3-1962

An Evaluation of the Performance of a Beef Taste Panel

Minnie Ruth Chambers
University of Tennessee, Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by Minnie Ruth Chambers entitled "An Evaluation of the Performance of a Beef Taste Panel." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Food Science and Technology.

Bernadine Meyer, Major Professor

We have read this thesis and recommend its acceptance:

Ruth Buckley, Jeanette Biggs

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

March 9, 1962

To the Graduate Council:

I am submitting herewith a thesis written by Minnie Ruth Chambers entitled "An Evaluation of the Performance of a Beef Taste Panel." I recommend that it be accepted for nine quarter hours of credit in partial fulfillment of the requirements for the degree of Master of Science, with a major in Foods and Institution Management.

Bernadine Meyer
Major Professor

We have read this thesis and
recommend its acceptance:

Ruth Buckley

Jeannette Bigg

Accepted for the Council:

Dean of the Graduate School

AN EVALUATION OF THE PERFORMANCE OF A BEEF TASTE PANEL

A Thesis

Presented to

the Graduate Council of

The University of Tennessee

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

Minnie Ruth Chambers

March 1962

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

CHAPTER I

INTRODUCTION

It is desirable that an individual whose sensory judgment is used in evaluating the quality of foods possess the ability to discriminate and the ability to duplicate his judgments. The present study was conducted primarily for the purpose of evaluating the performance of members of a beef taste panel in the Food Research Laboratory at the University of Tennessee with respect to these competencies. These judges had had considerable experience in scoring beef.

Tenderness is a quality which is universally desired in meat (Lowe, 1955). Methods of evaluating tenderness in meat which are described in the literature include: subjective or sensory testing methods which employ human senses to evaluate tenderness; mechanical devices such as shearing machines, gauges and penetrometers as objective methods; and, determination of the amount of connective tissue by chemical methods. Satisfactory evaluation of tenderness through the use of any one method alone is questionable because of the limitations of each method, but any method used should reflect the consumer's evaluation of tenderness. Therefore, use of objective methods which correlate with sensory estimates of tenderness is highly to be desired. A second purpose of the present study was to correlate two methods of measuring tenderness. A shearing machine was used as representative

of one of the objective methods. The correlation between this method of measuring tenderness and the subjective panel scores for tenderness was determined.

Through the use of duplicate sets of rib roasts representing 5 breeds of beef it was hoped to test the 2 attributes, ability to discriminate and to duplicate judgment, necessary in a qualified judge. It was anticipated that the variation in breed would provide a range of tenderness in order to test each judge's ability to discriminate and since duplicate roasts from each animal were used, there was opportunity to evaluate the ability to duplicate judgment.

A standardized method was used for cooking the test roasts. Samples from a comparable location of each roast were used for scoring by the taste panel and for testing on the shearing machine. Judges scored one sample from each roast for tenderness, flavor and juiciness and indicated a preference between randomized paired samples.

The ranking of breeds as to tenderness was compared and cooking losses were summarized.

It was hoped that recommendations for membership on future beef taste panels could be made on the basis of the findings of this study.

CHAPTER II

REVIEW OF LITERATURE

In general, there are 2 methods for measuring quality attributes in foods: (1) the use of a sensory panel and (2) the use of instruments designed to measure certain physical or chemical properties (Deatherage and Garnatz, 1952). A question arises concerning the use and value of the 2 methods. Since objective tests are few and limited in their application, in many cases dependence must be placed on subjective methods (Overman and Li, 1948).

Foster (1954) states that one conspicuous need in the use of the panel technique is the need for standardization. Boggs and Hanson (1949) state that there are few well-established facts regarding techniques and plans for sensory-difference tests. They list several limitations of the panel technique such as variation in response of individuals due to a large number of unknown factors; expression of results in relative rather than absolute terms; requirements of time and material in order to obtain valid results; and the insufficient personnel of small laboratories from which to select satisfactory panels. Although many problems exist in the use of taste panels, it is generally recognized that they fill a need for measurement of food quality which cannot be met by other approaches (Bennett et al., 1956; Foster, 1954; Knowles and Johnson, 1941).

