A COMPARISON OF WORD ASSOCIATION MEASURES AS PREDICTORS OF RECALL, AND AN ASSESSMENT OF THE RELIABILITY OF FREE ASSOCIATIONS

by

Brian Douglas Mackenzie

B. A., University of British Columbia, 1966

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS
in the Department
of
Psychology

L. C. [illegible]

SIMON FRASER UNIVERSITY

August, 1969

APPROVAL

Name: Brian Douglas Mackenzie

Degree: Master of Arts

Title of Thesis: A comparison of word association measures as predictors of recall, and an assessment of the reliability of free associations

Examining Committee:

(L.M. Kendall, Ph.D.)

(K.F. Koopman, Ph.D.)

(A.R. Blackman, Ph.D.)

(S.S. Lee, Ph.D.)
External Examiner
Assistant Professor
Department of Educational Psychology
Faculty of Education
University of British Columbia
Vancouver, B. C.

Date Approved: **October 20, 1969**

Abstract

The purpose of the present research was a) to make a comparative assessment of the power of several measures of word association to predict free recall, and b) to investigate the reliability and related properties of word association behaviour.

The subjects, first and second year university students, gave single written responses on two administrations of a word association test; they were then tested on free recall of one of a set of lists of words drawn from the word association stimuli. The response distributions on the word association test, pooled over the sample, were analyzed and compared by means of several association measures. Functions of the measures were then used to predict the average recall of each word list and the proportional frequency of recall of each word in the lists.

It was found impossible to differentiate meaningfully the predictors of average list recall, as the predictors were all highly correlated with each other and with the criterion. For prediction of recall of individual words, the best single predictor was a function of a correlation measure (a conditional probabilities correlation coefficient) not previously used to measure free associations.

The same correlation measure was used to determine the test-retest and scale, or split-sample, reliabilities of the response distributions. Both forms of reliability coefficients were found to be quite high, with means of over 0.90. Comparisons were made with previously reported reliability measures.

The assumption of the absence of individual differences on the emission of associations was tested by means of predictions derived from some of the reliability measures. It was concluded that most of the variance in responding on a single word association test is due to between-subject influences. On retesting, many subjects exhibit a consistency of response patterns which is independent of the between-subject influences.

Table of Contents

Page

Volume I

Volume II: Appendices

List of Tables

Chapter 1: Introduction and Review

A research paradigm which is often used in empirical investigations of cognitive processes is that of free association. In the most common experimental procedure, a subject gives written associations to visually presented words or nonsense syllables; oral presentation and response is a common variant. Instructions usually emphasize the need for speed and spontaneity of responding, so as to minimize the effects of higher level cognitive mediation. Experimentation in this area has a long and extensive history, dating at least from the work of Kent and Rosanoff (1910). It continues, however, to be plagued by the failure to develop valid and reliable measures of word association. The purpose of the present research is a) to make a comparative assessment of the power of several measures of word association to predict free recall, and b) to investigate the reliability and related properties of word association behaviour.

There are numerous measures of free association currently in use, and new ones are frequently proposed. A review article by Marshall and Cofer (1963), for example, describes eight separate measures for use with free associations; a study by Bousfield and Puff (1965a) lists eleven more. A few of the measures are to some degree related to theories of cognitive functioning (eg., Bousfield, 1953; Deese, 1965). Most, however, are simply attempts to provide empirical generalizations which concisely describe the observed response distributions in terms of the degree, strength, and pattern of associations. Few comparative studies have been performed, and it is difficult to compare results of studies using different measures.

Several different response measures have also been applied to the measuring of free recall. Common measures include the number of words recalled from a list, the number recalled from one or more sublists, the number of conceptual categories represented in recall, the frequency of recall of individual words from a list, and the extent to which words cluster together in recall. Recall measures have been more consistently related to specific problems and theories than have association measures; nevertheless, the number of measures in use contributes to the difficulty of comparing and interpreting different studies.

It is noteworthy that none of the currently popular association measures are explicitly derived from, or consistent with, statistical and measurement theory. Word association distributions are certainly amenable to rigorous statistical analysis, and such analysis does no more violence to the inner nature of the responses than does any other numerical manipulation. It seems surprising that measures of bivariate correlation (P.M. Jenkins & Cofer, 1957), variance (Brotsky & Linton, 1967; Horvath, 1963), and test-retest reliability (Gegoski & Riegel, 1967; Hall, 1966) have been formulated without reference to available and appropriate statistical and measurement models. As a result, the mathematical derivation of many measures is rather sketchy, and may sometimes be faulty. Elaborate statistical analysis performed upon the results of such measures (eg., Deese, 1962, 1964; Howe, 1966) may be questioned if based upon incorrect assumptions about the derived data. For purposes of factor analysis, for example, it is necessary to seek measures which behave in accordance with the requirements of the statistical model used in analysis.

The Theory of Associations

The fact that there are links, or associations, joining ideas to other ideas and words to other words is fundamental to the study of mental processes; observations and descriptions of the associative process have appeared in written form at least since the time of Plato (particularly in the *Phaedo*). Relatively "modern" theories of the cause and basis of associations were proposed by Hobbes, Locke, and Hume. Their basic theories, although elaborated and rephrased in contemporary language, provide the basis for much of the current theoretical and experimental work on association.

The "laws" of association developed by the early British empiricists are familiar to most students of verbal learning and philosophy. *Contiguity* is the fundamental law proposed by Hobbes, and states that if two ideas have been made to occur in temporal contiguity in a person's experience, he will come to make the association between them automatically.

The law of *frequency* was added by Locke, and states that the more often two ideas have been experientally joined, the more readily will they occur together in the mind. The law of *resemblance* or *similarity* was derived from the law of *contiguity* by Hume; the more similar two ideas are, the more likely they are to occur in temporal contiguity; hence, the more likely they are to be joined in the mind.

"Reinforcement" can be substituted for "contiguity" (with the understanding that reinforcement occurs through contiguity), and "resemblance" can be relegated to the status of a higher-order mediating process (Hume, in fact, intended it in a similar way). These changes in terminology and emphasis render the "laws" distressingly contemporary. As Deese (1968) has observed:

> ...the fundamental assumptions of the study of associative processes have been untouched by nearly an entire century of empirical investigation. It is hard to think of another discipline in which so long a period of investigation has not been accompanied by a fundamental change of assumptions. It is by no means obvious to many students of verbal behaviour that radically different assumptions about the nature of associations are even possible (p. 97).

This lack of theoretical progress may be due in part to the absence of an adequate methodology for observing, measuring, and analyzing associations. It is more explicitly due to the fact that the tenets of British empiricism have passed from the status of theory to that of underlying, often unrecognized assumptions. Several attempts have been made to formulate theories of association and verbal mediation (eg., Bousfield, 1953; Cofer & Foley, 1942), but they are mostly restricted to deriving the mechanisms of specific association paradigms from the general "laws" of association. The influence of these "laws" on theoretical accounts of verbal phenomena in the first half of the century is reviewed in detail by Goss (1961).

There are, of course, alternative views of association to the classical ones, and the alternatives are gradually coming to be presented with greater sophistication and generality. Two of the leaders in formulating distinctive association theories are Deese (esp. 1962, 1965, 1968) and Pollio (1966, 1968). Although their approaches differ, both Deese and Pollio focus on the connected structure formed by the associations given to a set of related words. Patterns of association, they agree, are not determined by a series of single-association reinforcements, but are based on superordinate conceptual organizations which are themselves determined by the individual's broad linguistic experience in a variety of language situations.

The development of association theories that are not dependent on the tenets of philosophical empiricism is, curiously enough, just beginning. Such a theory will necessitate far broader inferences and extrapolations from the observed association data than are required by a simple, chained response theory. In particular, the study of the structure of numerous associatively related words would seem to require considerable use of multivariate techniques of analysis. It is essential, therefore, that the measures applied to the basic data be as rigorous and logically defensible as possible. It is hoped that the analysis of association measures presented in this report will provide some of the logical and empirical underpinnings necessary.

Problems and Assumptions in the Measurement of Free Associations

Equivalence of Measures

In discussing the comparability of different measures, we are trying to assess their equivalence according to some criterion. Several levels of equivalence have been described, and are discussed at length by Gulliksen (1950, 1968). Gulliksen (1968) details three levels of equivalence that are relevant here. In the first and highest level, the equivalent measures are interchangeable. They have the same true scores, means, variances, and reliabilities. In the second level, the measures have the same true scores, but have different error variances, and hence different reliabilities. The third level is similar to the second, but is expressed in factor analytic terminology. At the third level, the same common factor is tapped by both measures, but they have different specific factors. The third level differs substantively from the first and second in that the common factor need not be tapped to the same extent by both measures. In many situations, the third level of equivalence may be the most practical and possible to attain.

Smith, Kendall, and Hulin (1969) offer a more pragmatic definition of equivalence:

> By equivalence of measures we mean the assurance that conclusions reached using one measure would be the same had another measure been used for the same purpose....The focus is on equivalence of relations with other variables. This requirement can be less demanding, however, than requiring complete comparability of factor composition. We require comparability only in terms of relationships with selected other variables or factors about which we are to draw conclusions (pp. 7-8).

This emphasis on "functional" equivalence suggests a distinction between internal and external equivalence. If two operationally different bivariate correlation measures are applied to the same body of data, and the resulting correlation matrices are both factored, the similarity of the two factor structures measures the internal equivalence of the measures. The focus is on similarity of description of the primary data. External equivalence refers to the relation of the measures to an external criterion. If two association measures predict recall equally well, they have a form of external equivalence. The focus is on using some characteristic of the measure as a tool to define, describe, or predict another, independently measured, phenomenon. Alternately, the focus may be on reaching a decision about whether the measure and the external phenomenon are in fact related at all. Does the first, for instance, discernibly influence the second? Can the second be predicted from the first? The external equivalence of the predictive measures may, at this level, be dependent simply upon their both yielding the same yes or no answer to these questions.

Internal and external equivalence, as described here, may operationally be somewhat different properties. Complete internal equivalence (at the first level described by Gulliksen, 1968) of course implies complete external equivalence, and the level of internal and external equivalence may be the same in any case. However, the degree or extent of equivalence at the second and third levels is not necessarily the same in the internal and external applications.

Level of Analysis

Word associations, like most other measures, are used as a tool to study and learn about something else. The "something else" may be cognitive processes, the relation of thought to memory, psychopathology, or the characteristics of language-in-use. Each purpose requires that the data be treated in a different way. Studying the characteristics of language, for instance, often requires factoring the matrix of intercorrelations derived from the words. Studying psychopathology, on the other hand, may require the detailed examination of single responses by individual subjects. In using free association measures to predict free recall, there is a choice of levels at which the data can be analyzed. If the criterion of interest is number of words recalled from a list, then some summary measure of the average association value of the list is appropriate as a predictor. If recall of single words from a list is being considered, the appropriate measure is one that applies to each list item individually. In many cases, the two levels of analysis can use different forms of the same measure--the average correlation and the squared multiple correlation can be used to predict list recall and word recall, respectively. It is necessary to remember that the same measure may have quite different properties at different levels of analysis. Conclusions about the worth of a measure as a predictor must be restricted to the kind of predictive situation in which it is performed. Caution is particularly advisable when mean scores are related; in some circumstances, means of predictors and criteria may be correlated much more highly than are individual scores. The relationship between group and individual observations is discussed well by Meltzer (1963) and Robinson (1950).

The Assumption of the Absence of Individual Differences in the Emission of Associations

Measures of word association are almost always based on the assumption that people do not differ significantly with respect to their internal hierarchies of association. It is assumed that a distribution of associations can be gathered from a large sample and be considered typical of every member of that sample. The probability that any person will emit a particular association to a given stimulus is considered equal to the proportion of subjects in the entire sample who gave that association.

The assumption is necessary if the association response distributions are to be considered meaningfully related to the internal associative mechanisms of the subjects; it is practically impossible to obtain a distribution of associations from individual subjects. However, the assumption is not so stringent as it may at first appear.

To begin with, it is recognized that people do differ with respect to their associations; the clinical use of free associations is based on analysis of some of the differences. In experimental work, however, the differences between subjects are assumed to be negligible compared to the similarities. The instructions in word association experiments moreover, are designed to minimize individual differences by reducing the effect of personal response styles. Idiosyncratic responses do of course occur, but are held to have little effect on the total response distribution.

Furthermore, the assumption cannot be considered to have completely general applicability. It may have adequate validity for homogeneous groups of subjects, such as the college freshmen usually found in association experiments, but not be valid for the population at large. Rosenzweig (1964) in fact found major differences in associations between separate cultural groups. These qualifications on the applicability of the assumption reduce, in part, the rigor of the conditions necessary for it to be satisfied, and make it slightly more tenable for application to specific experimental situations.

The assumption has not been substantiated definitively, as the procedure for obtaining an associative hierarchy from an individual would force many additional constraints upon his associative behaviour. The subject's memory of his previous responses would prevent the large number of trials necessary to obtain the distribution to any one stimulus from being independent. The subject might exhibit a false consistency of responding, or he might specifically try to make every response different. In neither case could it be assumed that the observed response distribution was typical of his hypothetical "inner" associative hierarchy.

It is possible to provide some support for the assumption by an approximation to the single-subject case. If the constraints specific to repeated associations are not appreciable until a fairly large number of responses has been given, then a few repeated associations to a stimulus given by each of a small number of subjects can be used as an estimate of the "true" distribution of associations to that stimulus. The extent to which such a distribution differs from a distribution collected from single responses given by a larger number of subjects estimates the extent to which individual differences are, in fact, implicated in the emission of associations. Such an experiment was performed by Garskof (1965), in which he found that in most cases the distributions collected from the small sample giving repeated associations closely approximated those from the larger sample giving single associations.

It has been stated by Deese (1965) that the assumption is partly self-correcting. Predictions requiring the assumption, he states, can be fulfilled only to the extent that the assumption is true. On the other hand, it may also be the case that some predictions which are believed to require the assumption do not, in fact, require it.

Further testing of the validity and applicability of the assumption is possible through examination of data measuring the reliability of word associations, and will be discussed in a later section.

The Representational Response

All bivariate indices of association are computed as some function of the number of common responses given to both of a pair of stimuli. If the two stimuli regularly elicit each other, but do not both elicit the same responses, they will appear to have no relation. Such is the case with the words "hard" and "soft" in the Minnesota norms (Palermo & Jenkins, 1964). "Soft" is the most common response to "hard" and "hard" is the most common response to "soft", but there are almost no responses common to both stimuli. There seems to be a close relation between the two words, but it cannot be measured unless the association measures take the reciprocal elicitation into account.

The solution offered by Bousfield, Whitmarsh, and Berkowitz (1960) requires the assumption that each stimulus elicits both itself and another response. The second response is typically different from the stimulus, and is the observed response in a word association test. The duplicate of the stimulus, which is assumed to be elicited, is called the representational response; the observed response is actually made to the representational response rather than to the stimulus as presented. Each subject, therefore, makes $2k$ responses to the k words on the word association test. Since each stimulus always elicits itself, it is possible to measure the response similarity between two stimuli that also elicit each other. The representational response assumption has been widely accepted, and has been used by most investigators using bivariate indices to measure word associations (eg., Cofer, 1957; Cowan, 1966; Deese, 1962, 1964, 1965).

The representational response effectively allows for the expression of the relation existing when two words elicit each other. However, it also adds serious constraints to the determination of the overall relation between two words. Two stimuli will have an association value of 1.0 (in all the bivariate measures to be considered) only if each stimulus always elicits the other one, and neither stimulus ever elicits any other response. Two stimuli which have identical response distributions, but which do not elicit each other, will have an association value of only 0.5. Without the representational response, the corresponding association values are 0.0 and 1.0, respectively.

The problem lies in trying to describe two separate forms of associative relation with the same measure. Stimulus equivalence (in generating response distributions) is one form of relation, and frequency of reciprocal elicitation is another; these two forms are quite separate, but are not independent. A theoretical proposal analogous to that of the representational response, but which allows the measurement of both types of relation while doing violence to neither, would be a welcome contribution. Others have proposed a complete separation of these two facets of associative relationship (Garskof, Houston, & Ehrlich, 1963), and it may eventually be necessary to do so.

Measures of Free Association

The following summary of word association measures is restricted to those derived from single-response free association data, and to those which have been, or reasonably may be, used to predict free recall. Two hypothetical response distributions are shown in Table 1. Values of the association measures, computed from the data of Table 1, are shown in Table 2.

Index of Total Association (*ITA*)

The index of total association (*ITA*), first proposed by Marshall and Cofer (1963) estimates the total associative linkage in a list of two or more words. It considers the response frequencies of all words which occur as responses to two or more stimuli (counting the representational response). All response frequencies, except those of responses which occur to only one stimulus, are summed and

Table 1

Distributions of Hypothetical Response Frequencies for 2 Stimuli

(n = 100)

Response	Stimulus	
	Cheese	Cottage
Burger	50	10
Cheddar	20	10
Cheese	0	20
Cottage	10	0
Cow	0	10
Cream	0	10
Dairy	0	10
Mouse	20	10
Rat	0	10
Swiss	0	10
Sum of frequencies	100	100

Table 2

Association Values from Hypothetical Data[a]

Measure	Word		
	Cheese	Cottage	Both
ITA			0.60
Nc	1.00	1.00	1.00
FE	0.20	0.10	
IIAS			0.15
MF[b]			0.30
MF[c]			0.30
*r**[b]			0.32
*r**[c]			0.45
*Cov**[b]			0.10
*Cov**[c]			0.09
Variance[b]	0.34	0.28	
Variance[c]	0.34	0.12	

[a] from Table 1

[b] with the representational response

[c] without the representational response

divided by the total number of responses elicited by the list; with k stimuli and n subjects the divisor is nk. *ITA* can be represented in an equation as

$$ITA = \frac{\sum_{1}^{k} \sum_{1}^{m} f_{common}}{nk}$$

where f_{common} is the frequency in any one response distribution of a response elicited by more than one stimulus,

k is the number of stimuli,

m is the number of responses given to the k^{th} stimulus,

n is the number of subjects.

ITA can range from 0.0 to 1.0. In the response distributions in Table 1, there are three words ("burger," "cheddar," and "mouse") which are elicited by both stimuli. The sum of the frequencies of these three responses is 120. As there are 200 responses in all (2 stimuli times 100 subjects), the *ITA* for this two item list is 120/200, or 0.6.

