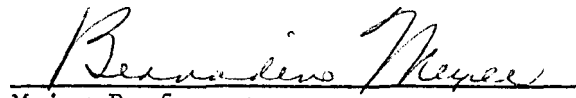


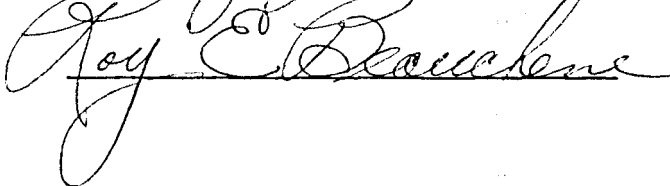
July 30, 1971

To the Graduate Council:

I am submitting herewith a thesis written by Constance Maria Hudgins entitled "The Training and Selection of a Panel for the Sensory Evaluation of Beef." 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 Food Science.


Major Professor

We have read this thesis and
recommend its acceptance:

Accepted for the Council:


Vice Chancellor for
Graduate Studies and Research

THE TRAINING AND SELECTION OF A PANEL FOR
THE SENSORY EVALUATION OF BEEF

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
Constance Maria Hudgins

August 1971

ACKNOWLEDGMENTS

Sincere appreciation is expressed to Dr. Bernadine Meyer for the patient guidance and encouragement with which she advised the planning, conducting, and writing of this study. The author also is indebted to Dr. Mary Jo Hitchcock and Dr. Roy E. Beauchene for their constructive suggestions in preparing this manuscript.

Many thanks are extended to Dr. William Butts and Dr. William Sanders for their assistance with the cooking plans and statistical analyses, to Dr. William Backus for procuring the meat and for his assistance with the sampling of the roasts, to the members of the experienced and "training" panels for their participation, to Mrs. Charles Haga for her cooperation in the earlier hours of the morning, and to Misses Ann Akin, Shelia Fason, and Eugenia Williams for checking data.

The financial assistance provided through a teaching assistantship in the Department of Food Science and Institution Administration has helped make this degree possible.

The author wishes to express deepest gratitude to her roommates, to her fellow graduate students, to the staff of the Department of Food Science and Institution Administration, and especially to her parents, Mr. and Mrs. William O. Hudgins, for their continual understanding, support, and encouragement which contributed greatly to the attainment of this advanced degree.

ABSTRACT

The training and selection of persons for a panel for sensory evaluation of beef proceeded in three phases - - preliminary screening by triangle tests, training sessions with hamburger patties with known differences in composition, and training with 6-7-8th rib roasts varying in USDA grades. Various statistical criteria were used to rank panelists' performance on scoring flavor, juiciness, and tenderness of hamburger patties and roasts. These criteria were the percentage of correct responses on the triangle tests; R^2 values and F ratios for flavor, tenderness, and juiciness scores and correlation coefficients for juiciness scores versus percent fat in cooked hamburger patties; and correlation coefficients for juiciness scores versus numerical USDA grades and for tenderness scores versus Warner-Bratzler shear values from the roasts. Panelists were ranked according to each of the above criteria and mean ranks with standard deviations and rank correlations were computed.

Results indicated that the triangle tests were not a good screening device for prospective panelists. Judges varied in their ability to maintain the same rank order in scoring different attributes of one product or similar attributes in two beef products. New judges for the panel on sensory evaluation of beef were selected on the basis of mean ranks.

TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION	1
II. REVIEW OF LITERATURE	3
Importance of Experienced Panelists	3
General Qualifications for Prospective Panelists	3
Selection Procedures to Obtain Potential Panelists	4
The Training of a Sensory Panel	6
The Effects of Environmental Conditions on Sensory Panel Performance	7
The Effects of Sample Presentation on Sensory Panel Performance	10
III. PROCEDURE	12
Source of Potential Panelists	12
Source of Meat	12
Plan of Study	13
Cooking Procedures	19
Sample Presentation and Environment Control	19
Analysis of Data	20
IV. RESULTS AND DISCUSSION	22
Preliminary Screening	22
Training to Score Using Hamburger Patties Varying in Levels of Fat and Source of Lean	24
Training to Score Using 6-7-8th Rib Roasts from Steers Varying in USDA Grades	29

