R 90-50

FOOD QUALITY AND PREFERENCE 2, (1990) 21–27

THE RELATIONSHIP OF RECENT AND RETROSPECTIVE FOOD ACCEPTANCE RATINGS

Barbara A. Jezior

US Army Natick Research Development and Engineering Center, ATTN: STRNC-YBH, Natick, Massachusetts 01760-5020, USA

Larry L. Lesher

Geo-Centers, Inc., 7 Wells Avenue, Newton Centre, Massachusetts 02164, USA

& Richard D. Popper

US Army Natick Research Development and Engineering Center, ATTN: STRNC-YBH, Natick, Massachusetts 01760-5020, USA

(*Received* 28 *September* 1989; *revised and accepted* 2 *April* 1990)

ABSTRACT

This large-scale ($N = 472$), 11-*day field test of acceptability of three military rations had a dual purpose*: (1) *to collect acceptability data at intervals during the test and* (2) *to determine if data collected retrospectively* (24 *h post test*) *adequately indicate ration acceptability. Regression analyses were conducted to determine how well the retrospective ratings predicted the field ratings. Findings showed high* R^2 *values and strong linear relationships.*

Keywords: Consumers food acceptance, retrospective.

INTRODUCTION

In developing a military ration there are many challenging criteria to satisfy. With cost being a constant reference point, military op-

erational needs also impose stringent nutritional, weight, size, packaging, and shelf-life requirements. The ultimate challenge is to satisfy the soldier's palate, given these constraints. If soldiers do not consume the ration in sufficient quantity, their performance and mission can suffer and large sums of money will be wasted. Sound, cost-effective methods are essential for measuring how well a ration is liked, along with the ability to apply these methods in field conditions (simulated combat).

The US military has an almost 40-year history of measuring like or dislike of food items to predict consumption. Hedonic evaluation of rations started in 1950 with Peryam at the Quartermaster Food and Container Institute in Chicago and now continues at the US Army Natick Research, Development and Engineering Center. One of his accomplishments was the development of a 9-point hedonic scale that has been widely used to obtain both food preference and food acceptance ratings (Peryam & Pilgrim, 1957).

Acceptance generally refers to the response to a food item a short time after it has been sampled, while preference as defined by Meiselman (1988) refers to an opinion about a food item in the abstract, i.e. the amalgamation of all experience with that item, and is not obtained in conjunction with any particular sample of the item. Both phenomena are detailed by Meiselman (1988) in his review of food acceptability measures.

It has been Natick's standard practice in field ration tests to obtain food acceptance ratings as one measure of how well a ration is liked. In one recent 11-day test, however, in addition to collecting acceptance measures at intervals throughout the test, we also had soldiers rate the ration items retrospectively, i.e. one day after the test ended. The retrospective ratings were not traditional acceptance ratings since they did not occur at the time the ration items were consumed. They were also not preference ratings, since the soldiers were rating the ration items with reference to a set of experiences with specific formulations.

The question of this investigation is whether the retrospective ratings are sufficient as indicators of acceptance. If so, the cost of field testing could be drastically reduced, and the logistical problems attendant to field tests and their impositions on military missions and training would be obviated.

METHODS

Test protocol

Three US Army infantry companies were fed three versions of an operational field ration for 11 consecutive days. The rations were their only diet, and each company was issued with only one version. One company served as a control and consumed the current configuration of the ration. Each of the other companies consumed versions in which some of the current items were produced to new specifications and some of the items were new additions. The troops consumed the rations while participating in a large-scale military exercise and no commercial food items were allowed to be consumed along with the ration items.

Each version of the ration offered 12 menus, and each menu packet contained all the components necessary for one meal. While no entree was repeated within any ration's menu cycle, some of the other meal components were, and some components, such as coffee and crackers, were common to every menu. Each version of the ration offered a total of no less than 35 different food items with a total of 107 food items across the three rations. Most items fell into the following categories: main dishes, fruits, desserts, spreads, starches, and beverages. Each soldier received three meals (menu packets) per day, which he could eat when and how he chose. He was also free to trade items. Since the test lasted 11 days the soldiers were issued each menu more than once.

The test population was entirely male ($N = 472$), average age 23. Racial composition was 65% white, 20% black, the remaining 15% representing various racial/ethnic groups. Over 96% were enlisted personnel with two to three years' service. The demographic balance was similar across companies and the companies were of almost equal size.