ITA may be considered to index the minimum conditions necessary for a list of words to be held together by their associations. Given a listing of all the different association responses to a set of stimuli, *ITA* answers the question, "What is the probability that a response to stimulus i, picked at random from this list, also occurs as a response to some other stimulus in the list?" It is the probability that any given response is not unique to one stimulus. A rationale for *ITA* might state that it was desired to develop an index which reflects as much of the associative linkage within a list as possible, given the minimum restrictions on the criterion for associative relationships.

Marshall (1967) investigated the relation between *ITA* and recall in sublists of four words each, embedded in lists of 24 words. In each list, all sublists had the same value of *ITA*; the level of *ITA* was varied across lists. Analysis of variance showed a significant effect of *ITA* on recall ($p<.01$). Kelley's epsilon (Kelley, 1935; Peters & Van Voorhis, 1940) applied to the results yields a correlation between *ITA* and recall of +.39. *ITA* was also related to clustering ($p<.01$; $\varepsilon = +.65$) and to recognition association, a task requiring the subjects to sort the randomized 24 item list into the six original sublists ($p<.01$; $\varepsilon = +.52$).

The effect of sublist *ITA* on sublist recall was significant but relatively small; *ITA* was more strongly related to clustering and recognition association. No information is yet available on the relation between *ITA* and recall of entire lists of words.

Forms of use. The *ITA* was developed strictly as a measure to be applied to lists or sublists of words. It could be adapted to individual words by taking the proportion of responses to any word which were also given as responses to some other word on the list. To date, such an application has not been made.

Cue Number (*Nc*)

The cue number estimates the associative relation between any one word and all the other words on a list. For any word *w*, *Nc* is defined as the number of list words which elicit *w* in free association, from more than one per cent of the subjects taking the word association test. In a list of *k* words, *Nc* can range from 0 to *k*. As it is rare for a word to elicit itself, *Nc* is more commonly considered to range from 0 to *k*-1. In Table 1, "cheese" and "cottage" are each elicited by the other, and so they each have an *Nc* of 1. Their average is, of course, also 1.

The extent to which a single word *w* is elicited by all the others in a list seems, intuitively, an appropriate measure of the direct associative relation of *w* to the rest. The number of words that elicit *w* is perhaps a rather crude estimate of the relation; because of its very crudeness, however, it may be relatively stable and invariant across subject samples or trials. It is, furthermore, very simple to compute, and may therefore by indicated when a quick, simple measure is desired.

Rothkopf and Coke (1961a, 1961b) introduced the *Nc* and computed its value for 99 words from the Kent-Rosanoff (1910) word list, according to the early Minnesota norms (Russell & Jenkins, 1954). When all 99 words were presented in a free recall task, *Nc* was correlated +.62 with frequency of recall of each word. However, 26 of the 99 words had an *Nc* of 0. As the authors did not report the correlation between *Nc* and recall with these words removed from the analysis, it is uncertain how the correlation was affected by the zero values.

Ninety-nine words is an exceptionally long word list, and it is possible that different associative mechanisms are involved in the recall of such a list from those involved in the recall of shorter lists. Pollio and Christy (1964) examined the relation between *Nc* and recall of each word from 28 item lists. Each of their three lists contained six buffer words *Nc* = 0 in all cases) and the 22 most common responses to "slow," "chair," or "music" (from unpublished norms). In the "slow" list, *Nc* correlated +.63 ($p < .01$) with recall of individual words. In the "chair" and "music" lists respectively, the correlations were +.44 ($p < .05$) and −.53 ($p < .01$). It is not clear why there was a negative correlation in the "music" list.

Forms of use. *Nc* was developed for prediction of recall of individual words, and has not been applied to recall of lists of words. It would be fairly simple, however, to use the average *Nc* in a list as a predictor of list recall. If desired, *Nc* can easily be rescaled so that it ranges from 0 to 1.0, without loss of information.

Frequency of Elicitation (*FE*)

FE is defined as the average proportional frequency with which a word occurs as a free association response to all the other words on a list. For any word *w*, *FE* is computed by summing its frequencies of occurrence to all other list words.

When this sum is divided by $n(k-1)$, the number of subjects times one less than the number of stimuli, *FE* can range from 0.0 to 1.0. A value of 1.0, however, indicates that w was the only response given to every other word on the list. The computation of *FE* is similar to that of *Nc*, but incorporates different information. *Nc* describes how many words elicit w, while *FE* describes how often it was elicited overall. *FE* can be described in an equation as

$$FE = \frac{\sum_{1}^{k} f_w}{n(k-1)}$$

where f_w is the frequency of occurrence of w as a response to any one stimulus,

k is the number of stimuli,

n is the number of subjects.

In Table 1, "cheese" is elicited by cottage 20 times. Since, in this case, $k-1 = 1$, the *FE* for "cheese" is 20/100, or 0.20. Similarly, "cottage" is elicited by "cheese" 10 times, and thus has an *FE* of 0.10.

FE has some intuitive appeal as a predictor of free recall. If recall is considered to result from chaining of associations (as Deese, 1961b, considers it to be), then *FE* seems a direct measure of the probability of an individual word being recalled. Rothkopf and Coke (1961b), however, found *Nc* to predict recall better than *FE*. The correlation between *FE* and recall was +.54 ($p < .01$), while *Nc* and recall correlated +.62. The superiority of *Nc* is marginally significant ($p < .10$). Deese (1959b) has used *FE* to predict the occurrence of w as an intrusion in free recall. He found that with short lists of 12 items, *FE* correlated +.87 with frequency of intrusion.

Forms of use. Like *Nc*, *FE* is designed for use with individual words. Again, the average *FE* could be used to measure associative linkage in a list of words. There would be little point in doing so, however, since the average *FE* is proportional to the next measure to be discussed.

Inter-Item Associative Strength (*IIAS*)

The *IIAS* was first proposed by Deese (1959a) and has become a widely used index. It is defined as the average frequency with which list items elicit each other in free association. Only those responses which also occur in the stimulus list are considered. Again, when the summed frequencies are divided by nk, *IIAS* can range from 0.0 to 1.0. The *IIAS* differs from the sum of the *FE's* by a proportional constant; *IIAS* times $k/(k-1)$ is equal to the sum of *FE's*. The formula for *IIAS* can be represented as

$$IIAS = \frac{\sum_{1}^{k} \sum_{1}^{m} f_w}{nk}$$

where w is any word from the stimulus list,

f_w is the frequency of occurrence of w as a response to any stimulus item,

k is the number of stimuli

m is the number of responses given to the k^{th} stimulus,

n is the number of subjects.

In Table 1, the summed response frequencies of "cheese" and "cottage" equal 30. As there are 200 responses in all, *IIAS* equals 30/200, or 0.15.

The *IIAS* is useful as an estimate of the extent to which items in a list are directly related to each other through elicitation of list members. It can be considered as the average probability that any association will be a list item. Its use in the prediction of recall of lists of words is analogous to the use of the *FE* in prediction of recall of individual words. If recall occurs through simple chaining of associations, and if the associative process is constant throughout the chain, then *IIAS* is the average probability of staying within the stimulus list at any point in the chain and should therefore be proportional to the average number of words recalled from the list.

In his first report of the measure, Deese (1959a) correlated *IIAS* and total recall over 18 lists of 15 items each, and reported a correlation of +.88. In a later experiment (Deese, 1961a), the correlation dropped to +.67 for the same word lists. Simon and Hess (1965) confirmed Deese's results with grade school children. For grades four through six, they obtained a correlation of +.65 between *IIAS* and total recall. Bousfield, Steward, and Cowan (1964) related *IIAS* to recall of eight-item sublists, embedded in lists of forty items. The rank order correlation between *IIAS* and recall of sublists was +.77. In similar experiments by Cohen (1963a, 1963b), three and four-item sublists were composed of exhaustive (*E*) or non-exhaustive (*NE*) instances of categories. An example of an *E* sublist is "winter, spring, summer, fall"--these names exhaust the category of "seasons." Four names of animals would constitute an *NE* sublist. The sublists were embedded in a 70 item list. In the first study (Cohen, 1963a), *IIAS* and recall correlated +.78 and +.63 in two different lists, across the combined *E* and *NE* sublists. In the second study (Cohen, 1963b), *IIAS* correlated +.59 and +.50 with recall of the *E* and *NE* sublists respectively.

IIAS has been related to other tasks besides recall. Willner and Reitz (1965) required their subjects to sort a list of words into sublists. They found a rank order correlation of +.74 between number of sublists and *IIAS* of the original lists.

Weingartner (1963) investigated the relation between *IIAS* and serial order in a serial anticipation task. His 16-item lists were composed of 4-item sublists; each sublist was high in *IIAS*. When the sublists were presented intact within the list, the number of trials to criterion (two errorless serial anticipation trials) was significantly lower than when the words were scattered randomly through the list. Postman (1967) suggests that there is an interaction between *IIAS* (pre-experimental associative bonding) and contextual effects (the artifical associations required in a serial anticipation task). When the words are to be learned in their "natural" order, with sublists intact, then high *IIAS* may facilitate learning. When the "natural" order is disrupted, high *IIAS* can impede learning. When the task involves a single relatively long list, with separate sublists, then high *IIAS* can be expected to interfere with learning the list in almost any specific order. Postman found that in a serial anticipation task, randomly arranged high *IIAS* lists took more trials to learn to criterion than did zero *IIAS* lists, on both original learning and relearning. The high *IIAS* lists were retained better over a seven day period, presumably due to their greater cohesiveness; once an item from the list was recalled, it was likely to trigger the other items into recall.

Forms of use. As stated above, the *IIAS* and the *FE* are very close to being alternate forms of each other. Little would be gained by adapting the *IIAS* for use with individual words.

Common Elements Correlation or Mutual Frequency Index (*MF*)

The *MF* index was first used to compare word association distributions by P.M. Jenkins and Cofer (1957). Since then it has been widely used, usually with the representational response, and has become the most common association measure in use. It has not, however, been extensively applied to the prediction of free recall. The *MF* for two stimuli S_1 and S_2 is defined as the number of responses given in common to S_1 and S_2, divided by the total number of responses given to them. The denominator is equal to n when the representational response is not included, and equal to $2n$ when it is. In both cases, *MF* can range from 0.0 to 1.0. It can be represented in an equation as

$$MF_{ij} = \frac{\sum_{1}^{m} f_w}{n}$$

where i and j are stimuli,

f_w is the common frequency of occurrence of w as response to both i and j (the common frequency is the smaller of the two frequencies),

m is the number of responses given to both i and j,

n is the number of subjects.

In Table 1, there are three responses--"burger," "cheddar," and "mouse" which are given as responses to both "cheese" and "cottage". All three are emitted less frequently to "cottage" than to "cheese". The common frequency of each response is thus the frequency of its occurrence to "cottage". The frequencies of the three common responses sum to 30; as there are 100 responses given to each word, the *MF* score, without the representational response, is 30/100, or 0.30.

With the representational response included, the common frequencies are supplemented by the frequency with which "cottage" elicits "cheese" (20) and the frequency with which "cheese" elicits "cottage" (10). The sum of the common frequencies is thus 60; as the total number of responses is raised to 200 ($2n$), the *MF* score with the representational response equals 60/200, or, again, 0.30. In this instance, the *MF* score is not affected by addition of the representational response.

In comparing the response distributions to two stimuli, the *MF* takes into account all of the responses common to the two stimuli, including, if desired, the representational response. Furthermore, it considers the frequencies of the common responses as they occur to both stimuli. It uses, therefore, almost all of the single-response associative information available, in comparing the two distributions.

Use of the *MF* without representational response has not been common. Bousfield and Puff (1965a) called it the mediator overlap ratio (*MOR*) when used without the representational response. In their experiment, 26 related pairs of words were scattered in a 52 item list; the criterion measure was the frequency with which the pairs were reunited in recall. *MOR* correlated +.88 with "forward clustering," i.e., clustering of word pairs in the same order in which they had appeared on the stimulus list. *MOR* did not correlate significantly with "backward" or total clustering.

Rosenzweig (1964; Rosenzweig & Miller, 1966) used the *MF* measure without the representational response to compare the associations of different cultural groups to the same stimuli. He found that French workmen and French students have quite distinct associative patterns. American workmen and students, conversely, have much more similar associative patterns.

Using the *MF* in its full form, with the representational response, Deese (1965) reported a correlation of +.82 between average *MF* and total recall of lists of 15 words. In the study by Bousfield and Puff (1965a), cited above, the *MF* with the representational response was also correlated with clustering. *MF* did not significantly correlate with either forward or backward clustering, but correlated +.82 with total clustering. In an earlier study, Bousfield, Whitmarsh, and Berkowitz (1960) found a correlation between *MF* and clustering of +.58.

In the first use reported of the *MF*, P.M. Jenkins and Cofer (1957) compared the response distributions to compound stimuli (eg., "quiet woman") with the response distributions to the components of the compound stimuli (eg., "quiet" and "woman"). They concluded that the compound stimuli functioned very differently from their components in generating associations. Cofer (1957) found a strong degree of relation between *MF* and rated similarity of pairs of adjectives.

Mathematical history and critique. Although the *MF* was first used as an association measure in 1957, it has a much longer history in statistical theory. Deese (1962) has pointed out that the *MF* is a simple algebraic derivation from the common elements form of the product moment correlation. This form of correlation was first suggested by Spearman (1904) and was developed by Thomson (1919, 1935, 1951) in connection with his objections to Spearman's theory of general intelligence. The mathematical rationale for the measure is presented in Kelley (1927) and the equation is derived in detail by Peters and van Voorhis (1940). The common elements formula is applicable in a situation in which the response to a stimulus variable may be considered a random sampling from a pool of available elements. The equation for the common elements correlation is

$$r_{AB} = \frac{n_c}{\sqrt{n_A + n_c}\,\sqrt{n_B + n_c}}$$

where A and B are stimulus variables,

n_c is the number of elements common to A and B,

n_A is the number of elements in A but not in B,

n_B is the number of elements in B but not in A.

If each free association response is considered as an element, then:

$n_A + n_c = n_B + n_c = n$, where n is the number of subjects.

The equation thus simplifies to $r_{AB} = n_c / n$.

The final form of the common elements equation closely resembles the formula for the *MF* index, given above. They cannot be assumed to be equivalent, however, as two rather stringent criteria must be met in deriving the common elements equation from the general equation for the product-moment correlation. Each response is first partitioned into common and unique parts, analogous to the partitioning of common and unique factors. Calling the common parts c, and the unique parts from responses A and B, a and b respectively, then $A = a + c$ and $B = b + c$. The first restriction is that a, b, and c must all be uncorrelated. The second restriction is that a, b, and c must all have equal variances.

Formally, the common elements correlation requires a model quite different from that found in word association distributions. The common elements are supposed to be exactly the same elements, simultaneously entering two score distributions. It is a little uncertain whether the same response elicited by two stimuli can be considered a single event in this way. Furthermore, the equal variances referred to above are binomial variances, rather than sums of squares. For the binomial variances to be the same in a word association distribution, each entry in the partitioned response vectors (assuming a word to be the "element") would have to be equal, or at least not differ significantly from equality. The derivation of the common elements correlation in fact assumes that the probability of occurrence is equal for each element. This requirement implies that every response should have the same frequency, within the limits of measurement or sampling error. A glance at any table of word association distributions, however, confirms that there are gross differences in the probabilities of occurrence of different responses.

An example might help to illustrate the appropriate use of the model. Suppose that there are an equal number of red, green, and blue balls in a basket, and that they are randomly picked, ten at a time, with replacement (or are picked from an infinite sample). It is possible to make score distributions of the number of red and green balls combined, and of the number of blue and green balls combined, selected on each trial. The distribution of red + green is A, and the distribution of blue + green is B; a is the distribution of red balls, b the distribution of blue, and c is the distribution of the common elements, the green balls. The overall proportion of green balls will approach 0.33 with repeated trials, 1/3 of the number of balls selected. This proportion is the common elements correlation.

It is unlikely that the common elements model can be successfully adapted to description of word association relationships. Response words may be theoretically considered as elements, but they do not behave as common elements should, according to the model. They are not equally likely to occur, and the binomial variances of the partitioned vectors are in no way constrained to be equal.

The failure of the *MF* index to satisfy the requirements of the statistical model implies simply that it cannot be considered a legitimate product-moment correlation. For many purposes, of course, it may be irrelevant whether it is a product-moment correlation or not. In relating the *MF* to recall or clustering, the success of the measure as a predictor is the criterion of its usefulness.

In cases where *MF* matrices have been factored to study the associative characteristics of language, the nature of the coefficients is a somewhat more serious problem. Factor loadings are not usually subject to empirical confirmation, and may be seriously distorted if based upon measures that do not qualify as correlation coefficients. Although *MF* matrices are probably Gramian (at least, no negative eigenvalues have been reported), they may result in faulty estimates of the common factor structure. To date, none of the reported studies that have included factoring of *MF* matrices (Deese, 1962, 1964, 1965; Howe, 1966) have examined the status of the *MF* as a correlation coefficient.

Conditional Probabilities Correlation (r^*)

In any task calling for single, discrete, separable responses to each stimulus, the response distribution can be considered a set of conditional probabilities. The responses, that is, are described as being conditional upon the occurrence of the particular stimulus. This situation exists with word association data, and with other data such as that from forced-choice psychophysical tasks. Treating associative response distributions as conditional probabilities highlights the necessity of the assumption of the absence of individual differences on the distribution of associations. In fact, of course, simple frequency distributions such as are used in computing the *MF* rely equally heavily upon the assumption.

With m unique responses to each stimulus, each set of conditional probabilities describes a vector in m-space. The cosine of the angle between any two vectors is the correlation between the two vectors and can be taken as a measure of the similarity of the two stimuli (Rosner, 1956). The equation for r^*, using Rosner's notation, has the form:

$$\frac{\sum_k p_i(k) p_j(k)}{\sqrt{\sum_k p_i(k)^2 \sum_k p_j(k)^2}}$$

where $p_i\ (k)$ is the probability of response k, given stimulus i,

$p_j\ (k)$ is the probability of response k, given stimulus j,

$\Sigma\ p_i\ (k)^2$ is the sum of squares of the proportions on the i^{th} response vector (i.e., the sum of squares of the proportional frequencies of the responses to the i^{th} stimulus).

