Reprinted from FOOD TECHNOLOGY, Vol. 12, No. 1, Pages 30-33

Variation in Preference Ratings for Foods Served at Meals*

David R. Peryam and
Norman J. Gutman

Quartermaster Food and Container Institute for The Armed Forces, Chicago 9, Illinois

(Manuscript received June 21, 1955)

COMPETITION among the food industries has brought about increasing awareness of the importance of consumer likes and dislikes, but there has been considerable lag in the development of adequate methods of measuring these variables and relating them to consumer acceptance. It is generally agreed that the prediction of acceptance requires the measurement of preferences, but there is less concurrence on how to answer such important questions as "Whose preferences?", "In what situations?", and "Under what conditions of control?" The answers are often determined arbitrarily rather than through analysis of the particular problem under consideration.

In many cases this is due to lack of knowledge about the situations which are utilized for testing, and about how changes in test conditions may affect results. Expressed preferences may sometimes appear highly variable, but they are also characterized by definite elements of stability. It is reasonable to assume that much of the variation is systematic, and therefore can be measured or accounted for in test design. The present experiment was designed to test this assumption by obtaining information about the factors affecting food preferences in a particular environmental situation. The specific findings will be applicable only under the same circumstances, but it is believed that the approach has general validity.

The situation investigated was that represented by a normal meal in Army mess-halls—a field situation as opposed to controlled laboratory preference tests. Preference results obtained here should be more valid as predictors of actual food acceptance than the results of laboratory tests simply because it represents normal feeding and, potentially at least, more of the conditions that might influence actual acceptance are allowed to operate without restrictions. On the other hand, one might expect results to be less reliable because of this very lack of control. The normal meal situation can play an important role in Quartermaster Corps food research, since it may be used as the second stage, following laboratory preference tests, in the pretesting of new food items for acceptability. Therefore it was important to learn its characteristics as an aid in test design and data interpretation.

The experiment had the general objectives of estimating the extent of variation in preference for a constant food that might be encountered in repeated testing and locating the sources of the variation. It was necessary to be selective in regard to the sources, since preference might be affected by a great many factors, and only a few could be included in one experiment. The following factors were investigated: (1) mess-hall, or, to be more exact, subject group, since the men constantly eat at the same mess hall, (2) meal (breakfast, dinner, or supper), (3) day of the week (Monday, Wednesday, or Friday), (4) successive weeks during the experiment, and (5) questionnaire form (whether one, two or four foods were to be rated).

PROCEDURE

A single test food was served repeatedly at regular meals in nine mess halls at Fort Lee, Va. in late August 1953, the 5 factors being varied according to a pre-planned pattern. Each time it was served, preference ratings were obtained from a sample of the men eating at the meal.

The test food was orange juice of reasonably good quality, and was all prepared from a single production lot of dehydrated juice using a uniform procedure. It was served chilled (about 50° F.) either in pitchers on the tables or on the serving line, according to normal practice in the particular mess. The temperature increased about 10° F. during the course of a meal but the change was fairly uniform between messes.

Nine messes, each of which fed one Company of about 200 men, were randomly selected to participate. The type of duty which the men were engaged in was uniform both among Companies and throughout the period of the test. It was also rather strenuous since practically all of the men in each Company were basic trainees. The testing was done by military test teams from the Quartermaster Research and Engineering Command Field Evaluation Agency at Fort Lee, Va. Respondents were obtained by randomly selecting 30 of the men served at each test meal. While a selected respondent was eating, the test monitor oriented him by means of a simple, uniform talk, then gave him a questionnaire and instructed him in its use. The test monitor remained available to answer questions.

The questionnaires which were used were similar to the laboratory form, as described by Peryam and Pilgrim *(2)*, but also provided the subject the opportunity to check *Did not try* as an alternative to rating any particular food. There were 3 different forms which included scales for rating 1, 2, or 4 foods, respectively. The food names were printed at the head of the scales prior to testing. Orange juice appeared on all forms; the other item(s) to be rated were selected by the test team from the menus for the test meals.

The design used was a 1/27th replicate as described by Kempthorne *(1)*. Its operational detail is shown in Table 1 which gives the combination of factors for each session. It

*This paper reports research undertaken at the Quartermaster Food and Container Institute for the Armed Forces, and has been assigned number 643 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 endorsement of the Department of the Army.

will be noted that each of the 9 messes was tested 3 times, and that these tests were on 3 different days of the week, one at each of the 3 daily meals, and with 3 different questionnaire forms. Also, each mess was tested once in each of the three weeks of the study.

The mess variable requires further explanation. The population of Company-size messes available at Fort Lee was determined and 9 of these were selected randomly. Two pseudo-variables, U_1 and U_2, were established, each at 3 levels, which together define the mess variable. The 9 messes were randomly assigned to the cells developed from the 3 levels of U_1 and U_2. Once this was done, the combinations of day, form, meal, and week of testing were determined by the confounding relation defining the design, which, however, is not shown. The analysis of variance table indicates the confounding of main effects and interactions with other main effects and 2-factor interactions (considering mess as one factor). Confounding with 3 or more factors is not shown.