I. TYPES OF TESTS

Several tests have been employed in taste panel studies. Boggs and Hanson (1949) list and describe in detail 5 types of sensory tests: ranking tests; paired sample tests; triangle tests; dilution tests; and, scoring tests. In the ranking test, judges are asked to rank samples in increasing or decreasing order of some characteristic. In the paired sample and triangle tests, judges are asked to indicate whether there is a difference in a particular characteristic or a difference of any kind between samples. Two samples are used in the paired test while in the triangle test 3 samples are used, 2 of which are duplicates. The dilution test is used to determine the smallest amount of an unknown that can be detected when mixed with a standard material. In the scoring test, the scorer is expected to detect quality differences in samples and to assign a value to each sample which represents the difference observed. Peryam and Swartz (1950) have proposed a sixth test, the duo-trio test, which is similar to the triangle test.

Comparisons of the results obtained by different sensory methods have been made. Dawson and Dochterman (1951) compared the paired and triangle tests in measuring differences in flavor of chocolate fudge made with and without vanilla flavoring. They found neither one to be more precise than the other as a basis for selecting reliable panel members. However, they did express more confidence in the results of the triangle test because of the opportunity to eliminate judges who were unable to identify duplicate samples. They also compared the ranking and scoring tests and found them to function

equally well in detecting differences in acid concentration of apple juice, but the scoring test indicated degrees of difference that were not shown by ranking.

Gridgeman (1955) compared the paired test, the duo-trio test and the triangle test. He found the paired test and the triangle test to be about equal in precision and appreciably superior to the duo-trio test.

II. PANEL SELECTION AND TRAINING

Small panels are usually used in food research laboratories. Boggs and Hanson (1949) suggest using panels of 5 to 10 persons. Since the number of persons on the panel is small, those who are members must have exceptionally good judging ability (Overman and Li, 1948). Therefore, selection of panel members is of major concern.

Techniques for selecting panel members have not been standardized (Foster, 1954; Overman and Li, 1948). Selecting panel members on the basis of taste sensitivity alone is of limited value (Foster, 1954; King, 1937; Mackey and Jones, 1954). Girardot et al. (1952) state that sensitivity to the 4 basic tastes or to various odors will only partially determine a person's value as a panel member. They include other factors such as rate of adaptation and recovery, memory for flavor properties, adjustment to the test situation, skill in handling flavor perceptions and the degree of interest and motivation. They suggest that tests for selecting panel members be closely similar to the type to be employed later in order to evaluate these factors.

Foster (1954) lists as possible factors sex, age, health, motivation, sensitivity and intelligence but states that little is known about their relative importance.

Boggs and Hanson (1949) discuss health, smoking, psychological factors and age as possible causes for individual variation in ability to distinguish differences in foods. They consider health and emotional stimulus as having definite effects while age and smoking are of doubtful influence.

A preliminary period of training is considered desirable (Boggs and Hanson, 1949; Lowe, 1955). Bennett et al. (1956) in an investigation of the value of training found that a 3-week period of training resulted in improved performance in judging aroma and flavor of varying concentrations of rancid beef. They also found that the consistency of performance was improved by training.

Training should include the presentation of samples differing in all the characteristics of importance in the investigation (Boggs and Hanson, 1949; Foster, 1954).

Because standards have not been established for all kinds of foods and panel techniques are not uniform, Overman and Li (1948) feel that methods are needed for comparing panel members as to discriminating ability and consistency of judgment. They suggest 2 methods of analysis: (1) a preliminary study and evaluation of data in which the range, number of duplicated judgments and absolute deviations from means are used and (2) the use of an analysis of variance.