The numerator is the sum of cross-products of proportional frequencies, so the measure is always positive or zero. It can range from 0.0 to 1.0. When the representational response is included, the formula is unchanged and appropriate additions are made to the data matrix. Rosner (1956) discusses the ways in which this measure is equivalent to and different from the traditional Pearsonian product-moment correlation.

In Table 1, the conversion of the response frequencies to conditional probabilities is accomplished simply by dividing the observed frequencies by 100, the number of subjects. The conditional probability of occurrence of "burger" as a response to "cheese" is 0.50; to "cottage," it is 0.10. The cross product of the conditional probabilities is 0.05. Similarly, the cross products of conditional probabilities to "cheddar," and "mouse" are 0.02 and 0.02 respectively. The sum of squares on the "cheese" distribution is $0.5^2 + 0.2^2 + 0.1^2 + 0.2^2$, and equals 0.34. Similarly, the sum of squares on the "cottage" distribution equals 0.12. The correlation is equal to the sum of cross products (0.05 + 0.02 + 0.02 = 0.09) divided by the geometric mean (square root of the product) of the sums of squares. The square root of the product is equal to 0.202; the conditional probabilities correlation, without the representational response, is thus equal to 0.09/0.202 or approximately 0.45.

When the representational response is included, the number of responses is doubled; therefore, the proportional frequency of all responses other than the representational response is halved. The sum of the proportions is 1.0 in any case. In Table 3, the response distributions of Table 1 are expressed in the form of conditional probabilities with the representational response included. The procedure for determining the value of the conditional probabilities correlation is exactly the same as in the previous case. Note that the conditional probabilities correlation drops to 0.32 when the representational response is added; in the case of the *MF* score, addition of the representational response made no difference to the value of the index. The *MF* is not necessarily less sensitive on that account; the two measures simply behave differently.

Like the *MF* index, the conditional probabilities correlation takes into account all of the common responses to two stimuli, and the frequency of the common responses. It also considers the variances of the response distributions, but in a compensatory way; that is, it negates the effect of them. The choice between the *MF* index and the conditional probabilities correlation is made on the basis of how they use the information coming from the response distributions, rather than on the basis of what information they use. If the conditional probabilities correlation performs better than the *MF* index as a predictor of recall, then it is clearly preferable to the *MF* on empirical grounds. Even if

Table 3

Distributions of Hypothetical Response Proportions with the Representational Response Included[a]

Response	Stimulus	
	Cheese	Cottage
Burger	.25	.05
Cheddar	.10	.05
Cheese	.50	.10
Cottage	.05	.50
Cow	.00	.05
Cream	.00	.05
Dairy	.00	.05
Mouse	.10	.05
Rat	.00	.05
Swiss	.00	.05
Sum of Proportions	1.00	1.00

[a] Based on data from Table 1.

it performs only as well as the *MF* index, it may still be preferable on logical grounds.

The conditional probabilities correlation is a general measure, and is statistically appropriate for use with any distribution of discrete responses. The appropriate variant on the standard product-moment correlation in all cases where the data are frequencies assigned to nominally described responses. It has been used in psychophysical experiments involving the method of absolute judgements (Rosner, 1956), and in experiments on psychological diagnosis, where it was desired to compare response distributions to each pair of a set of Rorscach stimuli (Kendall, 1962). No previous word association studies have been found in which it is used.

Conditional Probabilities Covariance (*Cov**)

Correlations are computed by dividing the covariance by the geometric mean of the variances of the two distributions; the division is necessary to make the correlation independent of the variances. Usually, it is desirable to correct for unequal variances, as the metric of the variables is frequently arbitrary (eg., inches or feet) and is often different for the two variables (eg., inches and seconds). In the case of conditional probabilities derived from word associations, the metric is the same for all variables and is defined by the response distribution itself. The variances of the distributions are therefore comparable and provide real information about the shape of the distributions. Very high variance indicates that each of a small number of responses was emitted by many subjects; very low variance indicates that many subjects each gave different responses to the stimulus word. Defining similarity between two distributions as their covariance, retains information (deliberately removed from the correlation measure) about the shapes of the distributions. The covariance formula is simply the numerator of the correlation formula, given above, and equals the sum of the cross-products of the conditional probabilities.

Without the representational response, the covariance can range from 0.0 to 1.0. A value of 1.0 indicates that two stimuli each elicited only one response from all subjects, and that it was the same response to both stimuli. In this same case, the variances of each distributions are also 1.0. With the representational response, however, the covariance can range only from 0.0 to 0.5, and the variance (sum of squares) can range only from 0.25 to 0.50. The reason for this limitation is that the representational response is different from the written response. For the variances and the covariance to equal 1.0, each response distribution must be composed of just one word. But with the representational response, every response distribution contains at least two responses.

A simple form of the conditional probabilities covariance has been used by Rothkopf (1960). He gathered continuous associations from pictures of tools, and restricted the associations to names of the tools, names of their parts, descriptive adjectives, and uses of the tools. He calculated the conditional probabilities covariance only from the ten most frequent associations to each stimulus. When the pictures were used as stimuli in paired-associate learning, the covariance was found to correlate +.54 to +.60 with substitutions of responses experimentally paired to other stimuli in the learning task.

Multivariate Statistics Derived from Bivariate Measures of Association

One of the most useful features of correlation and covariance measures is that they can be used as the basis for computing multivariate statistics. Appropriate multivariate statistics can be derived for many selected levels of analysis. Bivariate measures of association can, therefore, serve as the basis for summary measures of association within an entire list of words, or for measures of the relation of one word to all the rest. To the extent that the *MF* index behaves like a true correlation coefficient, it is appropriate to compute the multivariate statistics from it as well as from the conditional probabilities measures.

The Mean Correlation or Mean of the Squared Correlations in a Matrix

The average value of all the elements in a correlation matrix is a simple and obvious summary measure of the overall level of association between all the variables in the matrix. In using the *MF* index to describe the level of association in lists of words, the average of the *MF* scores in a matrix is the measure that has been generally used. For the sake of comparison with the results of previous studies, the average value in the matrix is therefore useful. In general, the mean of the squared indices might be a preferable measure to use. A squared product moment correlation is equal to the percentage variance common to the variables; the mean squared correlation is thus equal to the average percentage of common variance. The square root of the mean of the squared correlations could be taken as the average correlation within the matrix; as such, it provides an estimate of the overall association value of the matrix in the familiar metric of bivariate correlations.

The Determinant

The determinant of a correlation matrix is another summary measure applicable to the description of entire lists of words. As such, it serves a function similar to the average, or average squared, correlation in the matrix. The determinant, however, is unusually sensitive to extremely high correlations. If any variable in the matrix is perfectly predictable from the rest, the determinant of the matrix becomes zero.

Rozeboom (1965, 1968), among others, discusses the interpretation of the determinant of a covariance matrix as the generalized variance of the mulitvariate distribution. It is possible to rescale the metric of the determinant back to that of the original variables by taking the $2k^{th}$ root of the determinant, where k is the number of variables in the matrix. This transformation could be considered to result in the generalized average standard deviation within the matrix. Conceived as a variance or standard deviation measure, the determinant can be considered an estimate of the homogeneity of the elements within the matrix. If there are definable clusters of variables in the matrix, that are highly correlated within themselves but are not correlated with the remaining variables, the determinant may reflect the degree of clustering. Such a matrix has a smaller determinant than a matrix which has the same mean association value but in which the elements are more closely grouped around the mean. In relating free association to free recall, it may be desirable to compose recall lists which contain separate sublists, the sublists being more or less related in different total lists. The determinant provides a summary measure of association which takes into account the relations between the sublists as well as the overall level of association among all the words.

The Squared Multiple Correlation (SMC)

The squared multiple correlation, or SMC, of one variable with a set of others, is the proportion of the variance of the first variable which is predictable from a linear combination of the rest. The square root of the SMC is the multiple correlation, and can be interpreted in the same way as a bivariate correlation coefficient. Since the SMC equals the total variance accounted for in the criterion variable, it is a very useful measure of relation between one variable and the rest, provided that the bivariate measures are true correlations. The squared multiple covariance is a similar measure, and is computed from a covariance matrix in the same way the SMC is computed from a correlation matrix. There is a simple relation between squared multiple covarinace and SMC; the squared multiple covariance, divided by the variance of the criterion variable, is equal to the SMC.

It should be emphasized that whenever the bivariate association indices can be considered "true" correlations, the SMC is the most appropriate measure of the relation between one variable and all the rest. It makes full use of the information available in the data from the sample on which the measurements are taken. As it has a known sampling distribution, significance tests can be applied to it. In general, greater confidence can be attached to the SMC than to any alternative estimate of the relation.

The Mean of the Squared Correlations of one Variable With the Rest

The average squared correlation of one variable with a set of others is equal to the average variance of the criterion variable accounted for by all the other variables. It is always equal to or smaller than the SMC, since the SMC is equal to or larger than the largest squared correlation. It is a less exact measure than the SMC, as it does not consider the correlations among the predictor variables. As is the case with the mean of the squared correlations in an entire matrix, the square root of the mean of the squared correlations can be considered as an average correlation. The root of the mean squared correlation is, again, in the familiar metric of single bivariate correlations.

The Variance of the Response Distribution

The variance, or sum of squares, describes the compactness of the association distribution. If every subject gives the same response to a stimulus, the variance for that stimulus word is 1.0. If every subject gives a different response, the variance is $1/n$. If free recall of a word is related to the "tightness" of the association distribution, the variance should provide meaningful predictive information. The variance of an associative distribution is not itself a bivariate measure, of course, since it is computed from the response distribution to a single stimulus. It is derived from a bivariate measure however; the main diagonals of the conditional probabilities covariance matrices contain the variances for each response vector. It may be simplest to consider the variance as a special case of the covariance; the sum of squares is the sum of "cross-products" of each variable with itself.

The sum of squares differs from the usual computation of variance in that the sum of squares does not take into account the mean frequency; variances are customarily expressed as squared deviations from the mean. The sum of squares (or squared deviations from zero) is, however, customarily used as a variance measure of a distribution of proportions (Rosner, 1956).

There is a familial resemblance between the variance of a distribution of proportions and the concepts of "amount of information" and "redundancy" used in information theory. Miller (1956) has discussed the similarity of the concepts of variance and information in a general sense. The average amount of information necessary to describe a series of events (called H; for a discussion, see Attneave, 1959) has a demonstrably close mathematical relationship to variance (Rozeboom, 1968). The nature of the relationship is not altered when the variances are calculated from conditional probabilities; the relationship is slightly more apparent in such a case perhaps, as the basic data in information theory are also probabilities.

Additional Measures

Numerous other measures, not experimentally related to free recall, have appeared in the literature. There are described here only briefly, as they are not otherwise used in the present study.

Jenkins and Russell (1952; Jenkins, Mink, & Russell, 1958) related clustering of word pairs in free recall to their average frequency of eliciting each other in free association. Bousfield and Puff (1965a, 1965b) used a similar measure, but defined response frequency as 1.0 plus the log of the cultural frequency. Bousfield, Steward, and Cowan (1964) developed the measure of stimulus equivalence (*MSE*), defined as the number of responses elicited by two or more stimuli, and related it to the clustering of words in recall. *MSE* is similar to the *ITA* described previously, but the *ITA* considers the frequency of the responses as well as their occurrence. Marshall and Cofer (1963) developed the index of concept cohesiveness (*ICC*), defined as the number of responses elicited by all stimuli, divided by the *MSE*. Finally, Pollio (1966) used an index equal to the average cue number to structure the semantic space of associations in a rating task.

Non-Associative Factors in Free Recall

There are many factors relating to free recall besides associative ones. Additional factors may include, for instance, list length, word frequency, word pronounceability, approximation to prose or common speech, primacy, recency, distinctiveness of individual words in a list (von Restorff effect), and serial position of list items. In addition, other experimental variables are involved, such as rate and method of presentation, length of warmup period, amount of information given to subjects, etc. The number of relevant variables increases still more when repeated testing is introduced. Although each of these factors can be investigated, only word frequency and serial position were varied in the present study.

Serial Position

In free recall, the last items in a list tend to be recalled best and earliest. The first items are recalled next, and moderately well, and the middle are recalled last and worst (Deese & Kaufman, 1957; Bousfield, Cohen & Silva, 1956). The exact shape of the serial position curve depends, of course, on the material presented; the generalization applies best to relatively homogeneous lists of words (homogeneous, that is, with respect to their association value and distinctiveness). The effect is a strong one, and must be taken into consideration in predicting recall of single words, lest it obscure other factors of interest. Serial position can be controlled either statistically or experimentally. For statistical control, the serial position of the words in the list (expressed as distance from the middle of the list) can be used as one of the predictors of recall. Alternately, the effect of serial position can be partialled from the correlations of each predictor with recall. For experimental control, the first and last few words in the list can be chosen so as to have extremely low association value. The inclusion of these "buffer" words, which are deleted from the prediction analysis, serves to minimize any possible interaction between serial position and associative factors, especially since the serial position effect is strongest at the ends of the list.

Word Frequency

Deese (1960) varied the list length and average word frequency of several lists of words chosen from the Thorndike-Lorge (1944) general word count. He found that average list recall was affected by both word frequency and list length, and by their interaction. More detailed analysis of his results showed, however, that much of the effect of both variables was attributable to the *IIAS* levels of the lists. Although the lists were composed of randomly chosen words, the high frequency words, especially in the longer lists, imparted a significant *IIAS* level to the lists. The level of *IIAS* in the lists was closely related to recall. In lists with zero *IIAS*, word frequency was unrelated to amount of recall.

Waugh (1961) tested subjects on repeated trial free recall of a list of 48 words. The number of trials to the learning of each word in the list was unrelated to Thorndike-Lorge frequency.

These studies indicate that the independent effect of word frequency on free recall is probably not great. Nevertheless, since the question is not settled, it is advisable to control for word frequency in constructing recall lists. It is fairly easy to construct lists of words that have the same, or very similar, average word frequencies. In predicting recall of individual words, over lists, the word count frequency of each word is appropriately included as a predictor of recall.

Reliability of Word Association Behaviour

The reliability of word association behaviour has received relatively little experimental attention. Those studies which have investigated reliability have used only one form of reliability, and have used a measure which may not be the best available.

The Meaning of Reliability

Reliability is affected, in general, by three sources of variability, due to the instrument or measure, the observer, and the subject. Instrument variability often receives the greatest emphasis, in attempts to answer such questions as: How reliable is this measure under these conditions? How much random error does it allow into the data? Observer variability has become a prominent focus of interest in recent years; studies of experimenter bias (eg., Rosenthal, 1966) seek to determine the extent to which the experimental data are contaminated by factors of attitude, set, and attention within the observer or experimenter.

In the context of word association experiments, it is most appropriate to focus on subject variability, in an examination of the reliability of the behaviour itself that is being investigated. The data to be analyzed are the observed response distributions themselves, rather than a measure that seeks to summarize them and provide further information about them. The initial measurement of the primary data need go no farther, on the purely descriptive level, than compiling the response distributions. It is the stability or reliability of the emission of these responses that is of interest. Appropriate questions are such as: Is the observed distribution of responses consistent within the population? Is it affected by retesting? Do some words give rise to a more reliable response distribution than others? Different techniques are appropriate in answering the different questions.

Whether it is the measure or the behaviour that is being examined for reliability, reliability is usually considered as a proportion of true variance to total variance, where the total variance is the sum of true variance plus error variance (Guilford, 1956; Gulliksen, 1950). Different techniques for estimating the proportion, such as split-half or test-retest reliability, may result in rather different results, as they differ in the portions of the variance which they treat as error.

Proportion of Identical Associations

The proportion of subjects who give the same response to a stimulus on two administrations of a word association test is called the proportion of identical associations. All of the reported studies on the reliability of associations have used this measure.

Hall (1966) studied the one and three week test-retest reliability of 54 words chosen randomly from the Thorndike-Lorge (1944) word count. The proportion of identical associations from the 61 subjects was around 0.50, and was unrelated to retest interval or word frequency. Brotsky and Linton (1967) studied the ten week test-retest reliability of the 100 Kent-Rosanoff words and of 58 words from the Connecticut norms (Bousfield, Cohen, Whitmarsh, & Kincaid, 1961). They found a proportion of identical associations of 0.32. They attributed the lower proportion they found, compared with that of Hall (1966), to the greater period of time between their test and to the higher average word frequency in their lists. They found a correlation of -.34 between word frequency and proportion of identical associations. Gegoski and Riegel (1967) reported an average proportion of identical associations of 0.28 on a similar task.

Correlation Measures and Reliability

The proportion of identical associations is an appropriate reliability measure to use when the question of interest is the within-subject stability of responses over trials. In many situations however, it may be necessary to know the reliability of an entire distribution of associations over all subjects in the sample. Such a situation arises in conjunction with all of the bivariate association measures considered in this study. If two response distributions are found to be unstable in the sample, then the level of observed association between them is probably unstable as well, and not amenable to interpretation. Since the association measures are based on response distributions computed from an entire sample of subjects, reliability information about the distributions is appropriate and necessary. Reliability of the overall distribution within the pooled sample is not, however, necessarily related to the proportion of identical associations, and cannot be estimated from it *a priori.*

An appropriate measure to use in estimating the reliability of a sample association distribution is the conditional probabilities correlation, without the representational response. This measure provides a correlation between response distributions to the same stimulus, from two samples or testing sessions; it is similar to a product-moment correlation, and thus provides reliability estimates that can be compared meaningfully with those obtained from other forms of data. It may be used in two ways. First, the response distributions from the same set of subjects may be compared on two test administrations to obtain a test-retest reliability coefficient for each response vector. Second, on one administration of the test, the response distributions of half the sample may be compared to those of the other half. This latter method, called scale reliability or split-sample reliability, is somewhat analogous to split-half reliability, in which the responses to item subsets are compared. Noble (1955) has described the rationale for the measurement of scale reliability, and has shown how the Spearman-Brown correction applies.