RESULTS AND DISCUSSION

Table 1 presents the basic data in the form of the mean preference rating for each session, obtained by assigning the values 1–9 to the scale categories beginning at the *dislike* end *(2)*. *N's* less than 30 were due either to spoiled questionnaires or to a subject's not having tried the test food. The 27 means show considerable variability, ranging from 5.79 to 8.29. Statistically, this range of 2.50 scale points is very significant, despite the relatively small number of subjects at each test session. Evidence from other experiments *(3)* shows that differences of this order mean much more than just preference trends; usually they may be taken as indicating important differences in objective food behavior. In Table 2 the data are grouped in different ways to show the effect of each category of each variable.

Table 3 presents the analysis of variance for the entire experiment. An analysis of variance using all of the data would have been arithmetically very difficult because of the unequal number of ratings at various sessions. Therefore, for purposes of this analysis, ratings were discarded from various sessions in such a way that there were an equal number for all sessions while the mean and standard deviation remained invariant.

The effects of day, week, and questionnaire form were clearly insignificant. Little more can be said about them except to qualify the findings by taking into account the limited ranges of the variables. The day variable did cover the entire range which is practical for pretesting. Food tests are never run on week-ends simply because so many conditions may change at that time. The variation in number of foods included on the questionnaire was a minor, though practical, one. It is worth while to have demonstrated that stated preferences will remain constant in spite of this difference in the rating task; however, we obviously cannot conclude that other changes such as, e.g., changes in instructions or description of the scale categories would have no effect. Similarly, there was practical value in showing that no effect could be attributed to the general time factor (week) over a 3-week period, since most pretests are short range affairs. However, logic warns us against the general conclusion that we need never expect such effects. There may be changes in the weather, long-range seasonal changes, changes in personnel or their type of duty, or changes in morale. We can suppose

TABLE 1

Preference ratings for orange juice obtained at meals over a period of three successive weeks

Session	Mess	Week	Day[1]	Meal[2]	Questionnaire[3]	N	Mean rating
1	A	1	M	B	1	24	7.00
14	A	2	W	D	4	29	7.07
27	A	3	F	S	2	28	8.29
7	B	1	F	B	4	29	7.38
11	B	2	M	D	2	29	7.21
24	B	3	W	S	1	29	7.07
4	E	1	W	B	2	32	7.28
17	E	2	F	D	1	27	6.89
21	E	3	M	S	4	29	7.55
2	F	1	M	D	1	29	7.00
15	F	2	W	S	4	27	7.00
25	F	3	F	B	2	31	6.06
5	G	1	W	D	2	28	5.79
18	G	2	F	S	1	29	6.69
19	G	3	M	B	4	21	6.95
9	I	1	F	S	4	29	7.38
10	I	2	M	B	2	27	6.37
23	I	3	W	D	1	26	7.42
3	K	1	M	S	1	29	7.76
13	K	2	W	B	4	24	6.50
26	K	3	F	D	2	29	7.10
6	L	1	W	S	2	29	7.45
16	L	2	F	B	1	27	7.30
20	L	3	M	D	4	28	7.14
8	M	1	F	D	4	29	8.24
12	M	2	M	S	2	30	8.17
22	M	3	W	B	1	26	7.19
			Grand mean—all sessions				7.16
			Range of session means				2.50

[1] Monday, Wednesday, Friday.
[2] Breakfast, Dinner, Supper.
[3] Number of foods appearing on the questionnaire for the subject to rate.

TABLE 2

Preference ratings for orange juice obtained by grouping test meals for each category of each variable

Variable	Category	No. of test sessions	No. of ratings	Mean rating	Range of means
Day	Monday	9	246	7.26	0.28
	Wednesday	9	250	6.98	
	Friday	9	258	7.24	
Questionnaire	1 food	9	246	7.13	0.19
	2 foods	9	263	7.08	
	4 foods	9	245	7.27	
Week	First	9	258	7.26	0.23
	Second	9	249	7.03	
	Third	9	247	7.16	
Meal	Breakfast	9	241	6.89	0.58
	Dinner	9	254	7.10	
	Supper	9	259	7.47	
Mess	A	3	81	7.47	1.49
	B	3	87	7.22	
	E	3	88	7.25	
	F	3	87	6.67	
	G	3	78	6.40	
	I	3	82	7.06	
	K	3	82	7.16	
	L	3	84	7.30	
	M	3	85	7.89	

TABLE 3

Analysis of variance based on preference ratings for orange juice obtained at 27 different meals

Variable	Degrees of freedom	Sum of squared deviations	Mean square	F-ratio
Meal ($= FU_2$)	2	32.91	16.46	4.45[1]
Day ($= FU_1 = WF^2 = WU_1 = MU^2_1U_2$)	2	9.54	4.77	1.29
Form ($= DU_1 = MU_2 = WD = WU^2_1$)	2	2.65	1.32	<1
Week ($= DF^2 = DU^2_1 = FU^2_1 = MU_1U_2$)	2	6.40	3.20	<1
Mess	8	86.13	10.77	2.91[2]
MF^2 ($= MU^2_2 = FU^2_2$)	2	9.91	4.95	1.34
MD ($WU_2 = MU_1U^2_2$)	2	.35	.18	<1
MU_1 ($= WU^2_2$)	2	2.98	1.49	<1
MU^2_1 (DU^2_2)	2	36.09	18.04	4.88[2]
$DU_2 = MU^2_1U^2_2 = MW$)	2	1.91	0.95	<1
Within Cell (error)	540	1998.10	3.70	
Total	566[3]	2186.96		