III. DESIGN OF EXPERIMENT

In a discussion of the experimental design of panel studies, Boggs and Hanson (1949) state that several factors are to be considered in obtaining accuracy in tests. First, there are advantages to limiting the number of characteristics to be judged. Second, the method of cooking the samples should be standardized. Third, the quality of the test foods will affect panel accuracy. Fourth, the use of standards may stabilize the judges' scores. Fifth, the number of replications needed in a particular experiment must be considered. Variability in samples and in judges' performance, the magnitude of the difference between samples and the completeness of information desired will determine the number of replications needed.

Lowe (1955) states that the experiment should be so designed that data obtained from studies of quality differences in food could be evaluated statistically.

IV. TEST CONDITIONS

The conditions of testing are discussed at length by Boggs and Hanson (1949). They consider it desirable to conduct tests in an environment conducive to concentration. Constant temperatures, humidity, lights and background are considered. Utensils should meet the requirement of uniformity and impart no flavor to the food. Judges should be allowed as much of the sample as they need in order to reach a decision. Various temperatures have been found to be satisfactory depending upon the test sample. The usefulness of rinses

in increasing accuracy of flavor judgments is questionable. In most cases the amount of time permitted for judging samples is not limited.

Foster (1954) states that psychologically, any factor capable of influencing judgment is of primary importance.

In a study concerning the test environment, Mitchell (1957a) found that in a taste-difference test the degree of mental effort required was directly related to the amount of difference between samples. His study offers evidence of the necessity for concentration and he emphasizes the importance of psychological and physical conditions on the sensitivity of the taste-difference test.

In another study, Mitchell (1957b) found small differences in taste testing between days. He found that the early morning hours represented a period when the subject was not able to give full concentration and by the last hours of the day, the subject was past giving his best efforts.

Boggs and Hanson (1949) discuss the problem of fatigue. They state that the number of samples which can be successfully evaluated without fatigue depends upon the product and the judge.

V. RELATION OF OBJECTIVE TO SUBJECTIVE TESTS

Lowe (1955) states that objective tests can be used to substantiate subjective appraisals. Boggs and Hanson (1949) discuss the use of chemical and physical tests on foods as valuable supplements to panel tests. They suggest the desirability of showing that a chemical or physical test measures a characteristic that correlates

with something detected by panels.

Kropf and Graf (1959) used 334 beef carcasses representing a wide range of grades, classes and carcass weights to determine interrelationships of various subjective, chemical and sensory tests. They found that sensory tenderness had a highly significant correlation with mechanical shear value.

Satorius and Child (1938) found that palatability-tenderness showed a high correlation with pounds of shear force. However, they found no correlation between panel scores for juiciness and press fluid measurements.

The relationship between panel scores and shear values is not clearly established. Deatherage and Garnatz (1952) in a study to compare the results obtained by subjective and objective methods found a poor correlation between panel scores and shear values.

CHAPTER III

PROCEDURE

Two attributes desirable in members of a taste panel are the ability to discriminate and to duplicate judgments sometimes designated as consistency. The primary purpose of the present study was to evaluate the performance of a panel of judges who were experienced in the sensory testing of beef for tenderness. They were members of the beef taste panel in the Food Research Laboratory at the University of Tennessee. Ability to discriminate was determined by correlating each judge's scores for tenderness with shear force values. Consistency was determined by calculating per cent duplication of judgments in scoring and preference tests.

I. DESCRIPTION OF BEEF USED FOR TESTING THE PANEL

The test material consisted of 68 standing rib roasts, 2 from each of 34 animals, representing 5 breeds of beef. The animals were raised under controlled conditions for a project being conducted by the Department of Animal Husbandry and Veterinary Science. Aberdeen Angus, 2 groups of Herefords, Santa Gertrudis, Jersey and Holstein were the breeds of cattle for the project entitled, "Type and Breed as Factors Influencing Beef Carcass Characteristics and Consumer Acceptance." A predetermined final weight was the criterion for the time of slaughter.

