A major modification of the Bousfield (1966) measure of category clustering[1]

ROBERT L. HUDSON,[2] *Central Washington State College, Ellensburg, Wash. 98926, and* JAMES E. DUNN, *University of Arkansas, Fayetteville, Ark. 72701*

The measure of category clustering proposed by Bousfield & Bousfield (1966) was modified to take into account differences in the distribution of the expected value of number of repetitions. The possible effects of using the clustering index without this major modification are discussed.

Bousfield & Bousfield (1966) define an index of category clustering as the observed number of repetitions minus the expected number, where a repetition refers to the occurrence of two words from a category in succession. Thus a cluster of two words from a category is one repetition, a cluster of three words is two repetitions, etc. The general formula for the expected number of repetitions, assuming that the sequence is generated at random, is as follows:

$$E(R) = \left[(M_1^2 + M_2^2 + \cdots + M_k^2)/N\right] - 1 \qquad (1)$$

where the M_i are the number of items recalled from Category i as i takes the values from 1 to k, and N is the total number of items recalled.

The number of words recalled from each category is treated as a constant rather than a random variable, and the formula computes the chance number of repetitions given M_1 words recalled from Category 1, M_2 words recalled from Category 2, etc. One takes as the measure of clustering for an S the difference between this expected value and the actually observed number of repetitions.

The problem of determining the chance number of repetitions in a series of items is the same as the problem of determining the chance number of runs where a run is a consecutive series of items from the same set or category and is thus the same as a cluster. The total number of items is equal to the sum of the number of repetitions and the number of clusters (runs).

Barton & David (1957) developed a formula for the expectation of the number of runs in a series of items which, with a little manipulation, turns out to be the formula for expectation of the number of repetitions given above. The same authors derived a general formula for the variance of the distribution. This variance, of course, differs from case to case depending upon the number of categories and the number of items recalled from each category. The formula for the general case with k observed categories is as follows:

$$\text{Variance} = \frac{F_2(N-3)}{N(N-1)} + \frac{F_2^2}{N^2(N-1)} - \frac{2F_3}{N(N-1)} \qquad (2)$$

where $F_2 = \sum_{i=1}^{k} M_i(M_i - 1)$

$$F_3 = \sum_{i=1}^{k} M_i(M_i - 1)(M_i - 2)$$

and where M_i and N are as previously defined.

The distribution for the expected number of repetitions varies depending on the number of different kinds of items and the different number of each kind recalled. Therefore, it seems that a much more stable and meaningful index of clustering is provided by subtracting the expected from the observed number of repetitions and dividing by the SD for that particular distribution. This, of course, produces approximately normally distributed standard scores which allow direct comparison of Ss who recall words from different numbers of categories and who have different recall totals. This strategy is similar to that of Battig, Allen, & Jensen (1965) who developed an index to indicate the priority of newly-learned items in free recall; this gives comparable measures across recall lists of varying lengths.

Table 1
SD as a Function of Number of Words Recalled Per Category and Number of Categories

		Words Per Category				
		2	4	6	8	10
	12	.98	1.68	2.16	2.55	2.88
Number of	6	.95	1.62	2.07	2.44	2.76
Categories	4	.93	1.56	1.98	2.33	2.63
	2	.82	1.31	1.65	1.93	2.17

Table 1 shows the change in SD as the number of words per category and number of categories vary; Table 2 indicates the change in SD as the number of categories and total number of items vary. In computing the figures in Table 2 it was assumed that an equal number of words was recalled from each category for a particular total recall. The total-recall SD also varies as the number of words retrieved per category departs from equality. These changes are relatively minor, however, within the limits usually observed in recall protocols.

It would seem that use of the index without taking the SD into account could possible result in some misleading inferences. This is particularly likely, for example, in a study which varies the number of categories but keeps the total number of words the same. For example, Dallett (1964) reported an experiment in which variation in clustering was studied as a function of number of categories, keeping the total number of items at 24. He noted the following respective mean differences between observed and expected number of repetitions for 2, 4, 6, 8, and 12 response categories: 3.21, 3.50, 3.33, 1.24, 1.12. Since raw data were not provided, it is difficult to determine the effect of dividing by SD. However, it is obvious from Table 2 that the SD would probably decrease from two to 12 categories. This would tend to make differences in clustering diminish considerably.

Some hypothetical data will help illustrate the process. Suppose one S recalled 24 words, 12 words from each of two categories. Suppose, further, that 21 repetitions were observed in his recall protocol. From Formula 1, the expected number of repetitions for this case is 11 and the difference between the observed and expected number of repetitions is 10. Now assume an S in the 12-category group also recalled 24 words but these were distributed as two words per category. Assume

Table 2
SD as a Function of Total Words Recalled and Number of Categories

		Total Words Recalled			
		12	24	36	48
	12	0.00	.98	1.37	1.68
Number of	6	.95	1.62	2.07	2.44
Categories	4	1.28	1.98	2.48	2.90
	2	1.65	2.40	2.96	3.43

that 11 repetitions were observed in this S's protocol. By Formula 1 the expected number of repetitions is 1.17 and the difference between the observed and expected number is 9.83. Without going further one might conclude that the two Ss evidence almost the same amount of clustering. However, for the first S the SD (by Formula 2) is 2.40 and for the second S it is .98. Thus, when these SDs are used to divide the above difference scores, the resulting clustering indexes are 4.17 and 10.03, respectively. Both values are considerably above chance, the second much more than the first. So it is clear that if groups are to be compared for amount of clustering when the numbers of categories recalled differ from group to group, the standard score method should be used.