CHAPTER	PAGE
Rank and Rank Correlations for Prospective Panelists	
for Statistical Criteria Used	31
Selection of Panelists	34
Recommendations for Future Work	36
V. SUMMARY	37
Scope of Study	37
Principal Findings	38
REFERENCES	39
APPENDIX	42
VITA	44

LIST OF TABLES

TABLE		PAGE
1.	Plan for Triangle Tests in the Preliminary Screening Phase .	15
2.	Plan for Scoring Tests in the Training Phase with Hamburger Patties	17
3.	Plan for Scoring Tests of 6-7-8th Rib Roasts in the Second Training Phase	18
4.	Potential Panelists Performance in Ten Triangle Tests in the Preliminary Screening	23
5.	Rank of Panelists by R^2 values and F Ratios in Estimating Flavor in Hamburger Patties	25
6.	Rank of Panelists by R^2 Values and F Ratios in Estimating Juiciness in Hamburger Patties and by Simple Correlation Coefficients	26
7.	Rank of Panelists by R^2 Values and F Ratios in Estimating Tenderness of Hamburger Patties	28
8.	Rank of Panelists by Simple Correlation Coefficients on Scoring 6-7-8th Rib Roasts	30
9.	Rank Order of the Prospective Panelists by Statistical Criteria	32
10.	Mean Rank of Prospective Panelists on Scoring Hamburger Patties and 6-7-8th Rib Roasts	33
11.	Rank Correlation for Prospective Panelists	35

CHAPTER I

INTRODUCTION

Sensory testing is as essential as physical testing and chemical analysis to the comprehensive study of food. It is desirable to employ both subjective and objective tests in connection with research on the functional and organoleptic properties of food products. Sensory evaluation plays an important part in most experiments because it answers the necessary questions of how a food tastes, smells, looks, and feels. Systematic analysis of the sensory properties of foods involves the use of human subjects in a laboratory environment. Thus, selected individuals are considered a laboratory tool to measure and characterize differences in odor, flavor, texture, tenderness, juiciness, and other qualities of experimental samples (Boggs and Hanson, 1949).

Randomly selected and untrained individuals are variable in their judgments. By selecting stable and sensitive individuals, a small efficient panel can be obtained. Selection is important because individuals vary considerably in sensitivity, interest, motivation, and ability to judge differences (Amerine et al., 1965). It has been suggested that seldom a judge is equally proficient in testing all attributes of foods.

Panel members should be in good health. Conscientious observers who concentrate on the job are necessary. Performance of panels is improved by training, control of environmental conditions, and sufficient replications to reduce error (Lowe and Stewart, 1947).

The department of Food Science and Institution Administration of The University of Tennessee through its research laboratory on the agricultural campus cooperates with many departments in the Institute of Agriculture in making sensory evaluation of experimental products. For many years there has been a trained panel for the evaluation of beef and other meat products; however, for various reasons the number of trained people has reached a minimum. At times it has been very difficult to have enough members present for a valid sensory evaluation of experimental products; therefore, it has become necessary to obtain additional members.

The primary purpose of the present study was to train and select new panelists. Secondary purposes were to evaluate the use of triangle tests as a basis for the preliminary screening of prospective panelists and to determine if training and experience in sensory evaluation of ground beef patties can be transferred to beef roasts.

CHAPTER II

REVIEW OF LITERATURE

I. IMPORTANCE OF EXPERIENCED PANELISTS

The reliability of results attained by means of sensory tests is very important. To obtain valid findings, investigators agree that it is necessary to use experienced panelists when an estimate of small differences in food quality is the desired information. Ough and Amerine (1967) compared an experienced and inexperienced panel for their ability to judge color characteristics of rosé wines. They found that the consistency of the experienced judges was significantly better than that of the inexperienced subjects. In beer taste testing better results were obtained with experienced tasters than with inexperienced subjects (Helm and Trolle, 1946).