Duplicate series of tests were conducted using 2 standing rib roasts from each animal. The first series of roasts contained the 6th and 7th ribs while the roasts in the second series were adjacent cuts containing the 8th and 9th ribs.

For one day's tests, 4 roasts selected at random from the 34 animals were cooked in a rotary hearth oven by a standarized method. Each roast was cooked from the frozen state in an open pan at a constant temperature of 325° F. to the medium-done stage, 154° F. A meat thermometer was inserted into the center of the roast to determine the end-point of cooking.

Samples for testing were removed from each roast in the same manner. Twelve adjacent slices and 2 cores 1-inch in diameter were cut from all roasts. The slices were numbered so that each judge would receive a comparable slice from all roasts.

II. TASTE PANEL

A taste panel of 6 judges composed of 4 men and 2 women performed the 2 sensory tests. One of these judges was in training, therefore, his scores were not included in the panel's averages. The remaining 5 judges had had several years experience on beef taste panels.

III. TESTS EMPLOYED

Sensory Tests

To accomplish the purpose of this study, 2 sensory tests were

employed. A scoring test was used to evaluate the tenderness, flavor and juiciness of the beef. This provided a basis for correlating the judge's ability to evaluate tenderness with the objective method of evaluating tenderness. This test also permitted evaluation of each judge's ability to duplicate his judgments in the 2 series of tests. Use of a preference test permitted further evaluation of the ability to duplicate judgments.

Scoring test. At each test, each judge was given 4 slices of roast representing 4 different breeds to score for tenderness, flavor and juiciness. He was asked to assign a numerical score which represented his evaluation of each attribute. The following 9-point hedonic scale was the basis for scoring:

Excellent.....9
Very good.....8
Good.....7
Fair plus.....6
Fair.....5
Fair minus.....4
Poor.....3
Very poor.....2
Extremely poor.....1

A sample of the score sheet is included in the appendix, page 32.

Preference test. For the second test, the judges were given 4 samples paired at random and asked to select a preference from each of the 2 pairs. The same animal pairs were used in both series of tests. A sample of the form used in this test is included in the appendix, page 33.

Objective Measurement of Tenderness

A Warner-Bratzler shearing machine was used for the objective measurement of tenderness. Two cores, 1-inch in diameter, were removed from each roast. After cooling for approximately one hour, each core was sheared 3 times. The average pounds of force to shear each roast was calculated.

Other Tests

Routinely cooking losses were determined on the roasts to have a basis for comparing the cooking characteristics of the different breeds.

A comparison of the tenderness of the longissimus dorsi muscle of each breed was made on the basis of shear values and panel scores for tenderness.

IV. ANALYSIS OF DATA

For each series of tests, a correlation was run between shear values for tenderness of each roast and each panel member's score for tenderness of each roast. A correlation of the panel's average score for tenderness and shear values was also determined. In addition for each series of tests, the per cent duplication of judgments in scoring and preference tests was calculated.

CHAPTER IV

RESULTS

I. RELIABILITY OF PANEL

The purpose of the present study was to evaluate the performance of a taste panel in scoring beef. To accomplish this purpose, a panel of 5 judges scored the tenderness of 34 samples of 6th and 7th rib roasts from 5 breeds of beef in one series of tests (Series A) and 34 samples of 8th and 9th rib roasts from the same animals in a second series of tests (Series B). Samples comparable in location to the samples scored by the panel were tested on a shearing machine by cutting 3 shears on each of 2 cores obtained from each roast. A summary of the data obtained in these tests is presented in Table I. The percentage agreement between Series A and Series B and the direction of the difference are also presented in Table I.