The test protocol called for a number of physiological and food-related measures to be obtained at intervals, among which were self-reported consumption levels for ration items and acceptance ratings. Evidence that a food item had been consumed during the test was based on the self-report data on days 1, 3, 6, 7, and 11, which was corroborated by examination of food waste. On days 3, 7, and 11 the soldiers were also asked to rate the food items they consumed on those days. Typically the soldiers rated the items at meal time, but any delay between consumption and rating could not exceed 24 h, since the data were collected within that time period. On day 12, the day after the test ended, the soldiers were administered a final questionnaire which, in addition to general questions about the ration, also asked them to rate all ration items consumed during the field exercise. Here the time interval between consumption and rating could range from 24 h for the items consumed the previous day to a number of days (up to 11), depending on when a particular item was last consumed. For any item not consumed during the test, the soldiers were asked to check 'never tried' on the rating form. The rating scale used was Peryam's 9-point hedonic scale (9—like extremely, 1—dislike extremely). A separate rating form and final questionnaire were constructed for each version of the ration, but all questions and formats were kept as uniform as possible. A complete description of the test population and procedures can be found in Popper *et al.* (1987).

Analysis

The analysis was guided by the following considerations:

(1) The final questionnaire gave each of the soldiers an opportunity to rate all ration items, whereas the field acceptance ratings did not. In the field, each soldier rated only those items he ate on the three days of the test on which acceptance data were collected. While this situation did not allow for each ration item to be rated by each soldier in the field, the large sample of soldiers and the relatively small number of food items in the ration nonetheless provided an adequate number of field ratings for each ration item.
(2) The final questionnaire left open the possibility that soldiers would rate items they had not consumed during the test even though specifically requested not to do so.
(3) Consumption data were available for each soldier for eight of the 11 test days, on five of which there were no corresponding field acceptance ratings.

Based on these considerations, we conducted three regressions to examine the relationships between the average field and final questionnaire ratings for each food item. The three regressions differed with respect to the observations on which the average ratings were based. While there may be some concern with using parametric regression statistics with acceptance ratings, analyses of the relationships between field and final questionnaire ratings with nonparametric statistics, i.e. Spearman's Rho, showed the same trends we will show with parametric statistics. Furthermore, Traylor (1983) and Labovitz (1970, 1971) have argued that applying parametric statistics to ordinal or quasi-interoal data is justified unless the scoring system approaches gross, non-linear distortions such as found in dichotomous data.

The data set for each regression can be described as follows:

Regression 1. The entire original data set.
Regression 2. Any final questionnaire rating a soldier gave for any food item was deleted from the data set if we had no evidence he consumed the item during the test from either the field acceptance data set or the food consumption data set. The final ques-

tionnaire ratings thus reflect his opinion of items which we know he ate in the field, some of which he rated in the field and some he did not.

Regression 3. All ratings a soldier made on the final questionnaire were deleted from the data set used in regression 2 if we did not have corresponding field acceptance ratings. This data set can be considered paired data since for all items a soldier rated on the final questionnaire there are corresponding field ratings by that soldier.

RESULTS

The first regression analysis, using all the data, shows a significant linear relationship between all field and retrospective (final questionnaire) ratings. Figure 1 shows a scatterplot of the average ratings for all ration items. It can be seen that almost all the data points lie above a straight line describing a perfect linear relationship (slope = 1, intercept = 0) and tend to cluster between 5 and 8 on the hedonic scale. Despite this limited range of values, Table 1 reveals a strong, statistically significant linear relationship between the two types of ratings, with 84% of the variance in the field ratings explained by the final questionnaire ratings. The best fit (least squares) line fitted to all the data is

$$f = 2{\cdot}91 + 0{\cdot}66fq$$

where f is the field rating, and fq is the final questionnaire rating. Figure 1 also illustrates a strong tendency for items to be rated lower on the final questionnaire than in the field, especially if the item receives a low rating in the field (for example, less than 5). A number of these lower rated items show a discrepancy of larger than two scale points between the field and final ratings.

The scatterplot and regression statistics obtained in the second regression, when final questionnaire ratings for items were dropped if there was no evidence of these items being consumed in the field, are displayed in Fig. 2 and Table 2. Analysis of these data results in the regression line

$$f = 1{\cdot}95 + 0{\cdot}76fq$$

Again, a significant liner relationship between the field and final ratings was obtained, with an R^2 that accounts for 88% of the variance in field ratings. The slope and intercept are still significantly different from those of the ideal relationship of intercept = 0, slope = 1 (B_0; $t = 10{\cdot}27$, $p \leqslant 0{\cdot}001$; B_1; $t = 8{\cdot}66$, $p \leqslant 0{\cdot}001$), but Fig. 2 reveals a perceptible decrease in the discrepancy between in the two sets of rating when compared with Fig. 1. In particular, a shift in the data points for the less preferred items toward the ideal line can be seen. Significance tests for differences (Kleinbaum & Kupper, 1978) in the slope and intercept between the first and second regression compared to the values in Table 1 are significant (for B_0: $t = 3{\cdot}64$, $p \leqslant 0{\cdot}01$; for B_1: $t = 2{\cdot}65$, $p \leqslant 0{\cdot}01$). However, the R^2 were not significantly different (Fisher $Z = -1{\cdot}03$).

