Journal of Applied Psychology
Vol. 41, No. 1, 1957

Prediction of Soldiers' Food Preferences by Laboratory Methods [1]

David R. Peryam and John G. Haynes [2]
Quartermaster Food and Container Institute, Chicago, Illinois

Prediction of the acceptability of foods to potential consumers has become an important problem to the food industries in recent years, and is perhaps even more important in planning military feeding. The success of a commercial product may depend on the preferences of a loyal minority, but military rations must take into account those of the entire population of servicemen. The final criterion of the acceptability of foods must be that of consumption, but there are techniques of assessing acceptability besides the obviously valid one of recording eating behavior in the normal situation. The most common, the most efficient, and probably the most reliable is to measure the verbally expressed affective responses of a sample of consumers, and from these measurements establish the positions of various food items on some continuum from which acceptance behavior may be inferred.

Before this method can be used effectively certain questions must be answered. What task should be set for the consumer subjects? What kind of experimental situations will call forth responses which are valid for predicting acceptance? Such problems are particularly important to the Armed Services. Military consumers are a fairly homogeneous group, but the conditions under which rations are used vary widely. Many military feeding situations are totally inaccessible for conducting tests on foods, and others offer varying degrees of difficulty. The various types of pretesting that are used by the Quartermaster Corps may be conveniently classified according to whether testing is done in "artificial" or "natural" situations. The "artificial" situations include (*a*) laboratory testing under controlled conditions using civilian subjects and (*b*), soldier-consumer panel testing at military posts using laboratory-like procedures. The "natural" situations include (*a*) normal mess-hall feeding, (*b*) planned test exercises where rations are used by selected groups of soldiers, and (*c*) regular field maneuvers where rations are used under nontest conditions.

The relative value of these approaches will depend upon the criteria by which they are judged. If one demands experimental control, or is particularly concerned about economy of testing, the "artificial" situations have the advantage. However, if attention is concentrated primarily on test validity, the "natural" situations are superior, since one is entitled to assume that results become more valid as the test situation more closely approximates the actual conditions of consumption, granting, of course, that the test population is always a good sample of the population of interest. The laboratory method is the one most used by the Quartermaster Corps. Important decisions as to the selection or rejection of items are frequently made on the basis of laboratory results alone. However, there has been a tendency to distrust laboratory results and to require additional testing in the field. It became apparent that the lack of knowledge of the true value of the various types of pretesting, and of relationships among them, was retarding the ration development program and making it unduly expensive. The experiment reported here represented the initial phase of a program of research planned to remedy the situation.

The test subject variable was selected for first investigation since it represented one of the most obvious differences between laboratory tests and any field test conducted with

[1] This paper reports research undertaken at the Quartermaster Food and Container Institute for the Armed Forces, and has been assigned number 483 in the series of papers approved for publication. The views or conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the views or indorsement of the Department of Defense.

[2] Present address: Armour Research Foundation, Chicago, Illinois.

service personnel. The problem may be stated as follows: How well do the relative preference ratings of foods by groups of soldiers correspond with ratings by groups of civilians when the test situations are made to correspond as closely as possible? Referring to the classification scheme above, this represented comparison of the laboratory and soldier-consumer panel situations.

Procedure

The laboratory tests were conducted at the Quartermaster Food and Container Institute in the food acceptance laboratory, which is especially built for running sensory tests on foods. It is secluded, air-conditioned, and comfortable. Test subjects sit in panel booths separated from the food preparation room. Soldier-consumer panel tests were run at Fort Lee, Virginia, in a dining hall which was made available between regular meals for that purpose.

The 12 test foods (see Table 3) were selected so that distinctly different types would be represented and so that their preference ratings, as established in previous laboratory tests, would cover a wide range. Food materials for the two locations were drawn from a common source and methods of preparation were controlled to assure identity. Other controllable physical factors, such as the holding time before serving and the size of samples, were standardized at the two test locations. Time of testing in relation to regular meal-times was made comparable.