Discriminating Ability of Panel

The range of the scale used by the panel tends to indicate that the panel was somewhat discriminating in scoring tenderness. The difference between the highest and lowest scores was 2.2 in Series A and 1.8 in Series B. The shear values tend to substantiate the discriminations among the breeds of beef made by the panel. Good agreement as to the direction of the difference between the 2 series was shown for all breeds except the Blount Herefords. The breeds which

TABLE I

AVERAGE PANEL SCORES FOR TENDERNESS
AND SHEAR VALUES OF TWO SERIES
OF BEEF ROASTS

Breed of cattle	Number of animals	Average panel scores			Average shear values		
		Series A*	Series B*	Difference (per cent)	Series A*	Series B*	Difference (per cent)
Hereford (Blount)	5	7.5	7.6	+1.3	13.6	14.5	+6.6
Hereford (Alcoa)	6	7.7	7.9	+2.6	13.5	13.1	-3.0
Santa Gertrudis	6	5.9	6.1	+3.4	16.6	16.4	-1.2
Angus	5	7.6	7.1	-6.6	14.3	14.6	+2.1
Jersey	6	8.1	7.4	-8.6	13.0	15.6	+20.0
Holstein	6	6.6	6.2	-6.1	16.8	17.2	+2.4

*Series A consisted of 34 roasts from the 6th and 7th ribs.
Series B consisted of 34 roasts from the 8th and 9th ribs.

the panel scored lowest, Santa Gertrudis and Holstein, had the highest shear values. Both panel scores and shear values indicated that there was variation in the tenderness of adjacent roasts in the Jersey breed. On the basis of these findings, it appeared that the panel and machine were measuring the same attribute in the meat: tenderness.

To further evaluate the ability of the judges to discriminate in scoring beef, the judges' scores for tenderness were correlated with shear values for tenderness. The correlation coefficients for each judge and for the panel as a whole are presented in Table II. All correlations were significant at the 1 per cent level, thus indicating that each judge possessed the ability to discriminate degrees of tenderness among different breeds of beef. The correlations of the panel as a whole were higher than the correlations of any individual judge in both series of tests. Judge 1 would seem to be the most discriminating of the 5 judges because his correlations were the highest. Judge 4 would seem to be the second most discriminating judge as indicated by his correlation coefficients. Judges 3, 5, and 6 were about equally discriminating, with Judge 5 showing some improvement in the B series.

Further evaluation of the judges' scoring ability was done on the basis of their use of the full range of the 9-point scale in scoring tenderness. Table III shows the number of times each judge used the 9 values in scoring the tenderness of 68 samples of beef. The scores of Judges 1 and 6 ranged from 2 to 9, a slightly wider range than the values 3 to 9 used by the other 3 judges. Judge 5

TABLE II

COEFFICIENT OF CORRELATION* OF JUDGES' SCORES WITH SHEAR VALUES

Judge	Series A	Series B	Per cent agreement
1	-0.74	-0.74	100
3	-0.65	-0.66	98
4	-0.70	-0.72	97
5	-0.66	-0.71	93
6	-0.64	-0.66	97
Panel as whole	-0.81	-0.82	99

*Needed for significance at the 1 per cent level, -0.449

TABLE III

JUDGES' ABILITY TO DISCRIMINATE AS INDICATED BY THEIR USE OF FULL RANGE OF SCALE IN SCORING TENDERNESS

Judge	Number of times each value was used									Per cent of scores	
	1 Extremely poor	2 Very poor	3 Poor	4 Fair minus	5 Fair	6 Fair plus	7 Good	8 Very good	9 Excellent	Top third*	Bottom third*
1	0	2	0	2	7	21	15	17	4	52.9	2.9
3	0	0	1	1	4	19	16	15	12	63.2	1.5
4	0	0	2	0	2	12	22	17	13	76.5	2.9
5	0	0	2	0	0	3	25	27	11	92.6	2.9
6	0	2	1	2	3	9	20	21	10	75.0	4.4

*7.4 per cent of shear values were in top third.
54.3 per cent of shear values were in bottom third.

was possibly the least discriminating judge on the basis of his use of the full range of the scale. Ninety-three per cent of his scores were in the top third of the scale and within his 7-point range, he used only 5 of the values. Judge 1 appeared to be the most rigorous judge because of the fact that he used the top third of the scale only 53 per cent of the time. He also used the middle third of the scale more than any other judge. Analysis of shear data indicated that 54 per cent of the scores should have fallen in the top third of the scale and 7 per cent in the bottom third for complete agreement between the 2 methods of measuring tenderness.