The third regression, conducted on paired data (see Methods), resulted in the Fig. 3 scatterplot and the regression statistics of Table 3.

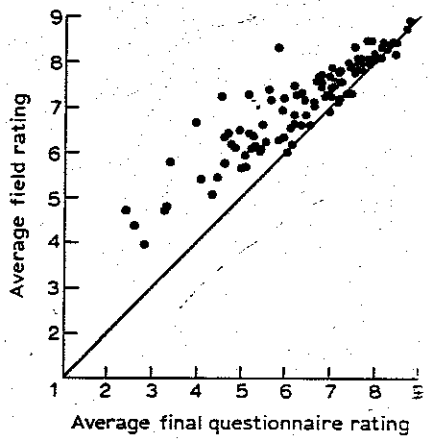

FIG. 1. Scatterplot showing the average ratings from the field and final questionnaires for the entire data set. The straight line depicts the perfect linear relationship.

TABLE 1. Regression Statistics Obtained for all Final Versus Field Ration Item Ratings[a]

R^2	f	df	p	Intercept (B_0)	p	Slope (B_1)	p
0·84	558·61	1105	$\leqslant 0{\cdot}001$	2·91	$\leqslant 0{\cdot}001$	0·66	$\leqslant 0{\cdot}001$

[a] Based on an average of 103 ratings per food item on the final questionnaire and 54 per item in the field for 107 food items.

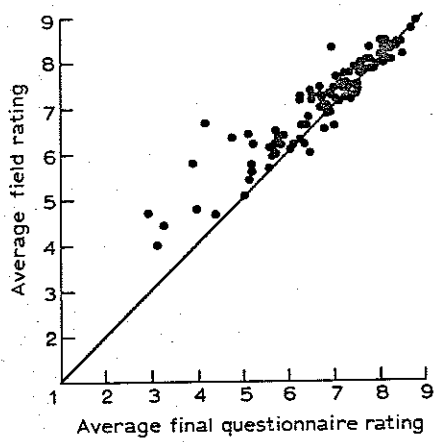

FIG. 2. Scatterplot showing the average ratings from the field and final questionnaires when ratings of foods deleted are for which there was no evidence of consumption. The straight line depicts the perfect linear relationship.

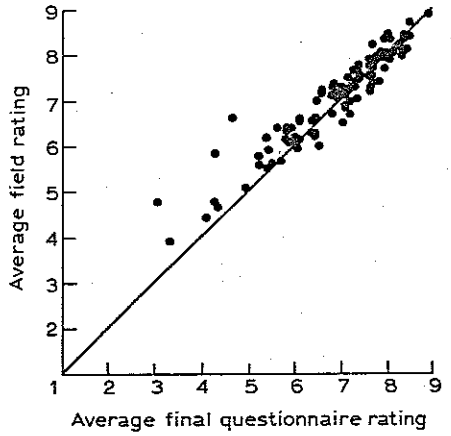

FIG. 3. Scatterplot showing the average ratings from the field and final questionnaires based on paired data. The straight line depicts the perfect liner relationship.

This analysis yields an R_2 of 0·90, and the regression line shows a further decrease in the intercept and a slight rise in the slope compared to the previous two regressions as the least squares line approaches that of the line of slope = 1 and intercept = 0

$$f = 1{\cdot}55 + 0{\cdot}80fq$$

The slope and intercept of this straight line are still significantly different from those of the ideal relationship of $B_0 = 0$, $B_1 = 1$ (B_0; $t = 8{\cdot}59$, $p \leqslant 0{\cdot}001$: B_1; $t = 7{\cdot}83$, $p \leqslant 0{\cdot}001$), but represent a significant shift in the direction of that ideal relationship when tested against the straight line of the first regression (for B_0: $t = 3{\cdot}73$, $p \leqslant 0{\cdot}01$; for B_1: $t = 5{\cdot}20$, $p \leqslant 0{\cdot}01$). The difference in R^2 between the two regressions (Fisher $Z = -1{\cdot}89$) is, however, not significant. The second and third regressions did differ significantly for any parameter (B_0; $t = 1{\cdot}40$, $p \leqslant 0{\cdot}16$; B_1; $t = 1{\cdot}03$, $p \leqslant 0{\cdot}30$).