Preference was measured by means of a nine-interval rating scale, commonly known as the "hedonic scale," which was developed at the Quartermaster Food and Container Institute in 1949 (3) and has been used extensively with satisfactory results (2, 4). The questionnaire used at Fort Lee was headed by these instructions:

> We want to find out how well certain foods are liked by Army men. You will be served three samples of food, one after another. As soon as you finish each, show how much you liked or disliked it by marking on the scale underneath the name of that food. Then have a drink of water and wait for the next sample. Please do not talk about the foods during the test. It is important to have each man give his own answers—peoples' likes and dislikes are expected to be different.

Three vertically oriented scales were arranged across the page below the instructions. Each was about five inches long with nine equally spaced intervals labeled with the following phrases, reading from top to bottom: "like extremely," "like very much," "like moderately," "like slightly," "neither like nor dislike," "dislike slightly," "dislike moderately," "dislike very much," "dislike extremely." The appropriate food name was rubber-stamped above each scale prior to testing. The form used at the Institute was identical except that the instructions were omitted since most of the Institute subjects were already familiar with the method. New subjects were given oral instructions. None of the Fort Lee subjects had ever participated in a test of this kind before.

It was not feasible to serve all 12 foods to one person at a single sitting. Experience has shown that if the number of foods is not strictly limited, the ratings of those served later may be affected, usually showing a decrement. Therefore, only three foods were presented to each subject in each test session, so that four sessions were required to test one replication of the 12 foods. Combinations of foods were established for four replications in such a way that no two foods appeared together more than once. Four replicates were run at Fort Lee, but only replicates 1 and 2 at the Institute.

Forty persons participated in each session at the Institute. They were selected each time from a pool of approximately 600 employees of the Chicago Quartermaster Depot by a standard procedure designed to obtain widespread participation. Most test subjects participated in only one session and none in more than two. Fifty soldiers participated in each Fort Lee session, each group being drawn from a different company. Selection within companies was not random since a small number of men were always unavailable for administrative reasons. Thus, there was not strict assurance that the groups were representative of the Army; on the other hand, no reason was known why their food preferences should have differed from those of the Army in general.

The Institute test subjects came to the laboratory in small groups. Each was given a questionnaire, with additional verbal instructions for those people who were new. They were seated in the panel booths and the three test items were presented one at a time in random order. At Fort Lee all 50 men were brought into the dining hall at the same time. They were seated two at a table, where places had been prepared with questionnaires, water, and necessary utensils, and were briefed by the test monitor before beginning the test. Again, the food items were served one at a time in random order.

Results and Discussion

The index of preference used here was that derived by assigning the values 1 to 9 to the scale categories, beginning at the "dislike extremely" end, and taking the mean of the resulting distribution of values. The mean rating and standard deviation were obtained for each food in each replicate. Thus there were six sets of ratings—four from Fort Lee and two from the Institute.

Product-moment correlations between sets of ratings for the 12 foods were obtained for all possible pairings of individual laboratory and field replicates and also between sets of

Table 1

Correlations Between Field and Laboratory Mean Ratings for Single Replicates and Mean Ratings Based on Combined Replicates

(N = 50 for field replicate. N = 40 for laboratory replicate. All correlations positive)

	Field	Laboratory Replicate 1	Laboratory Replicate 2	Laboratory Replicates 1 & 2
Single replicate	1	.88	.88	.90
	2	.92	.92	.95
	3	.84	.78	.87
	4	.80	.82	.83
Combined replicates	1 & 2	.91	.89	.92
	3 & 4	.99	.85	.86
	1, 2, 3 & 4	.86	.86	.92

Note.—Average (Fisher's hyperbolic arc-tangent transformation method) of 8 correlations between single replicates is .86.

ratings obtained by combining ratings from the individual replicates. These correlations, which are predictive validity coefficients in light of the purpose of the experiment, are shown in Table 1. (All of the correlations are positive.) Minimum validity is represented by the eight correlations grouped together in the upper lefthand corner of the table which were derived from the sets of ratings from single replicates. The remaining correlations all involve combinations of ratings from more than one replicate and demonstrate the expected improvement with increased length of test.