Consistency of Panel

The ability of the judges to duplicate their judgments was considered as a measure of consistency. This data is shown in Table IV and also indicated by data in Table I. Since data obtained from shearing indicated that there was considerable difference in the tenderness of the 2 series of Jersey roasts (20.0 per cent) and in the Blount Herefords (6.6 per cent), duplication in scoring by the panel probably should not have been expected for these 2 breeds. For this reason these 2 breeds were excluded in the analysis of the panel's consistency in scoring as presented in Table IV. The 4 breeds in which the shear values differed by 3 per cent or less between Series A and Series B were used to evaluate the panel's ability to duplicate judgments.

The panel as a whole was best able to duplicate judgments for flavor. Little difference was shown in ability to duplicate judgments

TABLE IV

PERCENTAGE DUPLICATION OF JUDGMENTS AND MAGNITUDE OF DIFFERENCE IN SCORES FOR TENDERNESS, FLAVOR AND JUICINESS ON TWO SERIES OF ROASTS FROM FOUR BREEDS

Judge	Number of possible duplications	Number of duplications	Per cent of times duplicated	Sum of difference between scores
		Tenderness		
1	23	7	30	25
3	23	7	30	20
4	23	9	39	20
5	23	11	48	13
6	23	11	48	25
Average			39	21
		Flavor		
1	23	10	43	18
3	23	11	48	13
4	23	15	65	8
5	23	11	48	14
6	23	9	39	17
Average			49	14
		Juiciness		
1	20*	8	40	16
3	23	8	35	16
4	23	8	35	18
5	23	10	43	17
6	23	7	30	19
Average			37	17

*Judge failed to score all samples.

for tenderness and juiciness. The sum of the differences between scores on duplicate samples was smallest for flavor scores and greatest for tenderness scores.

In evaluating the ability of the individual judges, consideration of both the percentage of times scores were duplicated and the sum of the difference between scores indicated that Judge 5 was probably the most consistent. The fact that Judge 5 was least discriminating was undoubtably a factor in his high percentage of duplication of judgments.

In general, the panel showed only a fair ability to duplicate judgments in scoring beef for tenderness, flavor and juiciness. Judges 1 and 3 in scoring tenderness, and Judge 6 in scoring juiciness, had less than 1/3 of the possible number of duplications. Judge 4 was the most consistent panel member when scoring flavor. He duplicated his scores for flavor in 65 per cent of the cases. This was the only case where a judge duplicated his scores more than 50 per cent of the time.

Consistency of Preference

As another measure of consistency, the panel was asked to express a preference among 17 pairs of samples duplicated one time. These samples were taken from the same 68 roasts representing the 5 breeds of animal. The panel members showed a greater ability to duplicate preference than the ability to duplicate scores for tenderness, flavor and juiciness. The judges repeated their preferences

as follows:

Judge 1 - 82 per cent
Judge 6 - 82 per cent
Judge 5 - 71 per cent
Judge 3 - 65 per cent
Judge 4 - 65 per cent

II. RANK OF BREEDS IN RELATION TO TENDERNESS

A summary of the subjective and objective ranking of the 5 breeds of beef is presented in Table V. The values are the averages of the 2 series of tests for panel scores and shear values. Both the panel scores and shear values indicated that the Santa Gertrudis and Holstein were the least tender breeds of beef. The panel scored the tenderness of both breeds Fair plus and each required 16 to 17 pounds of pressure to be sheared by the machine. The other breeds were rated Good to Very good by the panel and required 13 to 14 pounds of pressure to be sheared by the machine.