II. GENERAL QUALIFICATIONS FOR PROSPECTIVE PANELISTS

Mahoney et al. (1957) recommended that individuals selected as panel members be acquainted with the product through normal use, have no strong dislike for the product, be interested in the work, and be willing to give a conscientious, unbiased judgment of the attributes being studied. For sensory evaluation of food products investigators utilize persons who do not have colds or other illnesses which may impair taste or smell sensations.

Much of the work on selection of panelists has been conducted by distillery laboratories. Helm and Trolle (1946) selected panelists

by performance on four series of triangle tests in which there were differences in brews. They found no significant difference among tasters in regard to age, although the age group 30 to 39 had maximum sensitivity. When the data was analyzed in regard to smokers versus nonsmokers, no effect was noted in the results.

Qualified panelists should exhibit intelligence, comprehension, concentration, sustained interest, and motivation toward the sensory testing situation (Dawson et al., 1963). They also should be able to discriminate among foods of the same kind which are known to differ in a particular quality and should be able to duplicate their own judgments (Overman and Li, 1948). Tarver and Ellis (1961) reported that qualified panelists should have an inherent ability to duplicate a difference judgment, an inherent sensitivity to a particular flavor difference, and the ability to control bias by eliminating preferences when detecting flavor differences.

III. SELECTION PROCEDURES TO OBTAIN POTENTIAL PANELISTS

Since individuals vary in sensitivity and ability to detect differences in experimental samples, it is important to select those persons with demonstrated capabilities (Amerine et al., 1965). Panel members are usually required to deal analytically with complex situations. Girardot et al. (1952) stated that the simple factors, such as sensitivity to the four basic tastes or to various odors, are only partial determinants for potential panelists. The level of the panelist's performance varies with other factors such as adaption and

recovery, memory for flavor properties, adjustment to the test situation, skill in handling flavor perceptions, and the degree of interest and motivation. Motivation is considered extremely important. In tests involving indirect comparisons, panelists rely primarily on memory and emotional, mental, and physical conditioning (Byer and Gray, 1953).

Mackey and Jones (1954) found that the ability to arrange water solutions containing varying amounts of four basic taste substances in the proper order of concentration was not highly correlated with the ability to arrange varying levels of basic taste substances when added to foods. High sensitivity, as indicated by low threshold values, did not correlate significantly with the ability to arrange foods containing varying amounts of basic taste substances in order of concentration.

The procedure for obtaining efficient panels, that is workable from the standpoint of time and effort involved, requires at least two stages: first, screening or testing the potential panelists' abilities to make simple discriminations of differences between samples; and secondly, a training period to develop the ability to reproduce qualitative judgments. The second stage adds the more complex requirement of the ability to establish and use stable subjective criteria (Girardot et al., 1952).

Screening is employed by many investigators. The general approach has involved utilization of the same product and the same test that will be used in the actual study; adjustment of the task so that it is increasingly more difficult; and retention of only those

panelists who are able to discriminate between samples (Amerine et al., 1965). Girardot and his co-workers (1952) added the recommendations of using test materials having the same range of variation that will be encountered in the experimental product and of screening on the basis of relative achievement since arbitrary standards cannot be set.

The triangle test has been used by several investigators to detect individuals who are capable of recognizing differences. A triangle test involves the presentation of three samples, two of which are identical and one is different. The judge is requested to indicate the odd sample. In a study utilizing 86-proof whiskeys, Liebmann and Panettiere (1957) found that the triangle test was more sensitive than the duo-trio test when used by inexperienced tasters. The panelists remarked that they found it easier to make a judgment with the triangle test.

IV. THE TRAINING OF A SENSORY PANEL

Panelists must undergo systematic training to develop ability to detect and express differences that are unheeded by untrained individuals. Usually, training is directed toward securing recognition of small differences or toward concentrating on obtaining more consistent judgments. Training assures that panelists acquire a uniform understanding of the attributes and properties being evaluated, of the criteria of descriptive and numerical terms used in the test, and of the relationship between quality or intensity of sensory stimuli. Panelists learn to recognize the precautions that are necessary to minimize the effects of irrelevant factors in the scoring of experimental samples (Amerine et al., 1965).