The Assumption of the Absence of Individual Differences

Use of the proportion of identical associations can provide evidence for assessment of the assumption that people do not significantly differ in terms of their associative hierarchies, provided that the two administrations of the word association test can be considered statistically independent. If the test-retest reliabilities are high, statistical independence can be assumed fairly safely. The occurrence of any response to a given stimulus is an event mutually exclusive with the occurrence of any other response. The probability of occurrence of the same event twice (on two independent trials) is simply the square of the original probability. The predicted probability, therefore, that any subject will give a particular response to a given stimulus on both tests is simply the square of its conditional probability, assuming that the sample conditional probability in fact expresses each subject's conditional probability. This assumption is, of course, a restatement of the absence-of-individual-differences assumption.

The probability that any one of a set of mutually exclusive events will occur is the sum of their individual probabilities. Since the predicted probability that a particular response will be repeated on retesting is equal to the square of its probability of occurrence, the predicted probability that any response at all will be repeated is equal to the sum of squares of conditional probabilities in the response distribution collected from the entire sample.

The observed probability in the sample that any response will be repeated is of course the proportion of identical associations. The prediction, therefore, is that if the absence-of-individual-differences assumption is valid, the variance of each response distribution will equal the proportion of identical associations given to the eliciting stimulus. Furthermore, it should make no difference whether the proportions of identical associations are calculated from actual test-retest data on single subjects, or from random pairings of subjects from one administration to the next. If there are in fact no differences between subjects, the basis for the pairing is irrelevant.

If the predicted relation holds, then the assumption will receive quite strong support. If it does not, the credibility of the assumption is necessarily reduced.

The Present Investigation

Several questions have been raised with respect to experimentation on word associations and free recall. The present investigation attempts to shed light on some of these questions, and specifically addresses itself to the following:

1. A comparison of the ability of several summary measures of word association to predict average recall of lists of words.
2. A comparison of the ability of several related measures to predict the recall of single words from the lists.
3. The determination of several measures of reliability of word association behaviour, and a comparison of the measures.

4. An assessment of the validity of the assumption of the absence of individual differences on the emission of associations.

Chapter 2: Method

General Paradigm

The first purpose of this study is to determine the possible extent of prediction of free recall from word association information. A second purpose is to investigate several aspects of the reliability of word associations. It is necessary, for these purposes, to gather data from three experimental sessions with the same subjects, during the course of the main experiment. A first word association test provides the raw data for determination of the associative predictors of recall. A second word association test, with the same stimuli, provides the raw data for the determination of the test-retest reliability measures. Finally, a recall test provides criterion data for assessment of the predictive efficacy of the associative predictors.

Pilot Study

A pilot study was conducted four months before the main experiment to provide preliminary information for use in selecting items for the word association and recall tests. Subjects for the pilot study were 150 first and second semester psychology students, who gave single written free association responses to a list of 126 stimulus words.

Development of Word Association Test

The 126 words on the word association test were chosen so as to provide a reasonably broad sampling of associative relationships presumed to exist *a priori*. An attempt was made to "sample" three concepts--"cheese," "butterfly," and "student." Accordingly, approximately 35 words associatively related to each concept were included. Selection of words was from two chief sources. All words were included which were given as responses to "cheese" or "butterfly" by at least one per cent of the college sample in the Minnesota norms (Palermo & Jenkins, 1964). These items were supplemented, and the "student" items were chosen, by informal consensus that a word might appropriately be placed in a particular group. To fill out the list, words were added which did not seem to be related to each other or to the words previously selected.

All associative relationships among the words were analyzed using the conditional probabilities covariance and correlation measures, without the representational response. The resulting 126 x 126 matrices were analyzed with a cluster analysis technique (developed by McQuitty, 1957) to isolate sublists of words having common associative properties. The criterion for inclusion of a word in a cluster or type is quite simple; every word belonging to a type must be "more like some other member of that type (with respect to the data analyzed) than it is like any member of any other type (McQuitty, 1957, p. 209)."

Cluster membership was found to be quite a robust phenomenon; there were few major changes in the typal structure when the clustering was later performed on the *MF* matrices, and on all three association matrices with the representational response included.

The Main Experiment

Subjects

The subjects for all phases of the experiment were students in first and second semester psychology courses at Simon Fraser University. All subjects were native speakers of English or were fluent in English as a second language. In all, 362 subjects were tested in the three phases of the experiment: the two administrations of the word association test, and the recall test. The data from 71 subjects who were absent for one or more of the testing sessions were eliminated from the analysis. The data from eight others were eliminated according to one of the criteria used by Palermo and Jenkins (1964). The only criterion for rejection that proved necessary to apply in this sample was that of "response faults;" subjects whose word association tests included more than ten per cent omissions or illegible entries were dropped from the sample. All data analysis was done on the reduced sample of 283 subjects. None of the subjects in this reduced sample had taken part in the pilot study.

Stimulus Materials

Word association test. The cluster structure of the words used in the pilot study was examined to see if words clustered together in the way they had been expected to *a priori*. Several words were omitted which did not join clusters as they were expected to; a few others were eliminated because many subjects were unable to make any response at all to them. The items comprising the final form of the word association test are listed in Table 4.

The word association test was mimeographed (by Gestetner) on 8 1/2" x 11" white paper; fourteen words were printed double spaced on each page. Order of words on each page was fixed after a single randomization. A different randomization was used for the second administration of the test. Order of the nine pages in each test booklet was randomized manually.

Recall test. The items for the recall test were chosen from the clusters derived from the pilot study. Two ten-item sublists related to "cheese" were chosen, and two related to "butterfly." The second sublist for each concept was chosen so as to have a slightly lower average level of association than the first sublist. A fifth sublist was chosen of words having low correlations with each other and with the previously selected words.

From these five sublists, all possible pairs were combined to make ten recall lists. In this way, it was possible to vary between lists both the average correlations and the variance of the correlations. In addition, six extra words were added to each list. These "buffer" words, three on each end of each list, were the same for all lists and had

Table 4

Alphabetical Listing of Stimuli Used in Word Association Test

1. Animal
2. Ant
3. Apple
4. Arch
5. Bacon
6. Beautiful
7. Beauty
8. Bee
9. Best
10. Bird
11. Blue
12. Book
13. Bread
14. Brick
15. Bug
16. Burger
17. Butter
18. Butterfly
19. Cake
20. Canoe
21. Caterpillar
22. Caution
23. Cave
24. Chair
25. Cheddar
26. Cheese
27. Class
28. Cloth
29. Cocoon
30. College
31. Colour
32. Cottage
33. Course
34. Cow
35. Crackers
36. Cream
37. Dairy
38. Dart
39. Delicate
40. Dime
41. Eat
42. Effect
43. Essay
44. Exam
45. Fail
46. Fence
47. First
48. Flag
49. Flight
50. Flower
51. Fly
52. Food
53. Forget
54. Frail
55. Gift
56. Grade
57. Green
58. Ham
59. Harvest
60. Hedge
61. Holes
62. Host
63. House
64. Hut
65. Insect
66. Joke
67. Knife
68. Lamp
69. Learn
70. Lecture
71. Light
72. Mark
73. Metal
74. Milk
75. Mirror
76. Monarch
77. Moon
78. Moth
79. Mouse
80. Nature
81. Net
82. Number
83. Orange
84. Outdoors
85. Pass
86. Pie
87. Pizza
88. Pollen
89. Pretty
90. Rat
91. Research
92. Rye
93. Sand
94. Sandwich
95. Scholar
96. School
97. Sharp
98. Sky
99. Smell
100. Smile
101. Soft
102. Sour
103. Spread
104. Spring
105. Stomach
106. Story
107. Street
108. Student
109. Study
110. Subtract
111. Summer
112. Sun
113. Swiss
114. Table
115. Teacher
116. Text
117. Theory
118. Thesis
119. Tree
120. Trunk
121. Tutor
122. University
123. Wings
124. Worm
125. Write
126. Yellow

low correlations with each other and with all other list items. They were included to provide partial control for primacy and recency effects in recall. The recall test items are listed in Table 5.

The construction of the recall lists was done in such a way as to provide a fixed sample of word lists stratified on the bases of average association level and of homogeneity of associations. The stratification was performed to facilitate generality of conclusions about the word lists; if the stratification was appropriate, the conclusions should be applicable to other word lists that have average association levels and homogeneties of association anywhere within the range used in the present study.

Predictors of Recall

Average list recall and recall of single words within lists both were measured and predicted. The set of predictors differs, of course, in the two situation. For prediction of average list recall, no more than eight predictors could be used simultaneously, as the ten word lists comprise a sample of only ten observations on which to base the prediction analysis. Use of nine or more predictors would result in the total correlation matrix of the predictors (the association measures) and the criterion (average recall) becoming singular. That is, recall would have a multiple correlation of 1.0 with the association measures, due only to the small number of observations in the sample.

Table 5

Words in Recall Sublists

Sublists	Words	Sublist	Words
"Buffers"	Arch Best Frail Gift Smile Trunk	First "Butterfly" list	Bee Bird Bug Butterfly Flower Flight Fly Insect Moth Wings
First "Cheese" list	Burger Cheddar Cheese Cottage Holes Mouse Pizza Sharp Spread Swiss	Second "Butterfly" list	Hedge Nature Outdoors Pollen Pretty Sky Summer Sun Tree Worm
Second "Cheese" list	Bread Cake Cow Cream Dairy Knife Milk Rat Smell Sour	"Unrelated Words" list	Bacon Cloth Light Mirror Net Chair Forget Story Subtract Theory

For prediction of recall of individual words the number of predictors is not a problem, as the prediction analysis can be taken over all 260 observations (26 words in each of ten lists).

Predictors of average list recall. For prediction of list recall, nine predictors were used, although as explained above only eight were used at any one time. The buffer words were not included in the computation of any of the associative predictors in this analysis. The average value of each of the six association matrices for each 20-item list was calculated: the conditional probabilities covariance and correlation, and the *MF* index, all three calculated both with and without the representational response. Table 6 lists the abbreviations used to identify these measures in the tabular presentation of results in Chapter 3. To avoid ambiguity, the representational response is identified by the letter "I" (for "identity response," an alternative term). The expressions "*+I*" and "*-I*" refer to measures computed with and without the representational response, respectively. The other three predictive measures, with their abbreviations, are as follows:

7. Inter-item associative strength. *IIAS*.
8. Index of total association. *ITA*.
9. Average cue number. *Av.Nc*.

Table 6

Abbreviations for Multivariate Statistics used in the Prediction of Average List Recall

Representational Response	Bivariate Measure	Abbreviation for Average of all Elements in Matrix
Not included	cond. prob. covariance	$Av(Cov^*, -I)$
	cond. prob. correlation	$Av(r^*, -I)$
	MF index	$Av(MF, -I)$
Included	cond. prob. covariance	$Av(Cov^*, +I)$
	cond. prob. correlation	$Av(r^*, +I)$
	MF index	$Av(MF, +I)$

Predictors of recall of individual words. For the prediction of single words, a deliberate confounding of levels of analysis was introduced. As well as the measures appropriate for individual words, the average and determinant of each matrix were also included as predictors. It was hoped that these measures would partially account for any systematic differences existing between lists, in terms of both the average level of "memorability" of the lists and the differences between lists in variability of ease of recall of individual words from the lists.

As in the case of the prediction of average list recall, predictive measures included those which were derived from the six association matrices for each word list: those from the conditional probabilities covariance and correlation, and the *MF* index, all three calculated both with and without the representational response.

The association matrices in this case were formed from all 26 stimulus words in each recall list. The buffers were included in the prediction analysis of single word recall in order to provide a test for the full extent of the effect of serial position on recall.

From each of the six bivariate measures, four multivariate statistics were calculated for each word in each word list. The abbreviations assigned to these predictors are listed in Table 7.

The first was the squared multiple correlation of each word w with the remaining 25 in its list (in the case of

Table 7

Abbreviations for Multivariate Statistics used in the Prediction of Single Word Recall

Representational Response	Bivariate Measure	Multivariate Measure			
		Squared Multiple Correlation	Determinant	Average Value of all Elements in Matrix	Mean Squared Correlation
Not Included	cond. prob. covariance	SMC(Cov*, - I)	Det(Cov*, - I)	Av(Cov*, - I)	MSq(Cov*, - I)
	cond. prob. correlation	SMC(r*, - I)	Det(r*, - I)	Av(r*, - I)	MSq(r*, - I)
	MF index	SMC(MF, - I)	Det(MF, - I)	Av(MF, - I)	MSq(MF, - I)
Included	cond. prob. convariance	SMC(Cov*, + I)	Det(Cov*, + I)	Av(Cov*, + I)	MSq(Cov*, + I)
	cond. prob. correlation	SMC(r*, + I)	Det(r*, + I)	Av(r*, + I)	MSq(r*, + I)
	MF index	SMC(MF, + I)	Det(MF, + I)	Av(MF, + I)	MSq(MF, + I)

the covariance measures, the squared multiple covariance was calculated). Second, the mean of the squared correlations or covariances of each word *w* with the remaining 25 was calculated.

The other two multivariate statistics are ones which describe lists of words, and were included to reflect differences between lists, as described above. These two were the average of all the elements in the association matrix and the determinant of the matrix; each one had the same value for all the words in any one list.

Six predictors not derived from bivariate association measures were also included; with their abbreviations, they are as follows.

25. Cue number. *Nc*.
26. Frequency of elicitation. *FE*.
27. Variance of the response distribution to *w*, calculated without the representational response. *Var, - I*.
28. Variance of the response distribution to *w*, calculated with the representational response. *Var, + I*.
29. Serial position index. *Ser. Pos*.
30. Thorndike-Lorge general word count frequency. *T-L freq*.

Apparatus

The items for the recall test were recorded on a Uher 4000 Report-L tape recorder with AC adapter. The items were read into the tape recorder by a colleague who was judged by informal consensus to have a clear, distinct, unaccented voice. He was trained to read the words at a speed of one per second.

The items were presented to the subjects from the tape recorder through individual speaker headsets. All subjects wore Hosiden Model DH025 stereo headsets connected to a 16 terminal parallel circuit impedance matching unit manufactured for use with multiple headsets in conjunction with a Tenco Model 1026 ten watt amplifier. In this way up to sixteen subjects at once could hear the word list through individual headsets. All sixteen headsets were left connected to the distributor (impedance matching unit) regardless of the number of subjects being tested, so as to minimize variation in sound level. The peak sound level at the headsets varied between 78 and 82 decibels across different words, but did not measurably vary across different headsets. Sound level was measured with a General Radio Model 1551-B sound level meter with headphone attachment.

Procedure

Word association tests. Procedure and instructions for the two administrations of the word association test were based closely on those used by Palermo and Jenkins (1964) in collecting the Minnesota norms. Emphasis was placed on speed, clarity of handwriting, and the experimenter's lack of concern for perfect spelling. Subjects were tested in their regular tutorial classes, which ranged in size from five to sixteen students. The second administration of the test was given two weeks after the first.

Recall test. The recall test was given one week after the second word association test, and was also carried out in the tutorial classes. Three different classes were given each word list; presentation speed was one word per second. At the beginning of the testing session, the experimenter informally explained that the purpose of the tape recorder and headsets was to control for background noise and to ensure that different classes receiving the same word list would all hear the same thing. Subjects were then told that they would hear a list containing between twenty and thirty words, and would be asked to recall them immediately afterwards. They were instructed to write the words down in whatever order they recalled them. Subjects were encouraged to guess at words they were unsure of, and were assured that they would have ample time to remember the list. At a signal from the experimenter, subjects put on their headsets and listened to the list of words. As soon as the list was over, the experimenter said "Start writing please," and said nothing else for the duration of the experiment. Subjects were allowed ten minutes for recall, but all subjects finished well before the ten minutes was up.

Collection of Data

All word association test booklets and recall tests were initially checked for legibility of writing and correctness of spelling. Following the lead of Palermo and Jenkins (1964), misspelled words were corrected but in no case was a word changed that seemed to be misspelled but which also spelled another word; for example, as a response to "green," "blew" was not changed to "blue." Illegible words were treated as omissions. Some problems arose in the scoring of homophonic responses in the orally presented recall test. Should "aunt," for example, be considered a correct recall of the stimulus "ant"? It was finally decided that homophonic responses might legitimately be considered to reflect different associative recall processes from those involved in recall of the actual stimulus word. Homophones were therefore treated as incorrect responses. As it happened, extremely few homophonic responses occurred.

After the initial screening, all data were transferred to IBM cards for tabulation and statistical analysis.

Chapter 3: Results

Bivariate Correlations Between Words

The response distributions to the 126 stimulus words from the first administration of the word association test were intercorrelated by means of all six bivariate indices described in Chapter 1. Six 126 x 126 correlation matrices resulted from this analysis; they are reporduced in full as the first six tables in Volume 2 of this study.

Prediction of Average Recall of Lists

Each 26 item list in the recall test generated six submatrices of bivariate association indices. Table 8 reports the mean of the elements of these matrices for each list, and also the inter-item associative strength, index of total association, average cue number, and average recall of each list. All the predictors are scaled so as to range from 0.0 to 1.0. The buffer words were not considered in measuring average list recall, and did not enter into the calculation of any of the predictors.

The intercorrelations between the predictors and the criterion are shown in Table 9. The level of correlation within the matrix is generally very high, and suggests that detailed analysis of prediction may be unreliable. The extremely high correlations among the predictors makes any detailed analysis of the relative contribution of the predictors unreliable and relatively uninterpretable. The *MF*

Table 8

Summary Measures of Intra-list Association, and Average List Recall[a,b]

Measure	List									
	1	2	3	4	5	6	7	8	9	10
Av(Cov, - I)*	.02	.01	.01	.01	.01	.00	.00	.01	.01	.00
Av(r, - I)*	.10	.09	.06	.05	.08	.04	.03	.08	.05	.02
Av(MF, - I)	.06	.06	.04	.03	.06	.04	.03	.07	.04	.02
Av(Cov, + I)*	.01	.01	.01	.01	.01	.00	.00	.01	.01	.00
Av(r, + I)*	.04	.04	.02	.03	.03	.01	.01	.03	.02	.00
Av(MF, + I)	.05	.05	.03	.02	.04	.03	.02	.05	.03	.01
IIAS	.30	.31	.14	.13	.28	.12	.10	.25	.19	.02
ITA	.73	.74	.68	.51	.77	.67	.57	.75	.59	.55
Av.Nc.	.10	.13	.05	.03	.13	.05	.03	.16	.11	.03
Average Recall	9.30	8.99	8.23	8.08	8.68	8.90	6.86	9.74	8.57	6.87

[a] Total recall lists each included 26 items, but computation of association and recall measures excluded the 6 buffer words.