The panel scores for flavor and juiciness showed little difference among breeds. All breeds were rated Good to Very good for these 2 qualities.

III. SUMMARY OF COOKING LOSSES

A summary of the cooking losses for the 68 roasts is presented in Table VI. The total cooking losses for the various breeds in both series ranged from 13 to 21 per cent. The only noticeable difference was that the 2 dairy breeds, Jersey and Holstein tended to have low percentage of drippings and slightly lower total cooking losses than

TABLE V

SCORES FOR TENDERNESS, FLAVOR AND JUICINESS
OF FIVE BREEDS OF BEEF

Breed of cattle	Averages of Series A and B			
	Panel scores			Shear values
	Tenderness	Flavor	Juiciness	Tenderness
Hereford (Blount)	7.6	7.4	7.3	14.1
Hereford (Alcoa)	7.8	7.6	7.6	13.3
Santa Gertrudis	6.0	7.1	7.4	16.5
Angus	7.4	7.5	7.4	14.5
Jersey	7.8	7.4	7.8	14.3
Holstein	6.4	7.2	7.4	17.0

TABLE VI

AVERAGE PERCENTAGE OF COOKING LOSSES
FOR 68 BEEF ROASTS

Breed of cattle	Number of animals	Series A			Series B		
		Evaporation	Drippings	Total cooking losses	Evaporation	Drippings	Total cooking losses
Hereford (Blount)	5	12.1	6.2	18.3	14.0	7.1	21.2
Hereford (Alcoa)	6	12.9	6.4	19.3	13.1	7.1	20.2
Santa Gertrudis	6	11.3	4.4	15.7	13.8	5.1	19.0
Angus	5	12.8	6.6	19.4	13.2	7.8	21.0
Jersey	6	10.7	2.6	13.3	13.7	3.0	16.9
Holstein	6	11.9	2.3	14.2	14.2	2.6	16.8

the 3 beef breeds.

IV. DISCUSSION

The 5 judges whose sensory judgment was evaluated in this study showed a high degree of ability to discriminate on the basis of the correlation of their scores for tenderness with the values obtained through the use of a shearing machine. Although the judges demonstrated a lesser degree of ability to duplicate their judgments, consideration of the magnitude of difference between scores for tenderness, flavor and juiciness on duplicate samples tends to give increased confidence in the consistency of their judgments. Also, the judges showed that they were quite proficient in duplicating preferences.

The findings of this study tend to substantiate the value of the sensory judgment of this panel for use in determining the quality of beef and predicting consumer acceptance. Therefore, it is recommended that these 5 judges be retained as qualified members of future beef taste panels in the Food Research Laboratory at the University of Tennessee.

CHAPTER V

SUMMARY

Scope of Study

If the sensory judgment of a group of individuals is to be accepted and used as a valuable tool in measuring variations in quality characteristics of a food, the reliability of, or anticipated skill in exercising this sensory judgment must be known. The present study was designed to evaluate the reliability of performance of a beef taste panel in the Food Research Laboratory at the University of Tennessee.

Tests were conducted on 2 series of roasts from 5 breeds of beef. For the first series 34 roasts containing the 6th and 7th ribs were used while the second series consisted of 34 adjacent roasts containing the 8th and 9th ribs from the same animals. A standardized method was used for cooking the roasts.

Two sensory tests, a scoring test and a preference test, were employed in each series to evaluate the ability of the judges to discriminate and to duplicate judgments. A Warner-Bratzler shearing machine was used as an objective method of evaluating the tenderness of the meat. The judges' scores for tenderness were correlated with the shear values.

Principal Findings

An evaluation of the panel scores indicated that all members of the panel were discriminating in their judgments. A highly

significant correlation was found between each judge's tenderness scores and shear values. The correlation for the panel as a whole was significant also. The panel showed a fair ability to duplicate judgments and a good ability to duplicate preferences.