The correlation between averages of the 200 field ratings and 80 laboratory ratings was + .92. An equation expressing the relationship may be written as follows:

$$Y \text{ (field)} = 1.23\ X \text{ (lab)} - 2.30.$$

The assumption of linearity may be an oversimplification, subject to change on the basis of more extensive investigation; however, it seemed most appropriate for the present data. A scatter diagram of the data did not justify any other assumption.

The above equation suggests that the two groups of subjects were responding differently in ways that affected both level of rating and units of discrimination. The grand mean over all foods for the laboratory was 6.43 as compared to 5.61 for the field, while the respective ranges of means were 4.81 and 5.82 (Table 3). It is apparent that the soldiers responded to the low preference foods with more frequent and intense "dislike." The soldiers' comments written on the questionnaires gave further evidence of this tendency to respond more strongly and with fewer inhibitions than the typical laboratory subject, and suggested the possibility of differences in attitude toward the test situation as well as differences in attitude toward the foods. In spite of this, however, the high correlation shows that differences between foods produced differences in evaluation behavior that were proportional for the two groups of subjects.

Although secondary to validity in this experiment, test reliability was also considered. The "intralocation" correlations between sets of ratings provided a single estimate of laboratory reliability and six estimates of field reliability (Table 2). It was expected a priori that the laboratory results would be more reliable because of better control in the laboratory situation. The laboratory correlation was .84 while the average field correlation was .93; however, only one intralaboratory correlation was obtained and this figure may not have been generally representative. The Spearman-Brown prophecy formula (1) shows that to obtain a reliability comparable to that in the field, the number of laboratory subjects would have to be increased only from 40 to 120, i.e., considerably fewer than the 200 actually used in the field.

Another aspect of reliability is presented in

Table 2

Intercorrelations Among the Four Field Replicates

(N = 50. All correlations are positive)

	Replicate 1	Replicate 2	Replicate 3
Replicate 2	.95	—	—
Replicate 3	.94	.92	—
Replicate 4	.96	.80	.91

Note.—Average (Fisher's hyperbolic arc-tangent transformation method) is .93.

Table 3

Mean Preference Ratings and Standard Errors of the Mean for Laboratory and Field

	Laboratory*			Field†		
	Actual		Projected			
Food	$\bar{X}$	SE_m	SE_m‡	$\bar{X}$	SE_m	Differences§
Peaches	8.43	.066	.042	8.23	.047	.005
Salmon	8.08	.091	.058	7.12	.111	.053
Corn	7.14	.144	.091	7.38	.086	−.005
Corned beef	7.02	.159	.100	6.02	.144	.044
Ham and eggs	6.88	.175	.111	6.40	.140	.029
Bread	6.75	.154	.097	6.50	.139	.042
Carrots	6.66	.177	.088	4.54	.164	.076
Sauerkraut	6.31	.236	.149	6.18	.151	.002
Cheese bar	5.69	.223	.141	3.48	.156	.015
Milk	5.46	.186	.118	4.68	.178	.060
Cabbage	5.15	.252	.159	4.26	.168	.009
Meat bar	3.62	.230	.145	2.46	.130	−.015
Grand mean	6.43			5.61		
Range of means	4.81			5.82		

* Combined data for two laboratory replicates, $N = 80$.
† Combined data for four field replicates, $N = 200$.
‡ Projected to $N = 200$, assuming no change in variance.
§ Field SE_m minus projected laboratory SE_m.