The very process of seeking the ability to discriminate and reproduce judgments results in training. Bennett et al. (1956) used observers, who were previously untrained, to participate in a three week training period. In a controlled environment the subjects evaluated the aroma and flavor of fresh ground beef diluted with varying levels of rancid ground beef. They found that the ratings of aroma and flavor were inconsistent during the first week, but by the third week the observers had established a definite trend toward uniformity.

It is believed that persons of ordinary sensitivity, if adequately trained, become satisfactory judges, not through increased sensitivity to basic tastes and odors, but through the influence of training. Improvement may be due to increased familiarity with the product and the test method (Amerine et al., 1965). In the opinion of Moser et al. (1950) panelists should be given the opportunity to discuss the scoring of the samples as a way to maintain interest and motivation as well as a way to train the judges. A well-planned training program is essential to the development of the potential panelists. The best results in difference tests will be obtained from sensitive judges who are carefully selected and thoroughly trained (Amerine et al., 1965).

V. THE EFFECTS OF ENVIRONMENTAL CONDITIONS ON SENSORY PANEL PERFORMANCE

The precision of results depends a great deal on the precision of the laboratory instrument and the conditions under which the tool is used. When using a sensory panel as a laboratory tool, ratings of

samples which depend on human judgment may be influenced by physical conditions in the person or in the environment.

The results obtained in any particular study gives a relative measure of the sample with respect to the other samples judged at the same time. Reliability of taste test results depends in part on the maintenance of controlled environmental conditions in order to insure unbiased judgments. The environment should be conducive to optimum concentration. Systematic procedures for objective taste testing, particularly with screened or experienced tasters, make it possible to obtain reliable information concerning differences that are found in experimental samples (Byer and Gray, 1953).

The American Society for Testing and Materials Committee on Sensory Evaluation of Materials and Products, hereafter referred to as ASTM, (1968) has established guides for the sensory testing environment. They recommend that the testing area be completely partitioned into two areas, one for sample preparation and the other containing individual panel booths for the actual testing. To control odors, activated carbon filters should be installed in the air conditioning system. A slight positive pressure in the sensory testing room will reduce the inflow of air from the sample preparation room and other areas into the testing room. Odor-free or low odor materials and equipment should be used in the judging room. Unless color masking is necessary, a good lighting system should provide adequate and comfortable illumination. To mask irrelevant color differences special lighting, such as a low level of illumination or illumination from colored lights or colored filters, should be used. Amerine et al. (1965) also

recommended that a neutral gray background will prevent color distractions and that the light intensity should be 30 to 50 foot candles at the table surface in the booth. The preferred light source is incandescent lights with special blue filters or a source which gives light similar to that of a moderately overcast northern sky.

Mitchell (1957a) studied three sets of environmental conditions: condition A, a single judge was present in the testing room with no noise or disturbance occurring during a test session; condition B, two or more subjects were present per session with no noise or disturbance during the testing period; and condition C, one or more subjects present as well as noise and disturbances during the judging session. Results of this experiment showed a significant difference at the 1% level between all the conditions. A decrease in sensitivity between conditions A and C was expected and illustrated the assumption that quiet is necessary for complete concentration.

Mitchell (1957b) studied judge variability associated with time of day and day of the week. Although there was very little difference, he reported that performance was significantly better on Tuesday than the other days of the week (excluding Saturday and Sunday). He stated that the results supported the simple psycho-physical theory that subjects do better during the earlier part of the week when they are fresher, but poorer later in the week when they have become fatigued or bored. He also found that the middle of the day was significantly the best time. He concluded that human subjects are not capable of machine-like performance, but are susceptible to influence by physical and psychological conditions. However, Dawson et al. (1963) reported that there was no

significant difference between the results of morning (11 a.m.) and afternoon (3 p.m.) sessions when a comparison was conducted on bouillon reconstituted with different kinds of water. No evidence was found by Bengtsson and Helm (1946) to support the hypothesis that panel sessions should be held only in the morning. Mahoney et al. (1957) stated that taste panel sessions preferably are held between 9 and 11 a.m. and 2 and 4 p.m.