TABLE 2. Regression Statistics for Final Versus Field Ratings with Evidence of Consumption[a]

R^2	f	df	p	Intercept (B_0)	p	Slope (B_1)	p
0·88	759·26	1105	$\leqslant 0{\cdot}001$	1·95	$\leqslant 0{\cdot}001$	0·76	$\leqslant 0{\cdot}001$

[a] Based on an average of 80 ratings per food item on the final questionnaire and 54 per item in the field for 107 food items.

TABLE 3. Regression Statistics for Final Versus Field Ratings when Consumed Items were Rated in the Field and Final Questionnaire[a]

R^2	f	df	p	Intercept (B_0)	p	Slope (B_1)	p
0·90	978·37	1105	$\leqslant 0{\cdot}001$	1·55	$\leqslant 0{\cdot}001$	0·80	$\leqslant 0{\cdot}001$

[a] Based on an average of 53 ratings per food item on the final questionnaire and 53 per item in the field for 107 food items.

DISCUSSION

The preceding analyses show that the final questionnaire ratings predict the field ratings well, as shown by the high R_2 values in each of the three regressions. The results also show that the relationship between the field and final ratings improves with each more restrictive analysis condition. The figures reveal that for all three data sets discrepancies between the two sets of ratings were almost all in the same direction: the final questionnaire (retrospective) ratings were lower than the field ratings. In those cases, where the change was in the opposite direction, the differences tended to be small and of no practical importance.

Furthermore, the discrepancies are most pronounced for the less-liked items. While the discrepancies are progressively reduced over the three analyses, they narrow more markedly for the less-liked items. This phenomenon is most clearly seen with the second regression, in which the ratings for food items were dropped if there was no evidence of their consumption. By way of example, our first regression line predicts a field rating of 5·6 when the final questionnaire rating is a 4·0, while a field rating of 7·5 would be predicted from a final questionnaire rating of 7·0. In contrast, our second regression line predicts a field rating of 4·8 from a final questionnaire rating of 4·0, and a field rating of 7·3 from a final questionnaire rating of 7·0.

We feel that this change in the difference between the field and final ratings was due to soldiers who did not like a given item – be it a generalised dislike for a food or a specific one based on previous experience with that ration item. These soldiers probably avoided eating items during the test by trading or throwing them away, and then rated them poorly on the final questionnaire to make sure the dislike was registered, in spite of the fact they were requested to rate only items they had actually consumed during the test. It is these final ratings that tended to be excluded in the second regression, where evidence of consumption was required.

From a practical viewpoint, the consistent direction across the like–dislike continuum is important. Field ratings will not be lower than the final ratings, and confidence that the ratings will be no lower in the field allows for more decisive action on the part of the developer who only has collected final ratings.

While even the final ratings in the first analysis provided good prediction of field ration item acceptance, the data based on ratings when consumption was assured, i.e., those of the second and third regressions, more closely agree with the field ratings, as indicated by the differences in the regression parameters between the analyses.

In pursuing the validity of using final questionnaire ratings in place of field ratings as a measure of ration acceptance, one question that must be answered is whether a soldier's field ratings are a rehearsal, so to speak, for his final questionnaire ratings. The use of a control group in a future test that provides only the more retrospective, final questionnaire ratings but no field ratings would address this issue.

ACKNOWLEDGEMENT

The authors thank Dr Edward Hirsch for his technical guidance and editorial comments in the preparation of this paper.

REFERENCES

Kleinbaum, D. & Kupper, L. (1978). *Applied Regression Analysis and Other Multivariable Methods.* Duxbury Press, Boston, MA, USA

Labovitz, S. (1970). The assignment of numbers to rank order categories. *Amer. Sociol. Rev.*, **35**, 515–22.

Labovitz, S. (1971). In defense of assigning numbers to ranks. *Amer. Sociol. Rev.*, **36**, 521–2.

Meiselman, H. L. (1988). Consumer studies of food habits. In *Sensory Analysis of Foods, 2nd Edition*, ed. J. R. Piggott. Elsevier Applied Science, London, UK, pp. 267–334.

Peryam, D. R. & Pilgrim, F. (1957). Hedonic scale method of measuring food preferences. *Food Technol.*, **11**(9), Supplement 9.

Popper, R., Hirsch, E., Lesher, L., Engell, D., Jezior, B. & Bell, B. (1987). *Field Evaluation of Improved MRE, MRE VII, and MRE IV*, NATICK/TR-87/027, US Army Natick Research Development and Engineering Center, MA, USA.

Traylor, M. (1983). Ordinal and interval scaling. *J. Market Res. Soc.*, **25**(4), 297–303.