[1] Significant at 5% level.
[2] Significant at 1% level.
[3] N's of all test sessions were reduced to 21 for purposes of this analysis.
F—Questionnaire form; M—Meal; D—Day; W—Week; U_1, U_2—Components of mess.

that the present negative result was dependent on maintaining a temporary stability in regard to such relevant factors.

The effects of mess and meal were both significant. Mess had the greatest absolute effect, the means varying over a range of 1.49 scale points (see Table 2), and was significant at the 1% level. Such group differences seem to be typical of preference testing. Similar effects can nearly always be found whether one considers laboratory preference and trained panel tests or various field situations. On the whole, group effects are probably the most troublesome factor that must be dealt with in planning tests and interpreting results, since their causes are not precisely known. Much of this effect, perhaps all of it in laboratory testing, may be due to sampling variations within a population where preference itself is quite variable. However, additional factors are probably involved in the present situation. Each mess has a different group of food service personnel with varying skills and motivations, so that there will be opportunities for differences in food preparation and serving. The individuals constituting each group eat together and are usually together in many other situations. Also there may be basic differences among the groups of men. For example, they may be assigned to different types of duty or attitudes may develop in different patterns for other reasons. Such factors could cause physiological or psychological differences between groups which could affect food preferences. Even though a relatively large and significant amount of variation in preference was attributable to mess in this experiment, it is possible that this represents only a part of the potential range, since the number of messes was small, they were all at the same post, and all of the men were assigned to the same type of duty.

The variation due to meal is significant at the 5% level and, again, the range of 0.58 of a scale point may be considered important. One might expect differences in preference for many foods depending on the meal at which they are served, because custom dictates that certain foods are more appropriate for some meals than for others; however, this does not explain the present results. Orange juice would be considered as more appropriate for breakfast than for any other meal, yet here the breakfast mean lies at the low end of the range. Our search for explanations has to be speculative. The reason may have been the novelty of being served the drink at an unexpected time. Perhaps a better reason lies in the fact that the study was run during hot weather and most of the men were engaged in strenuous basic training activities. This suggests the explanation that the cool beverage had a greater refreshment value in the evening because the men were more thirsty after work. Preference was *not* found to be significantly correlated with temperature or humidity as measured at meal times or with daily average temperature or humidity; however, no data were available on relative degrees of activity so that a test of this factor was not possible.

The only interaction which proved to be significant was that between meal and one of the components of mess (MU^2_1), which is confounded with the interaction of day and the other component of mess (DU^2_2). It is impossible to make any specific interpretation of the finding on the basis of the present experiment.

This experiment was concerned with only one food, and one cannot assume that the pattern of variation established would be typical of all items. For example, it is possible that preference for many foods would remain constant from one meal to another, and that preference for a limited number of foods would vary during the week because of customs of serving. Here we did not investigate even a small sampling of foods, but had to be content with intensive examination of only one; however, since variation was found with orange juice, it would seem likely to occur with other foods.

The fact that one must expect such variation does not mean that the normal meal situation cannot be used for pretesting. Pretesting is concerned more often with establishing relative preference values within a group of similar items than it is with establishing absolute levels of acceptance. Typical problems might be: comparing dehydrated with fresh orange juice, comparing dehydrated potatoes made by different processes, or comparing a number of canned meat items. It is reasonable to assume that preference for similar items will be affected in the same way as elements in the test situation are changed, e.g., as the test moves from one mess to another, from dinner to supper, or from one week to the next. Whenever this is true the normal meal situation can safely be utilized to establish relative preference values simply by designing the test so that all items are subjected to the same set of conditions.

SUMMARY

A single food was tested for preference at regular meals in military mess-halls over a period of three weeks. The total range of variation was large, a result which suggests that a preference rating established in a single test of this type may be quite unreliable. The test design obtained information about the effect of five different factors on preference. Of these, the effects of mess hall (subject group), and of meal of the day were significant. The effects of questionnaire form, day of the week, and week itself were not significant within the ranges investigated.

This information has been useful in guiding the design of pretests utilizing the normal meal situation. It is suggested that the same general approach could be used to advantage to explore and evaluate any preference survey method.

LITERATURE CITED

1. KEMPTHORNE, OSCAR. *Design and Analysis of Experiments.* 1952. Wiley, New York, N. Y., pp. 390–429.
2. PERYAM, D. R., AND PILGRIM, F. J. The hedonic scale method of measuring food preferences. In *Methodology of Sensory Testing—A Symposium.* 1957. Pages 9–15. The Garrard Press, Champaign, Illinois.
3. SCHUTZ, H. G. Preference ratings as predictors of food consumption. *Am. Psychologist,* 12, 412 (1957).