VI. THE EFFECTS OF SAMPLE PRESENTATION ON SENSORY PANEL PERFORMANCE

Smooth and orderly presentation of samples aids the judge in making more accurate and consistent evaluations. Mahoney et al. (1957) recommended that no time limit be placed on individuals for judging, samples be large enough to permit evaluation, and arrangement order be random. Coding of the samples must avoid giving information to the panel; thus, it is preferable to use three digit numbers selected from a table of random numbers to insure complete randomness (Amerine et al., 1965) and to eliminate unconscious bias (Gray, 1962).

Samples should be presented in the same form, consistency, color, and appearance. Utensils should be the same size and color and should not impart taste or odor to the sample. Sufficient sample should be provided so that retasting is possible (Amerine et al., 1965).

Sensory organs become fatigued rapidly and cease to react when subjected to prolonged stimulations; thus, it is necessary to limit the number of samples per taste session (Bengtsson and Helm, 1946). A study comparing the results of scoring two and four samples per session

was conducted by Ward and Boggs (1951). They found that judges performed equally well when scoring either two or four samples concurrently. Byer and Abrams (1953) reminded their readers that the order and number of samples presented per session greatly influenced the impressions retained by a panelist as he proceeded from one sample to another.

CHAPTER III

PROCEDURE

I. SOURCE OF POTENTIAL PANELISTS

Fourteen people employed by The University of Tennessee, Knoxville, were invited to participate as potential panelists. Eight of these accepted and participated in the entire study. Since permanent panelists were desired, only faculty and staff were asked to participate. The group of potential panelists was composed of five men and three women from various departments in the Institute of Agriculture.

An experienced panel also participated in the present study. This five membered panel was composed of three men and two women who were faculty members at The University of Tennessee.

II. SOURCE OF MEAT

The test products for the screening and training phases were procured, packaged, frozen, and stored by the Animal Husbandry-Veterinary Science department of The University of Tennessee, Knoxville. For the preliminary screening and first training phase ground meat patties were selected as the test material. Foreign lean (ungraded boneless chuck) was obtained from a packing house in Managua, Nicaragua; it contained 5 to 10% fat on a fresh basis. Domestic lean (USDA Good chuck) containing 16% fat on a fresh basis and fat (beef plate) were secured through the East Tennessee Packing Company, Knoxville. . Calculated amounts of ground beef plate fat were added to both domestic and

foreign ground lean to give fat contents in the ground beef of 16, 25, 30, 35, and 45%. These mixtures were fabricated into eight-ounce patties that were 1/2-inch thick. Percent fat (ether extract) was determined on the cooked patties by the Animal Husbandry-Veterinary Science department.

For the final training phase 6-7-8th rib roasts were obtained from the left side of steer carcasses of varying USDA grades. Carcasses were purchased from the East Tennessee Packing Company, Knoxville. The grades were High, Medium, and Low Choice; High, Medium, and Low Good; and High Standard. Weights of the roasts varied from 6.6 to 12.4 pounds.

The test products were wrapped in freezer paper and held in a freezer that was maintained at -4°F . until they were needed for sensory evaluation. Storage time varied with the test material; the beef patties were held for a period of four to nine months and the beef roasts were held from four to twelve months.

III. PLAN OF STUDY

The present study was divided into three phases - - preliminary screening, training with ground beef patties, and training with beef roasts. The final selection of the permanent panelists was based on their performance in the two training phases.

For the first phase, the preliminary screening, a triangle test was used. This test involves the presentation of three samples, two of which are identical and one of which is different. The judge is asked to identify the odd sample. Presentation of the triads was

designed so that the first pair contained the greatest difference in fat levels (29%), presumably the least difficult, and the last pair presented to the judges represented the least difference in fat levels (5%), presumably the most difficult. Two triads were presented per tasting session (Table 1). The panel sessions were scheduled between 2:00 and 3:30 p.m. on Monday and Friday afternoons for a three week period.

Prior to beginning each training phase, two practice sessions were held in which the score card, terminology, and the attributes to be evaluated were discussed. Additional discussion periods were held when half of the panel sessions were completed. Questions were answered at these sessions, as well as any time during the study when they were proposed by a panelist.

