart misery tribulation sadness

RELIGION: church sacred truth Catholic Christianity faith belief

STREET: road walk residence corner crowd paving home

HEALTH: appetite sick happiness energy feeling Department robust

OCEAN: boat water ship horizon waves shore swim

BED: rest lazy soft enjoyment slumber sheet sleep

CHILD: infant baby growing school psychology boy blessing

TOBACCO: smoke pouch cigar smoking field farmer pipe

WOMAN: marriage wife man softness child femininity girl

CABBAGE: indigestion slaw kings garden head vegetable red

EARTH: planet land life agriculture sky ground dwell

CITIZEN: American man alien nation freedom U.S.A average

LION: beast man-eater fierce roar animal run kill

BUTTER: yellow jelly taste milk table bread food

MUSIC: scale listen notes sing tempo talent rhythm
APPROVAL SHEET

The thesis submitted by Richard C. Braun, S.J. has been read and approved by three members of the Department of Psychology.

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated, and that the thesis is now given final approval with reference to content, form, and mechanical accuracy.

The thesis is therefore accepted in partial fulfillment of the requirements for the Degree of Master of Arts.

Date

Signature of Adviser