Table 3 which gives the standard error of the mean (SE_m) for each food. Two figures are shown for the laboratory. Column 2 gives the actual value obtained from the distribution of 80 laboratory ratings and Column 3 projects this figure to $N = 200$, assuming no change in variance. For 10 of the 12 foods the projected laboratory SE_m is lower than the field SE_m, which indicates that a laboratory retest should reproduce its numerical indices more accurately than a field retest. This further suggests that the field reliability coefficients were higher because of the larger N and the greater range of the scale utilized in the field and not because the rating of each food was more precisely located on the scale.

Since the results reported here were based on the testing of only a small number of foods selected from the hundreds of items which may be of concern in military feeding, the possible effects of selection bias should be considered. The fact that the foods were not randomly selected detracts from the general applicability of the findings. Consumers tend to like, rather than dislike, the great majority of items that are available for use in military as well as civilian feeding. In the present experiment the test foods were selected to cover a wide range of preference; hence, there was considerably more loading with low preference foods than would have been the case had the items been randomly selected. At the same time, use of the wider range of the scale should have improved the correlation. However, another factor in the present experiment would have tended to lower, rather than raise the correlation, if we may assume that the probability of finding differences between laboratory and field would increase as the group of test foods became more heterogeneous. The attempt was made to maximize heterogeneity by selecting foods to cover a wide range of food types so that there was greater opportunity for differences to appear than would be the case with random selection of test foods.

Some further limitations on the significance of these results for the ration-testing program should be noted. First, neither group of test subjects was a random sample of a well-defined population. They were merely typical

of what might be expected on a continuing basis in the two test situations. Further, only certain ones of the many possible sources of variation between types of pretests were allowed to operate, e.g., quite a number of factors which would affect preferences in normal mess-hall feeding may have been entirely disregarded. However, this was deliberately accepted in designing the experiment. The intent was to compare two practical test situations where the two types of subjects could be reached, controlling only in regard to those factors which could be considered incidental, such as the rating scale, the number and combination of samples, and the food materials and their preparation and serving. Test subjects, test location, and certain conditions inseparable from test location varied independently. Under these conditions, representing what is normally attainable, good correlation was established. This both supports the inference that soldiers' food preferences are the same as those of the civilian population and demonstrates the practical equivalence of the two test procedures. The "intra-" and "inter-situation" correlations were of the same order, which suggests that any noncorrespondence between test results is just as likely to have been due to unreliability of the basic method as to differences between the subjects or the situations.

These results have very satisfactory implications for the methods of food acceptance evaluation currently being used by the Quartermaster Corps. It has been shown that laboratory ratings for a series of foods will accurately predict relative preferences as established by the soldier-consumer panel method. The fact that the validity of neither method for predicting actual food acceptance has been established does not detract from the importance of the finding. It represents significant progress toward rationalization and integration of methods for the pretesting of rations and serves as a sound basis for eliminating much expensive and unnecessary field testing.

Conclusions

The primary conclusion was a practical and specific one, namely, that pretesting of rations in the Institute laboratory may be considered equivalent to pretesting by the soldier-consumer panel method. Corollary conclusions were: (*a*) both laboratory and field preference ratings have satisfactory reliability and (*b*) the hedonic scale method is adequate for evaluating food preferences under varying conditions.

Received March 20, 1956.

References

1. Gulliksen, H. *Theory of mental tests.* New York: Wiley, 1950.
2. Peryam, D. R. Field testing of Armed Forces rations. In D. R. Peryam, F. J. Pilgrim, & M. S. Peterson (Eds.), *Food acceptance testing methodology, a symposium.* Washington, D. C.: National Research Council, Advisory Board on Quartermaster Research & Development, 1954. Pp. 75–85.
3. Peryam, D. R., & Girardot, N. F. Advanced taste test method. *Food Engng,* 1952, 194, 48–61.
4. Wood, K. R., & Peryam, D. R. Preliminary analysis of five food preference surveys. *Food Technol.,* 1953, 6, 248–249.