[b] Each association measure was calculated from 283 observations. The recall scorew were calculated from an average of 28 observations per list, with a range of 26 to 30.

Table 9

Intercorrelations of Summary Association Measures and Recall, Across Ten Recall Lists[a]

	Measure	1	2	3	4	5	6	7	8	9	10
1	$Av(Cov*, - I)$	1.00									
2	$Av(r*, - I)$	.86	1.00								
3	$Av(MF, - I)$	.66	.95	1.00							
4	$Av(Cov*, + I)$	.90	.98	.88	1.00						
5	$Av(r*, + I)$	.85	.99	.92	.99	1.00					
6	$Av(MF, + I)$	.72	.97	.99	.93	.96	1.00				
7	*IIAS*	.78	.97	.93	.98	.99	.97	1.00			
8	*ITA*	.53	.81	.90	.74	.78	.88	.79	1.00		
9	$Av.Nc.$	.40	.81	.91	.74	.80	.90	.85	.78	1.00	
10	*Average Recall*	.59	.83	.88	.76	.78	.86	.78	.75	.79	1.00

[a] Ten observations; based on data from Table 8.

index without the representational response is most highly correlated with recall, followed closely by the *MF* index with the representational response. The *MF* indices and the conditional probabilities correlations are all correlated higher than +.90 with each other. Every measure in the predictor set has a multiple correlation higher than .9999 with the rest of the predictors.

Use of the determinants to estimate the homogeneity of association within each word list was unsuccessful or irrelevant in this analysis. With one exception (the conditional probabilities covariance without the representational response), the determinants and the averages of each association measure were very highly correlated. In every case, the determinants were correlated slightly less with recall than were the averages, and added no new information in regression analyses to that already provided by the averages. They were not used further in this analysis.

To provide further information concerning the extent of similarity of the association measures, the ten word lists were ranked according to their overall level of association, as measured by each of the nine association measures listed in Table 8. In general, the measures ranked the word lists quite consistently. Table 10 reports the Kendall coefficient of concordance, *W*, for several subsets of measures. The best agreement is among the three average bivariate measures (the conditional probabilities covariance and correlation, and the

Table 10

Kendall Coefficient of Concordance (W) for selected Subsets of Predictors of Average List Recall[a]

Subset	W	Average ρ [b]
Averages of the three bivariate measures with the representational response	.98	+.97
Averages of the three bivariate measures without the representational response	.88	+.81
Averages of all six bivariate measures	.91	+.89
IIAS, ITA, Av.Nc.	.88	+.82
All nine predictors	.86	+.84

[a] Vased on rankings of ten word lists.

[b] Average $\rho + \dfrac{fW - 1}{f - 1}$

where f is the number of predictors (Siegal, 1956, p.229).

MF index), with the representational response. The average Spearman rank order correlation (ρ) among them is +.97. The three average bivariate measures without the representational response, conversely, show the worst agreement, with an average ρ of +.81. The average ρ for the whole set of nine predictors is +.84.

The extent of agreement among the predictors can be roughly estimated from Table 8. Lists 1, 2, and 8 are higher than the rest on most of the association measures, and are the three lists with the highest average recall. Lists 4, 7, and 10 are lowest on most of the association measures, and are the worst recalled lists.

Use of the best eight predictors in a regression analysis resulted in a multiple correlation of the association measures with recall of +.99. When corrected for shrinkage (Guilford, 1956), the multiple correlation dropped to +.96. Although almost all of the recall variance was accounted for, it was not possible to develop a prediction battery. The predictors are so highly correlated that any attempt to form useful predictor subsets would require, at the least, far more observations than are available here.

Factor Analysis of Summary List Measures

The correlation matrix shown in Table 9 was subjected to a principal component analysis with varimax rotation. The first six eigenvalues of the correlation matrix were 8.61; 0.83; 0.26; 0.25; 0.04; and 0.01. The remaining four were computationally zero to three decimal places. Four varimax factors accounted for 99.3% of the total variance. Their loadings and the percent variance accounted for by each are shown in Table 11; the highest loading from each variable is underlined. Factor I has the highest loadings from all four conditional probabilities measures and from the *IIAS*. Factor II has high loadings from both of the *MF* measures and from the average cue number. The only high loadings on Factors III and IV are from *ITA* and average recall, respectively. Again, the high level of correlations in Table 9 reduces the reliability of the measurements to a great extent. The first principal component accounted for 86% of the total variance; only the first eigenvalue exceeded 1.0. The separation of the variance into four distinct factors, while very interesting in terms of the divisions it creates, may well be spurious; in any case, it cannot be accorded very much confidence. It is probably best to conclude that, for this sample, a single factor solution is appropriate.

Prediction of Recall of Individual Words

Each of the thirty measures for prediction of recall was computed for all 260 observations. The complete data matrix appears as the seventh through fourteenth tables in Volume 2 of this study.

The intercorrelations among the predictors and the criterion are shown in Table 10. For comparison purposes, two measures of the recall criterion are included--the raw pro-

Table 11

Tentative Loadings of List Variables on Varimax-Rotated Principal Components[a]

	Variable	Factor			
		I	II	III	IV
1	Av(Cov*, - I)	.96	.07	.17	.20
2	Av(r*, - I)	.72	.49	.36	.32
3	Av(MF, - I)	.46	.61	.49	.40
4	Av(Cov*, + I)	.81	.45	.29	.24
5	Av(r*, + I)	.74	.52	.32	.24
6	Av(MF, + I)	.56	.61	.44	.34
7	IIAS	.67	.60	.33	.24
8	ITA	.32	.41	.81	.27
9	Av.Nc.	.22	.86	.33	.32
10	Average Recall	.37	.41	.31	.77
	% total variance accounted for	39.20	28.96	17.48	13.66

[a] Ten observations; based on data from Table 9.

portion of subjects who recalled each word, and the proportion as a deviation score from the list mean. The two forms of recall correlate very highly (+.98) and suggest, therefore, that there is much greater variance of recall between words than between lists.

The summary list measures, the means and determinants, are correlated only slightly with raw recall, and not at all with recall with the list means extracted. All the other associational variables are correlated slightly higher with raw recall than with recall without list means, even though the measures were derived from the association data for each list separately. The implication is that the predictor variables successfully reflect differences between lists, in addition to their function of predicting recall within lists.

The level of correlations with the criterion are not nearly so high in the prediction of single word recall as they are in the prediction of average list recall. The squared multiple correlations, with the representational response, have the highest simple correlations with raw recall; of these, the squared multiple conditional probabilities correlation is fractionally higher than the rest. The representational response seems necessary in predicting single word recall; the squared multiple correlations without the representational response are considerably less correlated with recall than are the same measures with the representational response.

Table 12

Intercorrelations among Measures used in the Analysis of Single Word Recall^a

(Decimals omitted)

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	_SMC(Cov*,- I)_	100															
2	_Det(Cov*,- I)_	-06	100														
3	_Av(Cov*,- I)_	21	-34	100													
4	_MSq(Cov*,- I)_	93	-06	22	100												
5	_SMC(r*,- I)_	69	-12	27	64	100											
6	_Det(r*,- I)_	-13	14	-66	-12	-21	100										
7	_Av(r*,- I)_	15	-47	85	13	29	-68	100									
8	_MSq(r*,- I)_	72	-12	28	71	97	-22	29	100								
9	_SMC(MF,- I)_	79	-13	25	78	93	-20	28	94	100							
10	_Det(MF,- I)_	-15	46	-84	-13	-29	82	-97	-29	-28	100						
11	_Av(MF,- I)_	09	-48	63	06	27	-58	94	25	26	-89	100					
12	_MSq(MF,- I)_	68	-15	25	65	93	-23	33	94	97	-32	32	100				
13	_SMC(Cov*,+ I)_	76	-10	23	67	64	-17	21	64	72	-22	17	67	100			
14	_Det(Cov*,+ I)_	-11	67	-65	-09	-25	65	-87	-24	-25	89	-89	-29	-18	100		
15	_Av(Cov*,+ I)_	16	-43	89	15	30	-72	99	30	28	-98	89	33	22	-83	100	
16	_MSq(Cov*,+ I)_	66	-09	24	62	58	-17	21	59	65	-21	15	60	96	-17	22	100

Table 12 (Continued)

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
17	SMC(r*,+ I)	56	-10	23	50	60	-18	23	59	65	-23	20	64	96	-20	24	96
18	Det(r*,+ I)	-18	39	-91	-18	-27	89	-88	-28	-26	95	173	-28	-22	80	-92	-23
19	Av(r*,+ I)	15	-43	84	12	30	-71	99	30	28	-98	93	33	21	-85	99	21
20	MSq(r*,+ I)	55	-10	23	50	61	-18	23	59	65	-24	20	64	94	-20	24	98
21	SMC(MF,+ I)	56	-13	24	46	67	-20	28	66	72	-28	26	74	93	-24	28	92
22	Det(MF,+ I)	-15	44	-84	-13	-29	81	-97	-29	-28	99	-89	-33	-21	87	-98	-21
23	Av(MF,+ I)	11	-47	70	07	28	-63	97	27	27	-93	99	33	19	-89	93	17
24	MSq(MF,+ I)	54	-13	24	47	71	-21	29	70	75	-29	27	78	91	-25	30	91
25	Nc	05	-14	12	00	37	-15	26	31	36	-25	29	45	56	-24	24	62
26	FE	01	-08	14	-03	26	-13	18	18	21	-18	18	26	59	-16	18	70
27	Var,- I	71	06	05	65	29	00	-04	34	39	03	-09	29	46	09	-02	38
28	Var,+ I	71	06	04	65	29	00	-04	33	39	03	-09	29	46	09	-02	38
29	Ser. Pos.	-07	00	00	-03	-19	00	00	-22	-21	00	00	-24	-13	00	00	-10
30	T-L freq.	-24	00	-07	-26	-15	06	-05	-19	-21	06	-03	-19	-20	03	-06	-19
31	Raw Recall	29	-11	10	28	28	-11	15	28	32	-15	16	31	51	-15	14	51
32	Recall-List Mean	28	00	00	28	25	00	00	25	28	00	00	26	49	00	00	49

Table 12 (Continued)

	Variable	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
17	_SMC(r*,+ I)_	100															
18	_Det(r*,+ I)_	-23	100														
19	_Av(r*+ I)_	23	-89	100													
20	_MSq(r*+ I)_	99	-23	24	100												
21	_SMC(MF,+ I)_	97	-25	28	96	100											
22	_Det(MF,+ I)_	-23	94	-98	-24	-28	100										
23	_Av(MF,+ I)_	21	-79	96	21	27	-93	100									
24	_MSq(MF,+ I)_	96	-26	30	96	99	-29	29	100								
25	_Nc_	71	-18	26	73	75	-25	29	77	100							
26	_FE_	76	-16	18	79	73	-18	18	74	88	100						
27	_Var, - I_	28	-01	-04	27	24	03	-08	22	-15	-17	100					
28	_Var, + I_	28	-01	-04	27	24	03	-08	22	-15	-17	99	100				
29	_Ser. Pos._	-14	00	00	-13	-18	00	00	-18	-16	-09	07	08	100			
30	_T-L freq._	-17	07	-06	-15	-17	06	-04	-16	12	07	-21	-21	-13	100		
31	_Raw Recall_	53	-13	15	52	49	-15	16	48	37	40	11	11	30	-21	100	
32	_Recall - List Mean_	50	00	00	50	46	00	00	45	32	38	12	13	30	-21	98	100

[a] Based on 260 observations.

A regression of raw recall on the complete predictor set accounted for 47% of the recall variance. Due to almost complete redundancy of the predictor set (almost every predictor is highly correlated with several others), very few predictors are necessary to account for most of the predictable variance. The squared multiple conditional probabilities correlation with the representational response, combined with the serial position index, account for 41.7% of the recall variance by themselves. In the complete regression, they were the only predictors which uniquely accounted for more than one percent of the recall variance. The contributions of the rest of the predictors to the variance accounted for are of the sort that would be expected due to sampling error: they are small, unordered, and much less than the contributions of either of the first two predictors. A parsimonious interpretation of the prediction analysis requires that only the variance accounted for by these first two predictors be considered meaningfully predictive.

Neither the averages nor the determinants of the bivariate measures matrices contributed significantly to the regression after the first two predictors were entered. This finding is somewhat surprising, particularly with respect to the determinants; the implication is that either the level of heterogeneity of association in different lists does not influence recall of individual words from the lists, or that the determinants are not successfully tapping the form of heterogeneity involved.

Several subsets of the full predictor set were used in separate regression analyses. The percentages of variance accounted for by each subset is shown in Table 13. Each regression was performed both with and without the two non-associative measures, serial position and word frequency, included. The amount of variance added to the regression equations by these two predictors is quite similar across subsets. The predictive power of some of the subsets is as high as what was suggested to be the reliable predictive power of the entire set. Any of the three squared multiple correlations or mean squared correlations with the representational response, with the addition of the non-associative measures, would appear to comprise a fairly good, small, predictive battery.

For the subsets containing only one predictor (each of the squared multiple correlations, the mean squared correlations, *Nc*, and *FE*) the regression was also performed with the addition of each of the two non-associative measures separately. The increment in the recall variance accounted for due to adding the word frequency measure is quite small in all cases except those involving the *Nc* and *FE*. All the other single predictors seem to incorporate most of the information contained in the word frequency measure; it is curious that *Nc* and *FE* alone are lacking in this regard.

Table 13

Variance of Single Word Recall Accounted for by Selected Subsets of Predictors[a]

Subset	f[b]	Percent Variance Accounted for			
		Without *Ser.Pos.* or *T-L freq.*	With *T-L freq.*	With *Ser.Pos.*	With *Ser.Pos.* and *T-L freq.*
No Associative Predictors	0	-----	4.41	8.90	11.87
All Squared Multiple Correlations	6	28.45			43.21
All Determinants	6	2.97			14.72
All Means	6	3.45			15.17
All Means and Determinants	12	4.41			16.01
All Mean Squared Correlations	6	27.77			41.81
Across Factors	7	-----			38.03
Nc	1	13.45	20.03	26.55	31.25
FE	1	16.22	21.92	27.53	31.45
SMC(Cov, - I)*	1	8.32	10.43	18.59	19.54
SMC(r, - I)*	1	8.05	10.93	20.99	22.24
SMC(MF, - I)	1	9.99	12.17	23.91	24.57
SMC(Cov, + I)*	1	25.55	26.76	38.86	39.13
SMC(r, + I)*	1	27.73	29.27	41.69	42.13
SMC(MF, + I)	1	24.36	25.94	40.00	40.36
MSq(Cov, - I)*	1	8.03	10.02	17.51	18.46
MSq(r, - I)*	1	8.03	10.57	21.56	22.47
MSq(MF, - I)	1	9.64	11.96	24.52	25.18
MSq(Cov, + I)*	1	25.75	27.12	38.03	38.44
MSq(r, + I)*	1	26.90	28.65	40.35	40.95
MSq(MF, + I)	1	23.46	25.23	38.98	39.45

---

[a] Based on 260 observations in each case.

[b] f = number of predictors.

Additional evidence is provided in Table 13 for the failure of the determinants to provide unique predictive information. The subset composed of the means and the determinants accounted for almost no more recall variance than did the subset composed of the means alone; in both cases, the variance accounted for was small.

The "Across Factors" subset was composed of the variables with the highest loadings on each of the varimax factors, to be discussed in the next section.

The six squared multiple correlations and mean squared correlations, with the representational response, were better predictors than any of the other measures. The squared multiple conditional probabilities correlation was slightly more effective than the others, but the difference was small.

Two additional regression analyses were performed on the combined word lists; both used the complete set of predictors. In the first, all the predictors derived from bivariate association measures were rescaled to the metric of the original bivariate indices. The squared multiple correlations and the mean squared correlations were transformed to the multiple correlations and root mean squared correlations, respectively, and the determinants were transformed to their $2k^{th}$ roots. The average of the elements in the matrix was not changed. When a complete regression was performed using these rescaled predictors, the variance accounted for increased 1.9% over that accounted for with the same predictors before rescaling. The multiple conditional probabilities correlation with the representational response and the serial position index were, again, the only predictors uniquely accounting for more than one percent of the recall variance. The two of them alone accounted for 43% of the recall variance, 1.3% more than they accounted for before rescaling.

In the second regression, the six words with the lowest scale reliabilities were removed from the prediction analysis. A complete regression was performed with the reduced data set; the total variance accounted for was just over 50%, 3% higher than in the regression with all data included. The same two variables, the squared multiple conditional probabilities correlation with the representational response, and the serial position index, were again the only predictors which uniquely accounted for more than one percent of the criterion variance. Together, they accounted for 43.59% of the criterion variance, 1.9% more than they accounted for in the regression with all the data included.

Both of the additional regressions accounted for more criterion variance than did the initial regression with the full data set and the predictors used before rescaling. For the same reasons as were advanced in the description of the initial regression, only the variance accounted for by the first two predictors to enter the regressions can be considered meaningful. The "meaningful" increment in the variance accounted for by the two additional regressions is rather small--1.3% and 1.9% respectively. Appropriate significance tests for the differences are not readily available, and it is thus rather difficult to conclude whether or not the increments are significant. They are, in any case, small.

In all the correlation analyses, there were six predictors which were more powerful than the rest; they were the three squared multiple correlations and the three mean squared correlations, all with the representational response. There were few predictive differences among these six variables, although the squared multiple conditional probabilities correlation was significantly more correlated with recall than was the squared multiple *MF* index ($p<.01$).

To provide a more detailed comparison, the correlations with single word recall of each of the six predictors was examined in each word list separately. Serial position is significantly related to recall but not, of course, to the association measures; the efficacy of the measures as predictors of recall is reduced by the serial position effect. To obtain correlations between recall and the association measures independent of the effect of serial position, serial position was partialled out of each correlation.