Dallett also found differences in number of words recalled with a maximum mean of about 13.5 for the two-category group and a minimum of 10.5 for the eight. From Table 2 the effect of this would be to reduce differences in clustering, although in this case, the difference would be relatively small. It is clear, however, that anytime independent or task variables effect differentially the number of words recalled, the standard score method should be used, even if the number of categories represented from group to group is the same.

Several indexes of clustering have appeared in the literature (Cohen, Sakoda, & Bousfield, 1954) and it is very possible that some are superior to the standard score index for specific purposes. However, when the Bousfield & Bousfield (1966) measure is to be used SD should be taken into account.

REFERENCES

BARTON, D. E., & DAVID, F. N. Multiple runs. Biometrika, 1957, 44, 168-178.

BATTIG, W. F., ALLEN, A. R., & JENSEN, A. R. Priority of free recall of newly learned items. Journal of Verbal Learning & Verbal Behavior, 1965, 4, 175-179.

BOUSFIELD, A. K., & BOUSFIELD, W. A. Measurement of clustering and of sequential constancies in repeated free recall. Psychological Reports, 1966, 19, 935-942.

COHEN, B. H., SAKODA, J. M., & BOUSFIELD, W. A. The statistical analysis of the incidence of clustering in the recall of randomly arranged associates. Technical Report No. 10, 1954, University of Connecticut, Contract Nonr-631(00), Office of Naval Research.

DALLETT, K. M. Number of categories and category information in free recall. Journal of Experimental Psychology, 1964, 68, 1-12.

NOTES

1. We would like to acknowledge the computational assistance of Mr. Richard Edwards.

2. Present address: Department of Psychology, Mississippi State University, State College, Miss. 39762.

Recall and anticipation methods in probabilistic associative learning[1]

JAMES F. VOSS, University of Pittsburgh, Pittsburgh, Pa. 15213

Six S-R_1, S-R_2 modified paired-associates were presented for 60 trials. A 2 by 6 factorial design was employed in which method, recall or anticipation, and the S-R_1, S-R_2 probability-ratio were varied. The S-R_1, S-R_2 ratios were: 1.00-0.00, .90-.10, .80-.20, .70-.30, .60-.40, and .50-.50. The results indicated that the recall procedure yielded superior performance in early acquisition. In later stages, recall yielded results approaching matching, whereas the anticipation procedure yielded overmatching of R_1 anticipations and undermatching of R_2 anticipations. The results were interpreted as showing that as probability-ratio varies from .50-.50 to 1.00-0.00, and hence, as the relative difference of response strengths of S-R_1 and S-R_2 increases, the selection factor in the anticipation procedure yields a heavy weighting of the stronger response, R_1, and the acquisition of the weaker association, S-R_2, occurs slowly; for the recall paradigm, however, the data suggest that R_1 is inhibited and S-R_2 is more readily acquired. The results also were interpreted as showing that one difference in the standard paired-associate recall and anticipation paradigms is that in the recall method, strong error tendencies may be inhibited, but in the anticipation method, strong error tendencies are weighted heavily.

In recent years the traditional paired-associate paradigm has been modified in order to study probabilistic rather than invariant occurrence of responses (e.g., Erdelyi, Watts, & Voss, 1964; Goss & Sugarman, 1961; Voss, Thompson, & Keegan, 1959). The procedure in such experiments has been to pair a stimulus with two (or more) responses, with the stipulation that only one of the responses is presented with the stimulus on any particular trial. Moreover, in the single-stimulus, two-response situation, S-R_1 and S-R_2 usually have a fixed probability-ratio, e.g., .70-.30, with S-R_1 and S-R_2 occurrence randomized over trials. In addition, the anticipation method always has been used.

The purpose of the present experiment was to compare the acquisition of S-R_1, S-R_2 associations under conditions of recall (R) and anticipation (A) and under probability-ratios varying from 1.00-0.00 to .50-.50. In the paradigm, S-R_1 or S-R_2 is presented on a particular trial and subsequently upon presentation of the stimulus, S is asked to try to recall the correct response, i.e., R_1 or R_2, that was presented with the stimulus on the same trial. Thus, the correct response is the response that most recently was presented with the stimulus. In the S-R_1, S-R_2 anticipation paradigm, the stimulus is presented on a particular trial and S is asked to try to anticipate the response that he thinks will be presented on that particular trial. In this case, the correct response also is designated as the response, R_1 or R_2, that occurs on that particular trial. However, with R_1 and R_2 randomized, S cannot know whether it is R_1 or R_2 that is correct when he makes his response. The major difference, therefore, in the S-R_1, S-R_2 recall and anticipation paradigms is that in the former, whether R_1 or R_2 is correct is designated before S responds, but in the latter R_1 or R_2 is designated after he responds. Furthermore, it also should be pointed out that in the recall procedure, 100% correct responses is possible; in the anticipation procedure, however, with R_1 and R_2 randomized, 100% correct responses is a virtual impossibility.

Because of the previous findings which have indicated overmatching in the .60-.90 recall probability conditions, e.g., Voss, Thompson, & Keegan (1959), and because of the differences in the recall and anticipation paradigms, it was expected that Condition R would yield more correct responses and performance more closely approximating matching behavior. In addition to these hypotheses, which are somewhat obvious from the comparison of the recall and anticipation paradigms, it was anticipated that the relative difference of the recall and anticipation conditions would increase as the probability-ratio varied from 1.00-0.00 to .50-.50. Such performance was expected on the basis of the possible increase