Panel scores and shear values indicated that the Santa Gertrudis and Holstein were the least tender breeds of beef but little difference in tenderness was found among the Hereford, Aberdeen Angus and Jersey breeds. Panel scores indicated little difference in flavor and juiciness of the meat among the 5 breeds of beef. Data on cooking losses indicated that the 2 dairy breeds, Holstein and Jersey, tended to have a slightly lower percentage of drippings and slightly lower total cooking losses than the 3 beef breeds.

I. CONCLUSION

The panel of 5 judges who participated in this study in the Food Research Laboratory at the University of Tennessee have been shown to be reliable judges of the qualities of tenderness, flavor and juiciness of beef. The panel as a whole seemed to be better qualified from the standpoint of discrimination than in ability to duplicate judgments.

BIBLIOGRAPHY

BIBLIOGRAPHY

Bennett, G., B. M. Spahr, and M. L. Dodds. 1956. The value of training a sensory test panel. Food Technol. 10, 205.

Boggs, M. M., and H. L. Hanson. 1949. Analysis of foods by sensory difference tests. Advances in Food Research 2, 219.

Dawson, E. H., and E. F. Dochterman. 1951. A comparison of sensory methods of measuring differences in food qualities. Food Technol. 5, 79.

Deatherage, F. E., and G. Garnatz. 1952. A comparative study of tenderness determination by sensory panel and by shear strength measurements. Food Technol. 6, 260.

Foster, D. 1954. Approach to the panel studies of foods and the need for standardization. Food Technol. 8, 304.

Girardot, N. F., D. R. Peryam, and R. Shapiro. 1952. Selection of sensory testing panels. Food Technol. 6, 140.

Gridgeman, N. T. 1955. Taste comparisons: two samples or three? Food Technol. 9, 148.

King, F. B. 1937. Obtaining a panel for judging flavor in foods. Food Research 2, 207.

Knowles, D., and P. E. Johnson. 1941. A study of the sensitiveness of prospective food judges to the primary tastes. Food Research 6, 207.

Kropf, D. H., and R. L. Graf. 1959. Interrelationships of subjective, chemical, and sensory evaluations of beef quality. Food Technol. 13, 492.

Lowe, B. 1955. "Experimental Cookery." 4th ed. p. 34. John Wiley and Sons, Inc., New York, N. Y.

Mackey, A. O., and P. Jones. 1954. Selection of members of a food tasting panel: discernment of primary tastes in water solution compared with judging ability for foods. Food Technol. 8, 527.

Mitchell, J. W. 1957a. Problems in taste difference testing, I. Test environment. Food Technol. 11, 476.

Mitchell, J. W. 1957b. Problems in taste difference testing. II. Subject variability due to time of the day and day of the week. Food Technol. 11, 477.

Overman, A., and J. C. R. Li. 1948. Dependability of food judges as indicated by an analysis of scores of a food-tasting panel. Food Research 13, 441.

Peryam, D. R., and V. W. Swartz. 1950. Measurement of sensory differences. Food Technol. 4, 390.

Pilgrim, F. J., and K. R. Wood. 1955. Comparative sensitivity of rating scale and paired comparison methods for measuring consumer preference. Food Technol. 9, 385.

Satorius, M. J., and A. M. Child. 1938. Problems in meat research. I. Four comparable cuts from one animal. II. Reliability of judges' scores. Food Research 3, 627.

APPENDIX

GRADING CHART FOR MEAT

Date _____ Name _____

Directions: Give full value for excellent quality.
Do not use fractional points.

Values:
9 - Excellent
8 - Very good
7 - Good
6 - Fair plus
5 - Fair
4 - Fair minus
3 - Poor
2 - Very poor
1 - Extremely poor

Sample No.				
Flavor				
Juiciness				
Tenderness				

Comments:

PREFERENCE TEST

Name _____ Date _____

Select a preference from each pair.

	Code		Preferred
Pair I			
Pair II			