Table 14 shows the correlations, with serial position partialled in each case. The highest partial correlations in each list are underlined. The squared multiple correlations of both the conditional probabilities correlation

Table 14

Partial Correlation of Six Associative Measures with Single Word Recall in Each of Ten Words Lists, Serial Position Partialled[a]

List	Measure						
	SMC (*Cov**, + *I*)	*SMC* (*r**, + *I*)	*SMC* (*MF*, + *I*)	*MSq* (*Cov**, + *I*)	*MSq* (*r**, + *I*)	*MSq* (*MF*, + *I*)	Root Mean Square (Average *r* for each list)
1	.71	.73	.73	.75	.74	.74	.73
2	.58	.60	.60	.57	.58	.60	.59
3	.74	.79	.79	.78	.79	.79	.78
4	.71	.73	.74	.73	.72	.74	.73
5	.51	.52	.42	.48	.48	.40	.47
6	.45	.46	.47	.44	.44	.44	.45
7	.60	.61	.52	.60	.60	.52	.58
8	.50	.49	.42	.45	.44	.37	.45
9	.46	.46	.50	.44	.44	.50	.47
10	.41	.41	.38	.41	.41	.37	.40
Root Mean Square (Average *r* for each measure)	.58	.59	.57	.58	.58	.57	

---

[a] Each partial based on 26 observations.

and the *MF* index had the highest correlation with recall in five out of the ten word lists, counting ties. The squared multiple conditional probabilities correlation has the highest average correlation (actually a root mean squared correlation) with recall over the ten word lists.

To determine whether any of the observed differences among the measures were significant, an analysis of variance was performed on the partials. The partial correlations were first normalized by converting them to Fisher z scores.

The standard error of a Fisher z score is $1/\sqrt{n-3}$, where n is the number of observations on which the z was calculated. One degree of freedom is lost with each variable partialled, so with one variable partialled, in the present case, the standard error of each z score is $1/\sqrt{n-4}$. The harmonic mean number of subjects tested on each recall list is 28.2; the average standard error in the cells of the data matrix is thus $1/\sqrt{24.2}$, or 0.203. This quantity is appropriate for use as an *a priori* estimate of the error residual in the analysis of variance; its use makes possible a two-way analysis of variance, which can provide an assessment of the effect of lists, and of the interaction between lists and measures.

Since the word lists were constructed according to definite experimental criteria, it should be possible to treat lists as a fixed variable. However, it is uncertain how successful the experimental criteria were in generating lists which were adequately differentiated in terms of effects on individual words. The differentiation was not, in any event, successfully related to differences in single word recall in the ten word lists combined. The more stringent test resulting from treating lists as a random variable was therefore employed.

The analysis of variance is summarized in Table 15. The list effect is highly significant; recall seems considerably more predictable in some lists than in others. There is no main effect for measures; the differences in predictive power between measures were not even close to significant. There is a significant lists x measures interaction, indicating that different measures predicted recall better in different lists. The size of the interaction is fairly small, however, and would not be significant were not infinite degrees of freedom associated with the *a priori* estimate of the error residual.

Of greater interest than the interaction are the findings of the two main effects. The *F* for measures is extremely low, below 1.0. No confidence at all can be attached to any predictive differences among the measures in this particular analysis.

The highly significant *F* for lists, on the other hand, suggests that it is worthwhile to examine the characteristics of the lists themselves to determine what properties of the word lists are related to increased predictive ability of the associative measures. A detailed analysis of the relevant properties of the lists would require more observations

Table 15

Analysis of Variance of the Predictive Power of Six Associative Measures on Single Word Recall from Ten Recall Lists[a]

Source	*df*	*MS*	*F*
Lists	9	0.3087	7.72[b]
Measures	5	0.0016	1.0
Lists x Measures Interaction	45	0.0016	1.79[b]
Error		0.0413	

[a] Based on data from Table 14, after transformation to Fisher z scores for normalization.

[b] p .01

than are available in the present study. In Table 14, the root mean squared correlations for each list, averaged across predictors, are presented in the final column. The most salient characteristic of the four most "predictable" lists (lists 1, 2, 3, and 4) is that they all contain the words from the first "cheese" sublist!

To test the stability of the relations between the predictors, a complete regression was performed on each word list separately. The amount of recall variance accounted for in each list ranged from 55% to 96%; in every list, the percentage of variance accounted for was higher than in the regression on the concatenated word lists. The regression equations for the different lists are quite dissimilar, however. In the concatenated word lists, only the squared multiple conditional probabilities correlation with the representational response, and the serial position index, uniquely accounted for more than one percent of the recall variance. In each of the regressions of single word recall in separate word lists, almost half of the predictors uniquely accounted for more than one percent. Furthermore, every variable excepting *FE* was one of the major predictors in at least one regression equation. The dissimilarity of the regression equations may be related to the differential predictability of recall in the separate lists. Conversely, it may rather illustrate the dangers of sampling only one or two lists of words, and expecting the resulting prediction equations to be universal.

Factor Analysis of Single Word Measures

The thirty predictive and descriptive measures for each word were factored with a principal component analysis and subsequent varimax rotation. The factoring was done on the correlation matrix of Table 12, based on the full words x lists sample of 260 observations. The first 24 eigenvalues of the correlation matrix were 12.904; 7.428; 3.303; 1.613; 1.036; 0.996; 0.769; 0.599; 0.421; 0.360; 0.158; 0.133; 0.074; 0.073; 0.042; 0.030; 0.025; 0.014; 0.007; 0.006; 0.004; 0.002; 0.001; and 0.001. The remaining 6 eigenvalues were computationally zero to three decimal places.

The varimax-rotated factor loadings are shown in Table 16; the highest loading from each variable is underlined to highlight the patterns of "affiliations" between variables and factors. All rotated factors were retained if they had varimax roots (i.e., sums of squares of factor loadings; analogous to the eigenvalues of the correlation matrix) greater than 1.0. The decision on the number of factors to keep thus required successive rotations in order to examine a two factor solution, a three factor solution, etc. Seven factors were kept by this procedure; together, they accounted for 93% of the total variance.

In most cases the separation between loadings is very clear. Factor I has highest loadings from the means and determinants of all the association matrices, with but one

Table 16

Loadings of Single Word Measures on Varimax-Rotated Principal Components[a]

Measure	Factor						
	I	II	III	IV	V	VI	VII
1 *SMC(Cov*, - I)*	-.06	.26	-.61	-.67	-.07	-.08	-.01
2 *Det(Cov*, - I)*	.39	-.05	.04	-.04	.89	-.02	.00
3 *Av(Cov*, - I)*	-.85	.09	-.08	-.07	-.02	-.02	-.01
4 *MSq(Cov*, - I)*	-.05	.18	-.63	-.61	-.09	-.10	.04
5 *SMC(r*, - I)*	-.17	.33	-.89	-.11	.01	.00	-.04
6 *Det(r*, - I)*	.81	-.07	.06	.00	-.25	.00	.01
7 *Av(r*, - I)*	-.97	.11	-.11	.00	-.11	-.01	.00
8 *MSq(r*, - I)*	-.17	.29	-.90	-.16	.00	-.04	-.07
9 *SMC(MF, - I)*	-.16	.36	-.87	-.24	-.03	-.06	-.07
10 *Det(MF, - I)*	.98	-.10	.10	-.01	.07	.01	.00
11 *Av(MF, - I)*	-.89	.10	-.11	.05	-.15	-.01	.01
12 *MSq(MF, - I)*	-.20	.38	-.86	-.11	-.01	-.06	-.10
13 *SMC(Cov*, + I)*	-.09	.80	-.37	-.39	-.05	-.09	-.03
14 *Det(Cov*, + I)*	.86	-.09	.09	-.06	.38	.01	-.01
15 *Av(Cov*, + I)*	-.97	.11	-.11	-.01	-.05	-.01	.00
16 *MSq(Cov*, + I)*	-.09	.87	-.29	-.31	-.05	-.08	.00
17 *SMC(r*, + I)*	-.11	.91	-.30	-.19	-.03	-.07	-.03
18 *Det(r*, + I)*	.93	-.09	.08	.03	.00	.01	.01
19 *Av(r*, + I)*	-.97	.11	-.11	.00	-.05	-.01	.00
20 *MSq(r*, + I)*	-.11	.92	-.30	-.17	-.02	-.06	-.01
21 *SMC(MF, + I)*	-.15	.88	-.39	-.12	-.02	-.08	-.07
22 *Det(MF, + I)*	.98	-.11	.10	.00	.04	.01	.00
23 *Av(MF, + I)*	-.93	.10	-.11	.04	-.12	-.01	.01
24 *MSq(MF, + I)*	-.16	.86	-.44	-.09	.00	-.06	-.05
25 *Nc*	-.15	.84	-.10	.25	.01	.17	-.04
26 *FE*	-.09	.93	.08	.24	.02	.09	.01
27 *Var, - I*	.05	.07	-.17	-.96	.05	-.05	.03
28 *Var, + I*	.05	.07	-.17	-.96	.05	-.05	.04
29 *Ser. Pos.*	-.02	-.09	.15	-.07	.00	-.07	.98
30 *T-L freq.*	.03	-.05	.12	.13	-.01	.98	-.07
% Total variance accounted for	32.73	22.86	16.10	10.86	3.63	3.53	3.33

[a] 260 observations; based on data from Table 12.

exception.

Factor II strongly represents the squared multiple correlations and mean squared correlations calculated with the representational response, and also *Nc* and *FE*. Factor II, therefore, contains all the associative measures most useful in prediction of recall.

Factors III and IV split several loadings, and provide the only cases of poor separation. Factor III has the highest loadings from the squared multiple correlations (with one exception) and from the mean squared correlations of the bivariate measures, all calculated without the representational response. Factor IV seems almost specific to the variances of the response distributions, but also has the marginally highest loading from the squared multiple conditional probabilities covariance without the representational response.

Factors V, VI, and VII appear to be specifics to the determinant of a covariance matrix, the word frequency, and serial position, respectively.

Several features of the factor loading matrix should be emphasized. First, except for measures derived from the covariance without the representational response, all the variables fit very neatly and solidly into their respective factors. Second, it is noteworthy that among the associative measures derived from bivariate indices, the major split was not between measures themselves, but between, on the one hand, presence or absence of the representational response, and on the other, measures relating to individual words (multiples and mean squares) and measures relating to lists (means and determinants). These findings are consistent with the overall patterns of correlations observed in the correlation matrix. The means and determinants correlated very poorly with the criterion but very highly with each other, the multiples and mean squares with the representational response correlated best with the criterion and highly with each other, and the multiples and mean squares without the representational response correlated moderately with the criterion and highly with each other. All three groups of variables had their highest correlations with other variables within their group.

The inclusion of recall in the data matrix for factoring does not change the relative loading patterns of the other variables appreciably. With seven factors, recall has as its highest loading a moderate +.66 on Factor VII (serial position). In a nine factor solution, recall loads +.81 on Factor IX and only +.27 on Factor VII. Adding the other recall measure (recall with list means extracted) to the data matrix makes little change except to double the recall variance. Some reordering of the factors takes place, and both recall measures load +.91 on the new Factor IV. Again, the composition of the other factors is not appreciably altered. In all cases, the next highest loading of recall is on Factor II, the factor which also contains the good associative predictors of recall.

Reliability of Word Associations

Six reliability and variance measures are reported in Table 17 for each of the 126 items on the word assO-ciation test. They are the scale reliability coefficient with the Spearman-Brown correction, the test-retest reliability coefficient, the proportion of identical associations calculated from the actual responses of each subject on the two test administrations, the proportion of identical associations calculated from random pairings of first and second administration tests, the variance of the response distributions on the first administration of the test, and the variance on the second administration. The correlations between the measures are reported in Table 18.

In this context, the scale reliability coefficient indicates the stability of the response distributions across samples of subjects. The test-retest reliability coefficient indicates the stability of the response distributions across two test administrations with the same subjects. It should be clear that both these coefficients pertain to group data. All of the reliability measures are of course calculated from group data and are appropriate for the assessment of reliability within the pooled sample; however, the proportion of identical associations from real subjects is somewhat closer than the rest to a measure of subject-based reliability.

The scale reliability, with the Spearman-Brown correc-

Table 17

Reliability and Variance Measures for Individual Words[a,b]

Word		Measure					
		r_{scale}	r_{trt}	Prop. *IAs*[c] (real)	Prop. *IAs*[d] (random)	*Var*, - *I* (first test)	*Var*, - *I* (second test)
1	Arch	.77	.83	.28	.02	.02	.03
2	Best	.95	.99	.40	.18	.15	.23
3	Frail	.85	.92	.27	.05	.03	.05
4	Gift	.99	.99	.47	.27	.25	.25
5	Smile	.96	.97	.30	.10	.10	.12
6	Trunk	.99	.98	.38	.20	.23	.17
7	Burger	.99	.99	.47	.25	.23	.31
8	Cheddar	.99	.99	.94	.92	.92	.92
9	Cheese	.93	.88	.20	.04	.05	.05
10	Cottage	.97	.89	.37	.14	.12	.17
11	Holes	.89	.84	.16	.01	.02	.02
12	Mouse	.96	.94	.31	.09	.09	.09
13	Pizza	.97	.97	.39	.10	.12	.13
14	Sharp	.95	.95	.29	.10	.11	.10
15	Spread	.97	.97	.36	.14	.10	.15
16	Swiss	.99	.99	.58	.50	.40	.57
17	Bread	.99	.99	.44	.32	.34	.35
18	Cake	.95	.96	.25	.07	.07	.07
19	Cow	.98	.97	.31	.22	.18	.17
20	Cream	.97	.78	.20	.06	.09	.10
21	Dairy	.95	.97	.37	.15	.16	.15
22	Knife	.97	.95	.39	.16	.15	.16
23	Milk	.97	.95	.26	.09	.09	.09
24	Rat	.98	.97	.31	.11	.13	.12
25	Smell	.89	.89	.24	.04	.04	.05
26	Sour	.99	.98	.52	.22	.22	.27
27	Bee	.97	.96	.33	.08	.11	.09
28	Bird	.97	.94	.28	.09	.11	.15

Table 17(Continued)

	Word	Measure					
		r_{scale}	r_{trt}	Prop.*IAs*[c] (real)	Prop.*IAs*[d] (random)	*Var, - I* (first test)	*Var, - I* (second test)
29	Bug	.96	.94	.26	.04	.07	.06
30	Butterfly	.94	.90	.23	.05	.05	.06
31	Flower	.95	.91	.21	.03	.04	.03
32	Flight	.97	.97	.34	.13	.12	.12
33	Fly	.92	.92	.18	.03	.03	.03
34	Insect	.97	.97	.31	.16	.14	.24
35	Moth	.95	.98	.34	.12	.11	.11
36	Wings	.99	.90	.40	.26	.25	.29
37	Hedge	.96	.92	.38	.09	.07	.07
38	Nature	.88	.81	.24	.04	.02	.03
39	Outdoors	.98	.98	.31	.14	.14	.15
40	Pollen	.96	.94	.29	.07	.10	.12
41	Pretty	.96	.96	.34	.09	.10	.10
42	Sky	.98	.98	.42	.26	.24	.36
43	Summer	.94	.93	.22	.07	.09	.07
44	Sun	.97	.88	.27	.11	.10	.09
45	Tree	.94	.91	.17	.05	.05	.06
46	Worm	.94	.91	.27	.05	.04	.04
47	Ant	.97	.97	.36	.12	.11	.15
48	Beautiful	.93	.96	.24	.05	.06	.06
49	Beauty	.94	.90	.29	.06	.05	.07
50	Blue	.98	.98	.31	.20	.20	.18
51	Caterpillar	.95	.94	.27	.10	.06	.07
52	Cocoon	.97	.97	.36	.12	.13	.11
53	Delicate	.92	.80	.26	.05	.07	.07
54	Monarch	.99	.97	.47	.25	.24	.29
55	Moon	.96	.91	.24	.07	.06	.06
56	Soft	.98	.98	.32	.15	.18	.18

Table 17 (Continued)

	Word	Measure					
		r_{scale}	r_{trt}	Prop. *IAs*^{c} (real)	Prop. *IAs*^{d} (random)	*Var*, - *I* (first test)	*Var*, - *I* (second test)
57	Spring	.95	.91	.25	.11	.10	.10
58	Bacon	.99	.96	.37	.17	.23	.16
59	Cloth	.92	.91	.20	.02	.04	.03
60	Light	.97	.92	.25	.10	.13	.09
61	Mirror	.98	.93	.35	.12	.12	.10
62	Net	.99	.98	.45	.27	.30	.27
63	Chair	.99	.97	.39	.18	.18	.16
64	Forget	.99	.99	.52	.32	.31	.31
65	Story	.98	.98	.40	.17	.18	.20
66	Subtract	.99	.99	.74	.62	.62	.60
67	Theory	.84	.88	.27	.03	.03	.04
68	Apple	.97	.96	.32	.12	.12	.13
69	Book	.95	.94	.20	.09	.06	.12
70	Class	.92	.92	.18	.06	.04	.04
71	College	.97	.96	.26	.09	.07	.09
72	Course	.84	.83	.20	.06	.03	.03
73	Essay	.94	.94	.28	.09	.07	.11
74	Exam	.98	.93	.28	.12	.10	.11
75	Fail	.99	.99	.46	.31	.31	.33
76	First	.97	.99	.50	.31	.27	.30
77	Grade	.93	.90	.24	.06	.06	.07
78	Learn	.98	.95	.25	.12	.09	.09
79	Lecture	.84	.88	.22	.04	.04	.05
80	Mark	.95	.80	.19	.04	.07	.03
81	Number	.96	.94	.27	.11	.07	.10
82	Pass	.99	.99	.42	.26	.24	.33
83	Research	.94	.87	.25	.04	.04	.05
84	Scholar	.95	.96	.28	.10	.06	.10

Table 17 (Continued)

Word	Measure					
	r_{scale}	r_{trt}	Prop.*IA*s^c (real)	Prop.*IA*s^d (random)	*Var,* - *I* (first test)	*Var,* - *I* (second test)
85 School	.88	.85	.19	.02	.03	.04
86 Student	.95	.96	.25	.09	.10	.08
87 Study	.95	.90	.17	.06	.07	.06
88 Teacher	.96	.90	.23	.08	.11	.06
89 Text	.99	.99	.79	.94	.68	.80
90 Thesis	.97	.97	.37	.08	.10	.11
91 Tutor	.97	.91	.23	.10	.09	.07
92 University	.95	.96	.27	.08	.07	.12
93 Write	.95	.95	.25	.05	.08	.07
94 Animal	.95	.90	.19	.08	.08	.06
95 Butter	.98	.90	.27	.13	.15	.12
96 Crackers	.98	.97	.44	.17	.18	.21
97 Eat	.98	.98	.28	.14	.14	.15
98 Food	.98	.97	.30	.11	.13	.18
99 Ham	.96	.95	.31	.15	.11	.19
100 Harvest	.96	.93	.45	.14	.12	.16
101 Orange	.96	.95	.22	.07	.09	.07
102 Pie	.98	.98	.31	.15	.19	.15
103 Rye	.99	.99	.58	.38	.32	.39
104 Sandwich	.96	.92	.27	.06	.09	.09
105 Stomach	.97	.94	.37	.10	.11	.10
106 Table	.98	.98	.40	.22	.18	.23
107 Brick	.98	.95	.39	.12	.13	.15
108 Cave	.97	.95	.28	.09	.10	.12
109 Fence	.95	.94	.29	.06	.08	.06
110 House	.96	.94	.22	.08	.09	.07
111 Hut	.97	.95	.32	.10	.10	.12
112 Lamp	.99	.99	.45	.31	.32	.28
113 Canoe	.99	.98	.45	.16	.15	.14

Table 17 (Continued)

Word		Measure					
		r_{scale}	r_{trt}	Prop. *IAs*[c] (real)	Prop. *IAs*[d] (random)	*Var, - I* (first test)	*Var, - I* (second test)
114	Caution	.96	.93	.32	.04	.05	.06
115	Colour	.97	.97	.33	.14	.14	.14
116	Dart	.97	.95	.37	.13	.09	.17
117	Dime	.99	.89	.38	.16	.14	.23
118	Effect	.97	.98	.47	.20	.18	.19
119	Flag	.93	.96	.27	.10	.07	.09
120	Green	.95	.86	.19	.11	.11	.14
121	Host	.98	.96	.45	.20	.16	.23
122	Joke	.99	.97	.46	.29	.26	.31
123	Metal	.94	.93	.27	.04	.04	.03
124	Sand	.97	.95	.30	.10	.12	.11
125	Street	.96	.85	.23	.05	.07	.07
126	Yellow	.94	.89	.20	.06	.04	.06
	Mean	.96	.94	.33	.14	.14	.15
	S. D.	.03	.04	.12	.13	.12	.13

[a] Each measure is based on 283 observations or pairs of observations.

[b] Truncated at the second decimal.

[c] Prop. *IA*s (real) = the proportion of identical associations actually emitted by the subject sample on the two test administrations.

[d] Prop. *IA*s (random) = the proportion of identical associations obtained from random pairings of sets of responses from the first and second test administrations.

Table 18

Correlations Among Reliability and Variance Measures[a]

Measure	1	2	3	4	5	6
1 r_{scale}	1.00					
2 r_{trt}	.61	1.00				
3 Prop. *IA*s (real)	.50	.60	1.00			
4 Prop. *IA*s (random)	.49	.53	.92	1.00		
5 *Var*, - *I* (first test)	.51	.52	.91	.98	1.00	
6 *Var*, - *I* (second test)	.49	.53	.92	.98	.97	1.00

[a] Based on 126 observations.

tion, is reasonably good. Of the 52 words used in the recall test, 46 have scale reliabilities greater than 0.90. Of the remaining 70 words, 67 have scale reliabilities greater than 0.90. The average of the scale reliabilities is 0.96. For most words, therefore, it can be concluded that the distribution of associations is reasonably consistent for samples of subjects from this population.

The relation of scale reliability to predictability of recall is not very strong in this sample of observations. As was stated previously, the removal from the data set of the six words with scale reliabilities below 0.90 increased the accuracy of prediction, but only slightly. Furthermore, when all the reliability measures were correlated with the error residuals of recall, after the accountable variance was removed, none of the correlations were greater than + or - .003.

The test-retest reliability has almost as high an average (0.94) as the scale reliability, indicating that for most words the distribution of associations is also reasonably stable for groups over time within the population sampled. The correlation between test-retest and scale reliability is only moderate, suggesting that stability of response distributions over time may be due to somewhat different factors than those accounting for stability over subject samples.

The moderate correlations of variance with test-retest and scale reliability, when the effects of identical associations are partialled out, drop to -.07 and +.14, respectively. On the other hand, the correlation of identical associations with variance, when the other reliability measures are separately partialled out, drops only to +.87 in both cases.

The proportion of identical associations, considered as a reliability measure comparable to other reliability indices, presents a very different picture of the reliability of the response distributions than do either the scale or test-retest reliability coefficients. The proportion of identical associations has a mean of only 0.33, far less than the mean of the other reliability measures. The obtained proportion is, however, consonant with the proportions reported by other investigators (Brotsky & Linton, 1967; Gegoski & Riegel, 1967).

The correlations among the test-retest reliability, scale reliability, and proportion of identical associations measures are only moderate in this sample. The three measures certainly cannot be considered equivalent; the forms of stability of the response distributions indexed by the three measures are apparently under the control of somewhat different factors. If a researcher was to make a selection of words for experimental purposes on the basis of their associative reliabilities, he would have to make an *a priori* decision as to what form of reliability was most relevant to his needs.

The Assumption of the Absence of Individual Differences

It was suggested previously that some of the data on reliability of word associations could provide evidence for or against the assumption of the absence of individual differences on the emission of associations. A fairly strong test of the assumption is provided by the relation between the variances of the reponse distributions and the proportions of identical associations, calculated from real subject-pairings of first and second administration test results. The correlation between the two is +.91, and thus provides considerable support for the assumption. The levels of the two distributions however, are significantly different ($p < .01$); the mean proportion of identical associations is 0.19 higher than the mean of the variances. The logic of prediction requires them to be the same. Other factors besides those that have been hypothesized seem to be implicated in the emission of identical associations. The differences in the emission of associations on any single testing session; it may be a phenomenon relating solely to repeated testing.

The proportion of identical associations is related to individual response tendencies more strongly than the other reliability measures. The individual response tendencies can be removed from the proportions by calculating them from random pairings of first and second administration tests. The random-pair proportions have a mean which is only 0.0026 greater than that of the variances; they are correlated +.98

with the variances. This correlation is significantly higher than that between variances and real subject proportions ($p < .01$). The partial correlation between random-pairs proportions and variances with real subject-pair proportions partialled, drops to +.91. The partial correlation of real subject-pair proportions with variances, with random-pair proportions partialled, drops to a negligible +.03.

The random-pair proportions behave, in their distributions, very much like the variances. The real subject-pair proportions behave somewhat less like the variances, and as if an additive constant was applied to them. In addition, the correlation of +.97 between variances on the first and second test administrations lends further confirmation both to the independence of the testing sessions and to the legitimacy of assuming that the first test variances can be used as the predictors of the proportions of identical associations.

The question that remains is how much of the variance of the proportions of identical associations is due to individual response tendencies not shared with the rest of the subject sample. Since the complete absence of individual differences would imply a perfect correlation between variances and real subject proportions, the amount of variance not common to the two distributions can be considered an estimate of the degree to which individual differences, in this subject sample, determine the emission of associations in general.

Several different estimates of the variance due to individuals can be made. The simplest is the proportion of sample variance not common to the variances and the real subject proportions, 1.0 minus the squared correlation between them. The estimate thus derived is 18.0%. Another is 1.0 minus the squared partial correlation between random-pair proportions with variances, real subject-pairs partialled. The estimate from these calculations is 17.4%. A third estimate is derived from the proportion of variance common to the variances and the real subject-pair proportions, with random-pair proportions partialled. This estimate, the square of the partial, is an extremely low 0.09%. Without further evidence, the higher estimate of around 18% is a safer and more appropriate estimate to make.

Chapter 4: Discussion

Evaluation of Summary List Measures

Although the nine predictors of average list recall have predictive validities ranging from +.59 to +.88, the high level of correlations between the predictors and the small number of observations jointly prevent any of the differences from being statistically significant. Any comparative assessment of the measures as predictors is, therefore, necessarily tentative. Nevertheless, it is possible to make some general observations.

The average covariance without the representational response is substantially worse as a predictor than all the others. As neither of the covariances are amongst the best predictors, it seems likely that the variances of the response distributions (which the covariances consider) are not necessary as predictive information, at least when considered in conjunction with the bivariate associative measures. On the other hand, the representational response does not seem necessary either, in this context.

There is some relation between the variance and the representational response. With the representational response, the variances are restricted to a range of 0.25 to 0.50. It is possible that keeping information about variances actually reduces the predictive power of the measure. If this is so, then the representational response restricts the confounding effect of the variances and allows the covariance with the representational response to have a relatively high correlation with recall.

To test this interpretation, average variance of each list was correlated with recall and with average covariance without the representational response. Variance, surprisingly, correlates -.44 with recall; however, it correlates +.30 with average covariance without the representational response. The correlation of the covariance measure with recall when the effects of average variance are partialled out rises to +.85.

Covariance can be considered as a product of correlation times the geometric mean of the variances of the two variables. Thus it seems that a fairly powerful predictor, the average correlation without the representational response, is confounded when multiplied by another variable negatively correlated with recall. The case is somewhat more complicated of course, as it is the individual correlations that are multiplied by the mean variances. It is more customary to consider correlations as derived from covariances than the other way around; the explanatory model, however, is not significantly different in either event. It is not clear why variance and recall are negatively correlated.

It is possible that the average variance of the variables in a matrix relates to the degree of redundancy amongst the variables. If all the items have a very high variance, each one has very few associative pathways. High variance, therefore, implies greater likelihood of associating to one of a relatively few possible words. If none of those few words are on the list of words to be recalled, high variance will impede recall. This interpretation is, at this time, highly speculative; it is somewhat supported, however, by the fact that the partial correlation between average variance and recall, average covariance partialled, rises to -.81.

Between the conditional probabilities correlation and the *MF* index there seems little basis for preference on predictive grounds. Both measures, with and without the representational response, are correlated quite highly with recall. The slight differences between their single predictive validities are too small to be considered of much substance in this sample.

Use of the representational response does not increase the accuracy of prediction of average list recall. It does, however, increase agreement or homogeneity among the predictors. The rankings of the ten word lists according to rated level of association are considerably more similar when the ranking agents are the three average bivariate measures with the representational response than when they are the three averages without the representational response. Whether or not the increase in homogeneity is desirable depends, of course, on whether the differences in ranking are due to sampling error or to genuine differences in the information provided by the measures. Does the representational response smooth out error variance or does it obscure real differences among the measures? The small number of observations in the present study again precludes a definite conclusion.

The predictive validity of *IIAS* is +.79. This figure falls within the range of those previously reported, and confirms its usefulness when used as the only predictive measure. Reports have not previously appeared relating *ITA* to recall of complete lists of words. The correlation of +.75 observed between *ITA* and recall is much higher than was previously reported between *ITA* and recall of sublists (Marshall, 1967). The average cue number also has not been reported as a predictive measure; its predictive validity of +.79 indicates that it might be useful in situations where a quick, simple summary measure is desired. Because of the ambiguous findings of Pollio and Christy (1964) regarding use of *Nc* as a predictor of single word recall, it would be advisable to submit it to further testing before accepting it for use in any form.

In the factoring of the summary list measures, it is interesting that the conditional probabilities measures loaded on different factors from the *MF* indices. The factoring cannot, however, be considered reliable since the first principle component accounts for so much of the total variance. Varimax rotation can sometimes force variables into separate factors when they do not really belong there; the forcing can most easily occur when the variables are all highly correlated, as is the case here.

Evaluation of Single Word Measures

The assessment and comparison of measures is not so difficult in the present case as in the previous one, because the predictors are not so highly correlated with each other or with the criterion, and because there are a greater number of observations.

The best associative predictors are the squared multiples and the mean squared correlations with the representational response. The squared multiple conditional probabilities correlation has a significantly greater predictive validity than the squared multiple *MF* ($p < .01$). There is no significant difference between the correlation and covariance, or between the covariance and *MF*. The greater predictive power of the correlation indicates that it is slightly preferable to the *MF* index on predictive grounds. This conclusion must be qualified, however, since the differential predictive power of the conditional probabilities correlation was not significant when prediction was considered on the ten word lists separately. The main effect for measures in the analysis of variance reported in Chapter 3 was negligible.

The representational response is clearly appropriate for prediction of single word recall. The correlations with the criterion of predictors without the representational response are all considerably lower than of those computed with it. The question still exists, of course, of whether the representational response could be replaced by an equivalent construct which more precisely reflects the different kinds of associative bonding.

In selecting optimal subsets of predictors, a combination of two predictors appears to be adequate. Serial position and any one of the three squared multiples or mean squared correlations with the representational response are appropriate. The conditional probabilities correlation is again more effective than any of the others, but the difference is not as great as when the association measures are used alone. However, the decrease in the differential effectiveness of the squared multiple correlation measure is probably an artifact due to chance differences in the correlations of all the association measures with serial position. The "true" correlation is of course zero, and the differential effectiveness of the correlation measure should remain exactly the same when serial position is added to the predictor set. Whichever of the six measures is used, the addition of any more predictors does not afford a large increase in the percentage variance accounted for, and probably lowers the reliability of the prediction battery.

It is of interest to note that Deese's (1960) finding concerning word frequency is confirmed. Word frequency correlates significantly with recall ($p < .01$), but does not increase the accuracy of prediction over that obtained with the associative measures alone. Apparently, the predictive information it conveys is almost fully duplicated by the other measures.

The finding of Rothkopf and Coke (1961b) that the cue number predicts recall more effectively than the frequency of elicitation is not confirmed in this study. Frequency of elicitation predicts recall as accurately as the cue number does; the slight difference in favor of *FE* is not significant. The correlation of both variables with recall is somewhat lower than that reported by Rothkopf and Coke (1961b). They reported validities of +.62 and +.54 for cue number and frequency of elicitation respectively, compared with +.37 and +.40 found in the present study. Their word list was rather unusual, however, being 99 items long; the difference in length might be related to the difference in validity. In addition, the cue number appears to be rather unreliable as a predictive index (Pollio & Christy, 1964).

Variance of the response distributions is ineffective as a predictor of single word recall. This finding, although negative, is of interest as it is somewhat surprising that the variance is not a relevant variable, since the average variance is in fact relevant to the prediction of average list recall. In particular, the interpretation of the variance as a measure of redundancy, which could inhibit recall, should apply at least as well to the prediction of single word recall as to the prediction of average list recall.

The prediction of recall with list means extracted is consistently slightly worse than that of raw recall. Most of the regressions that have been described were also conducted on recall without list means; these regressions have not been reported. The variance accounted for was uniformly two to three percentage points below that accounted for in the corresponding regressions of raw recall. The reason for the decline in performance is undoubtedly that recall without list means has removed from it a source of variance extremely well predicted by the associative measures, namely the average list recall. The percent of total recall variance due to list means is very low, approximately 3.75%.

The factor analysis of the predictors is of interest chiefly because it confirms the previous findings regarding the nature of the splits between groups of variables. All of the good associative predictors load on the same factor; the major split is between presence and absence of the representational response. The means and determinants almost all load on the same factor, and have little to do with any other variables. The variables that are somewhat isolated from the rest, such as the variances, serial position, and word frequency, tend to load on their own factors.

Assessment of Reliability Measures

The test-retest and scale reliabilities of the response distributions are impressively high. Considerable confidence can be placed in the response distributions of a great majority of the stimulus words. The correlation of only +.62 between test-retest reliability and scale reliability is unexpected. Evidently, there are different factors governing similarity of the response distributions over time as opposed to over subsamples; it is not clear just what the differences are. The probable cause of the correlations between variance and the two reliability measures has previously been explained; variance does not, by itself, seem to significantly affect the reliability of the response distributions.

The reliabilities of the response distributions do not seem strongly related to predictability of recall. The improvement in prediction with the low-reliability words removed from analysis is only slight. Both forms of reliability were correlated with the recall error residuals, but neither of the correlations were higher than .003. Low reliabilities might restrict the prediction of recall of course, but they were sufficiently high in the present situation that they did not do so.

The magnitude of the proportions of identical associations offers striking contrast to that of the other reliability measures. The average is far smaller, and the variance is substantially greater. Very different conclusions concerning reliability would be reached if the proportion of identical associations was interpreted as a test-retest reliability coefficient. As was explained previously, it is not fully appropriate to do so if interest is centered on the stability of the response distributions collected from the pooled sample.

Assumption of the Absence of Individual Differences

The test-retest reliability coefficients are sufficiently large that the two administrations of the word association test can be considered statistically independent samplings of the same behaviour domain within groups of subjects. It is therefore appropriate to examine the relation between the variance of the response distributions and the proportion of identical associations in order to assess the validity of the assumption of the absence of intra-individual constraints on the emission of associations. As was stated in the preceding chapter, the correlation between the two sets of observations is sufficiently high as to lend considerable support to the assumption. The assumption is not completely justified by the data, however, even within the limits of the observed correlation. The differences in the levels of the two distributions are quite different; the assumption and the logic of prediction require them to be the same.

Two of the estimates of the extent of individual determinants on the emission of associations agree that slightly less than 18% of the total variance is individually and uniquely determined. If these estimates are valid, then the assumption of the absence of individual differences is true with regard to most of the variance. The remaining

18% can be considered to somewhat reduce the validity of all measures which rely on the assumption.

Besides the variance in a single set of response distributions due to individual differences, there seem to be a degree of stability in the associations of individual subjects which is independent of the between-subject constraints. This stability is reflected in the mean level of the proportions of identical associations; it does not affect the variance of the proportions.

Reliability of Free Recall

The very moderate predictability of single word recall in this study is of considerable interest, as it focuses attention on the other determinants of recall besides the associational ones used here. Before searching for better predictors, however, it is necessary to consider the extent to which recall may be simply unpredictable from group measurements. The reliability of any dependent variable sets an upper limit on its predictability; as yet, the reliability of free recall is not known.

The reliability of recall, measured across different samples of subjects, would presumably index the degree to which recall is determined by stable between-subject factors. The amount of variance common to two recall distributions could easily be estimated with a form of the scale reliability technique. Either the proportional recall of single words by two subject samples or average list recall by two samples could be correlated with an ordinary Pearson product-moment correlation. The conditional probabilities correlation is not appropriate in this context as the recall data are not conditional probabilities; they are not discrete, do not sum to 1.0, and each response is not mutually exclusive with all other responses.

To the best of the author's knowledge, the split-sample reliability of free recall has not been investigated. Repeated testing on the same word lists is of course common, and could be used to provide an estimate of test-retest reliability. Test-retest recall data, however, are usually obtained specifically to investigate the change in recall as a function of practice and other experimental variables. The practice effect would tend to confound the reliability estimates.

Estimates of the reliability of free recall would be valuable for at least two purposes. In their own right, they would provide information concerning the extent to which recall is determined by group factors and the extent to which it is determined by unique within-subject factors. As a practical consideration, the same estimates would indicate the upper limit to which recall could be meaningfully related to any group-based measures.

Implications for the Study of Association

The findings of this study make possible several strong recommendations concerning the study of associations and of associative structure.

The Choice of Measures of Association

The conditional probabilities covariance and correlation have shown themselves at least equally effective as any other measures in the prediction of both average list recall and single word recall. They are, therefore, at least as valuable as any other measures on strictly empirical grounds.

On logical grounds they are preferable to any other measures. Only in the case of product-moment correlations and covariances does the matrix of bivariate relations contain the information necessary to determine the trivariate, quartovariate, and every succeeding level of multivariate relations among the variables. This property is essential if multivariate statistics computed from the bivariate measures are to be accorded any confidence. As legitimate variants on the product-moment measures, the conditional probabilities correlation and covariance share this important property. The property should not be assumed to be present in the *MF* index, as the derivation of the *MF* index from the product-moment correlation is highly questionable. The property is absent, of course, in the non-bivariate measures such as *IIAS*, *Nc*, etc.

The choice of the correlation or the covariance measure as a tool for multivariate analysis cannot be made independently of a specific experimental situation. Both measures are appropriate, but they differ on the important characteristic of the use they make of the variance of the response distributions. The correlation effectively equates the variances of every response distribution; the covariance does not. The variance of the response distributions may well be an important factor in determining the structure of associations. If so, the covariance may be preferable to the correlation.

In any situation where a measure of the overall relation between one word and a number of others is required, the squared multiple correlation or covariance is the preferred measure. Empirically, the squared multiple correlation and the mean of the squared correlations are the most effective multivariate statistics for the prediction of single word recall. As the squared multiple correlation is, in any sample, the exact measure of the common variance of one variable with a set of others, it is logically preferable to the mean of the squared correlations. The SMC gives the most precise estimate available of the relation. It may be more sensitive to sampling error and resulting shrinkage (Guilford, 1956); however, as it has a known sampling distribution the error can be more effectively estimated and corrected for than in the case of the mean squared correlation.

If the study of word associations is to progress in the development of sophisticated theories of the structure and basis of associations, it is inevitable that, to make full use of the association data, multivariate analysis will have to be used more and more extensively. The conditional probabilities correlation and covariance are the only measures reviewed in this study which can safely be used as the basis for multivariate analysis. They are, therefore, strongly recommended.

The Choice of Response Measures

In the present study, all the association data were collected from single-response free associations. It is quite possible that this experimental technique is inadequate to tap the structural characteristics of word association hierarchies. The alternatives are continuous or repeated association on the one hand, and restricted association on the other. In restricted association, the experimenter forces external constraints upon the subjects' association, such as "What is the first noun (animal, emotion, person, etc.) you think of when you see the stimulus word?" The semantic differential (Osgood, Suci, & Tannenbaum, 1957) can be considered as an extreme case of restricted association: "Does this word make you think more of "hot" or of "cold"?

Both continuous and restricted association have been used in several association studies, some of which have been reviewed in this report. Most investigators of associative structure, however, have concentrated on the single-response free association technique (e.g., Deese, 1965; Pollio, 1966). This emphasis on an experimentally simple paradigm is appropriate in the beginning phases of an investigation, but may be quite inappropriate later on. The idea of hierarchical structure certainly implies that the content of associations is in part determined by factors other than those specific to the eliciting stimulus. It is reasonable to expect that the psychological or situational context of the associations, and their position in an associative chain, may be influential.

If such considerations are valid, then it may be that far more of the variance in single-word recall is potentially accountable than was in fact accounted for in the present study. If recall is largely determined by the structure of associations, then response measures which more adequately reflect this structure may greatly increase the power of prediction.

All of the bivariate association measures reviewed in the present study are appropriate for the measurement of continuous and restricted associations, and have been used in such applications in the past (continuous restricted associations with the conditional probabilities covariance, Rothkopf, 1960; restricted associations with the conditional probabilities correlation, Kendall, 1962; continuous associations with the *MF* index, P. M. Jenkins & Cofer, 1957).

In summary, the study of the structure of associations will require a greater emphasis on more sophisticated paradigms than the single-response free association technique. More elaborate paradigms, incorporating repeated, continuous, and restricted associations are still amenable to analysis with the conditional probabilities measures.

Individual vs. Group Measurements of Association

The analysis and experimental evidence in this study of the assumption of the absence of individual differences in the emission of associations is of considerable significance. It is of much less value to construct a theory of associative structure if the theory is irrelevant to the determinants of association in the individual, than if it applies to both individual and group processes. If the evidence and analysis presented in this report are valid, then it is possible to make far more precise estimates of the extent of individual and group determinants than has been possible previously. The analysis of the assumption is dependent on a rather elaborate logical chain, and the evidence in favor of it, although strong, is necessarily indirect. A critical appraisal of the logic used is clearly necessary, as is confirmation of the empirical findings. If the findings withstand the appraisal and the attempts at replication, the next necessary step is to extend the investigation to the more elaborate paradigms of association mentioned above.

In combination, the estimate of the extent of group influences on association, and the development of appropriate correlational techniques for the measurement of bivariate and multivariate associative relationships, present a potentially very powerful set of tools for the construction of associative theories.

Concluding Note: On Measurement and Methodology in the Study of Associations

The quotation from Deese (1968) on page 4 of this report testifies to the conservatism of studies of association, and to their isolation from the mainstream of psychological research. The causes of the isolation are difficult to determine, but undoubtedly include the ready-made theories of and assumptions about associations provided by the British empiricists. One of the most curious facets of the isolation has been the failure of word association studies to partake of the powerful tools of analysis employed in other areas of psychological research since the turn of the century.

The measures of word association that have appeared in the literature have for the most part been "home grown" measures, only marginally related to developments in statistics and measurement theory. Inevitably, the reliability, validity, and generality of the measurements of association suffer as a result. There is no need, however, to use a special and unique set of basic measures in determining associative patterns. The more powerful techniques developed for general use in psychology and other disciplines are at least equally appropriate. In the area of measurement, at least, it is time that studies of association were brought more consistently into the mainstream of psychological research.

Conclusions

1. The prediction of average list recall is made about equally well by several measures. The *MF* index predicts slightly better than the correlation and covariance measures, but not significantly so. Presence or absence of the representational response is irrelevant to prediction. On the other hand, scaling the predictors by a function of the variances decreases the accuracy of prediction, as the variance is negatively correlated with recall.

2. The prediction of single word recall is best with the squared multiple or mean squared correlation with the representational response, derived from any one of the conditional probabilities covariance, conditional probabilities correlation, or *MF* index. The serial position of the word in the recall list provides the only significant additional predictive information. In all the effective bivariate measures, the representational response is necessary to maximize the accuracy of prediction. Word frequency is correlated with recall, but provides no predictive information not also provided by the bivariate association measures.

3. The response distributions of the great majority of the items on the word association test, pooled within samples, are highly consistent across subsamples and stable across repeated tests. Scale reliability and test-retest reliability are affected by somewhat different factors, but it

is not clear what the factors are. The proportion of identical associations given on retesting is not an appropriate measure of test-retest reliability of the response distributions characteristic of the entire sample. Differences in the reliability of response distributions are not, within the observed range of reliabilities, related to accuracy of prediction of recall.

4. The assumption of the absence of individual differences on the emission of associations is given considerable support. The amount of variance dependent on individual differences is estimated at about 18%. In addition, there is a degree of stability over time of responding for each subject, independent of the constraints upon single-trial emission of associations.

5. The conditional probabilities covariance and correlation are the most appropriate measures of association on logical grounds. As the conditional probabilities measures are the proper variants on the product-moment correlation and covariance, they are uniquely appropriate for use in multivariate analysis. The squared multiple correlation or covariance is the most appropriate statistic for measuring the relation of one word to a set of others, as it provides a precise estimate within the sample of the common variance.

References

Attneave, F. *Applications of information theory to psychology.* New York: Holt, Rinehart, & Winston, 1959.

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology,* 1953, *49*, 229-240.

Bousfield, W. A., Cohen, B. H., & Silva, J. G. The extension of Marbe's law to the recall of stimulus words. *American Journal of Psychology,* 1956, *69,* 429-433.

Bousfield, W. A., Cohen, B. H., Whitmarsh, C. A., & Kincaid, W. D. The Connecticut free associational norms. *Technical Report No. 35,* Contract Nonr 631 (00), Office of Naval Research, 1961.

Bousfield, W. A., & Puff, C. R. Determinants of the clustering of taxonomically related word pairs. *Journal of General Psychology,* 1965, *73,* 211-221. (a)

Bousfield, W. A., & Puff, C. R. Relationships between free associational and judgmental measures of word relatedness. *Psychological Reports,* 1965, *16,* 279-282. (b)

Bousfield, W. A., Steward, J. R., & Cowan, T. M. The use of free associational norms for the prediction of clustering. *Journal of General Psychology,* 1964, *70*, 205-214.

Bousfield, W. A., Whitmarsh, G. A., & Berkowitz, H. Partial re response identities in associative clustering. *Journal of General Psychology,* 1960, *63,* 233-238.

Brotsky, S. J., & Linton, M. L. The test-retest reliability of free association norms. *Psychonomic Science,* 1967, *8,* 425-426.

Cofer, C. N. Associative commonality and rated similarity of certain words from Haagen's list. *Psychological Reports*, 1957, *3*, 603-606.

Cofer, C. N., & Foley, J. P. Mediated generalization and the interpretation of verbal behaviour. I. Prolegomena. *Psychological Review*, 1942, *49*, 513-540.

Cohen, B. H. An investigation of recoding in free recall. *Journal of Experimental Psychology*, 1963, *65*, 368-376. (a)

Cohen, B. H. Recall of categorized word lists. *Journal of Experimental Psychology*, 1963, *66*, 227-234. (b)

Cowan, T. M. A Markov model for order of emission in free recall. *Journal of Mathematical Psychology*, 1966, *3*, 470-483.

Deese, J. Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, 1959, *5*, 305-312. (a)

Deese, J. On the prediction of occurrence of particular verbal intrusions in immediate free recall. *Journal of Experimental Psychology*, 1959, *58*, 17-22. (b)

Deese, J. Frequency of usage and number of words in free recall: The role of association. *Psychological Reports*, 1960, *7*, 337-344.

Deese, J. Associative structure and the serial reproduction experiment. *Journal of Abnormal and Social Psychology*, 1961, *63*, 95-100. (a)

Deese, J. From the isolated verbal unit to connected discourse. In C. N. Cofer (Ed.), *Verbal learning and verbal behaviour.* New York: McGraw-Hill, 1961. (b)

Deese, J. On the structure of associative meaning. *Psychological Review,* 1962, *69,* 161-175.

Deese, J. The associative structure of some common English adjectives. *Journal of Verbal Learning and Verbal behaviour,* 1964, *3,* 347-357.

Deese, J. *The structure of associations in language and thought.* Baltimore: Johns Hopkins Press, 1965.

Deese, J. Association and memory. In T. R. Dixon and D. L. Horton (Eds.) *Verbal behaviour and general behaviour theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Deese, J., & Kaufman, R. A. Serial effects in recall of unorganized and sequentially organized verbal behaviour. *Journal of Experimental Psychology,* 1957, *54,* 180-187.

Garskof, B. E. Relation between single word association and continued association response hierarchies. *Psychological Reports,* 1965, *16,* 307-309.

Garskof, B. E., Houston, J. P., & Ehrlich, N. J. Inter- and intra-hierarchical verbal relatedness. *Journal of Verbal Learning and Verbal Behaviour,* 1963, *2,* 229-233.

Gegoski, W. L., & Riegel, K. F. A study of the one year stability of the Michigan free and restricted association norms. *Psychonomic Science,* 1967, *8,* 427-428.

Goss, A. E. Early behaviourism and verbal mediating responses. *American Psychologist,* 1961, *16,* 285-298.

Guilford, J. P. *Fundamental statistics in psychology and education.* (3rd ed.) New York: McGraw-Hill, 1956.

Gulliksen, H. *Theory of mental tests.* New York: Wiley, 1950.

Gulliksen, H. Methods for determining equivalence of measures. *Psychological Bulletin,* 1968, *70,* 534-554.

Hall, J. F. The reliability of free association responses. *Psychonomic Science,* 1966, *4,* 79-80.

Horvath, W. J. A stochastic model for word association tests. *Psychological Review,* 1963, *70,* 361-364.

Howe, E. S. Associative structure of quantifiers. *Journal of Verbal Learning and Verbal Behaviour,* 1965, *5,* 156-162.

Jenkins, J. J., Mink, W. D., & Russell, W. A. Associative clustering as a function of verbal association strength. *Psychological Reports,* 1958, *4,* 127-136.

Jenkins, J. J., & Russell, W. A. Associative clustering during recall. *Journal of Abnormal and Social Psychology,* 1952, *47,* 818-821.

Jenkins, P. M., & Cofer, C. N. An exploratory study of discrete free associations to compound verbal stimuli. *Psychological Reports,* 1957, *3,* 599-602.

Kelley, T. L. *Interpretation of educational measurements.* New York: World Book, 1927.

Kelley, T. L. An unbiased correlation measure. *Proceedings of the National Academy of Sciences,* 1935, *21,* 554-559.

Kendall, L. M. The Rorschach, psychodynamics and psychophysics. Paper read at the 21st Annual Meeting of the Canadian Psychological Association, Hamilton, May, 1962.

Kent, G. H., & Rosanoff, A. J. A study of association in insanity. *American Journal of Insanity*, 1910. pp. 67, 37-96, 317-390.

McQuitty, L. L. Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. *Educational and Psychological Measurement*, 1957, *17*, 207-229.

Marshall, G. R. Effect of total association and conceptual cohesiveness among words on recall, clustering, and recognition association. *Psychological Reports*, 1967, *20*, 39-44.

Marshall, G. R., & Cofer, C. N. Associative indices as measures of word relatedness: A summary and comparison of ten methods. *Journal of Verbal Learning and Verbal behaviour*, 1963, *1*, 408-421.

Meltzer, L. Comparing relationships of individual and average variables to individual response. *American Sociological Review*, 1963, *28*, 117-123.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, *63*, 81-97.

Noble, C. E. Scale reliability and the Spearman-Brown equation. *Educational and Psychological Measurement*, 1955, *15*, 195-205.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning.* Urbana, Illinois: University of Illinois Press, 1957.

Palermo, D. S., & Jenkins, J. J. *Word association norms grade school through college.* Minneapolis: University of Minnesota Press, 1964.

Peters, C. C., & van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.

Pollio, H. R. *The structural basis of word association behaviour.* The Hague: Mouton & Co., 1966.

Pollio, H. R. Associative structure and verbal behaviour. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behaviour and general behaviour theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Pollio, H. R., & Christy, E. G. Variation of inter-item associative strength and some measures of free recall. *Psychological Reports,* 1964, *15,* 527-534.

Postman, L. The effect of interitem associative strength on the acquisition and retention of serial lists. *Journal of Verbal Learning and Verbal Behaviour,* 1967, *6,* 721-728.

Robinson, W. S. Ecological correlations and the behaviour of individuals. *American Sociological Review,* 1950, *15,* 351-357.

Rosenthal, R. *Experimenter effects in behaviour research.* New York: Appleton-Century-Crofts, 1966.

Rosenzweig, M. R. Word associations of French workmen: Comparisons with associations of French students and American workmen and students. *Journal of Verbal Learning and Verbal Behaviour*, 1964, *3*, 57-69.

Rosenzweig, M. R., & Miller, K. M. Comparison of word association responses obtained in the United States, Australia, and England. *Journal of Verbal Learning and Verbal Behaviour*, 1966, *5*, 35-41.

Rosner, B. S. A new scaling technique for absolute judgments. *Psychometrika*, 1956, *21*, 377-381.

Rothkopf, E. Z. Two predictors of stimulus equivalence in paired associate learning. *Psychological Reports*, 1960, *7*, 241-250.

Rothkopf, E. Z., & Coke, E. U. Intralist association data for 99 words of the Kent-Rosanoff word list. *Psychological Reports*, 1961, *8*, 463-474. (a)

Rothkopf, E. Z., & Coke, E. U. The prediction of free recall from word association measures. *Journal of Experimental Psychology*, 1961, *62*, 433-438. (b)

Rozeboom, W. W. Linear correlations between sets of variables. *Psychometrika*, 1965, *30*, 57-71.

Rozeboom, W. W. The theory of abstract partials. *Psychometrika*, 1968, *33*, 133-167.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *Office of Naval Research Technical Report*, 1954, No. 11 (Contract No. N8-ONR-66216).

Siegel, S. *Nonparametric statistics for the behavioural sciences.* New York: McGraw-Hill, 1956.

Simon, S., & Hess, J. L. Supplementary report: Influence of inter-item associative strength upon immediate free recall in children. *Psychological Reports,* 1965, *16,* 451-455.

Smith, P. C., Kendall, L. M., & Hulin, C. L. *The measurement of satisfaction in work and retirement: A strategy for the study of attitudes.* Chicago: Rand McNally, 1969.

Spearman, C. The proof and measurement of association between two things. *American Journal of Psychology,* 1904, *15,* 72-101.

Thomson, G. H. The hierarchy of abilities. *British Journal of Psychology,* 1919, *9,* 337-344.

Thomson, G. H. On complete families of correlation coefficients and their tendency to zero tetrad differences: Including a statement of the sampling theory of abilities. *British Journal of Psychology,* 1935, *26,* 63-92.

Thomson, G. H. *The factorial analysis of human ability.* (5th ed.) London: University of London Press, 1951.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Waugh, N. C. Free versus serial recall. *Journal of Experimental Psychology,* 1961, *62,* 496-502.

Weingartner, H. Associative structure and serial learning. *Journal of Verbal Learning and Verbal Behaviour,* 1963, *2,* 476-479.

Willner, A. E., & Reitz, W. E. Association strength as a mediating variable in verbal categorizing and recall: Preliminary report. *Psychological Reports*, 1965, *16*, 310.