It is believed that some knowledge of the problem is an aid in maintaining the interest of the panelists; therefore, the panelists were informed during the practice sessions that the ground beef patties contained two sources of lean beef and varying levels of fat and that the roasts varied in USDA grades. Panelists were permitted to discuss among themselves how they had rated the samples only after the taste session had been completed. At any time after he had completed his judgments on the samples, a panelist was allowed to compare his scores with those of other judges. Permitting the judges to view the score cards, familiarizing them with the problem, along with a dessert party that was held at the end of the training phase with the beef patties were efforts to maintain interest and to motivate the panelists.

Table 1. Plan for triangle tests in the preliminary screening phase.

Test number	Fat level difference	Samples by percent of fat	Source of lean
1	29%	16% vs. 45%	Domestic
2	20%	25% vs. 45%	Domestic
3	29%	16% vs. 45%	Foreign
4	20%	25% vs. 45%	Foreign
5	14%	16% vs. 30%	Domestic
6	10%	25% vs. 35%	Domestic
7	14%	16% vs. 30%	Foreign
8	10%	25% vs. 35%	Foreign
9	5%	30% vs. 35%	Domestic
10	5%	30% vs. 35%	Foreign

In the first training sessions a scoring test was used to evaluate the attributes of flavor, juiciness, and tenderness of ground beef patties. Scoring was done on a nine point scale for flavor and juiciness and an eleven point scale for tenderness. Descriptive adjectives were used to define the numerical points in the scale. A sample form is included in the Appendix. Ground beef patties with two sources of lean and fat levels of 16, 25, 30, and 35% were employed as the test product. The plan was designed so that all treatments were paired with every other treatment and a total of seven judgments were obtained for each source of lean and level of fat. The interior sections of broiled beef patties were divided into one-inch cubes for presentation to the judges. Samples were scored in pairs with two pairs presented at each sensory testing session (Table 2).

Again scoring tests employing the same descriptive scale were used in the training sessions to evaluate 6-7-8th rib roasts. Sixty-seven roasts were evaluated according to the plan shown in Table 3. The sixth and seventh rib areas were sliced with a Hobart electric meat slicer into 3/16-inch slices which were numbered from the sixth rib end so that each judge received a half slice from the same relative location in each roast. The scoring sessions for the roasts were scheduled at 11:00 a.m. on Monday, Wednesday, and Friday. To obtain an objective measurement of tenderness for correlation with the judges scores for tenderness, a three-inch section from the eighth rib end of the longissimus dorsi muscle was removed from each roast and two one-inch cores were obtained for shearing with the Warner-Bratzler shear. Each core was sheared three times.

Table 2. Plan for scoring tests in the training phase with hamburger patties.

Test number	First pair ^a		Second pair ^a	
1	25D	30D	30D	30F
2	30F	25D	16D	35D
3	35F	30D	35D	16F
4	16F	35F	30D	25F
5	25F	16D	16D	25D
6	35F	35D	35D	30F
7	16D	16F	25F	25D
8	16F	30F	35F	16D
9	30F	35F	25F	35F
10	35D	25D	35D	25F
11	35F	25D	30D	35D
12	30D	16F	30F	16D
13	25D	16F	16D	30D
14	16F	25F	25F	30F

^aNumbers indicate percent fat in the ground beef samples; D, domestic lean; F, foreign lean.

Table 3. Plan for scoring tests of 6-7-8th rib roasts in the second training phase.

Test number	Numerical USDA grade ^a			
1	8	11	12	13
2	8	11	11	12
3	10	12	14	13
4	9	11	12	14
5	9	11	12	14
6	9	--	13	14
7	9	10	13	14
8	9	11	11	14
9	8	11	13	14
10	10	10	13	12
11	10	12	13	12
12	10	10	14	13
13	9	13	12	13
14	10	10	12	12
15	9	11	14	12
16	8	10	13	12
17	8	12	14	13

^aNumbers represent the following grades: 8 = High Standard; 9 = Low Good; 10 = Medium Good; 11 = High Good; 12 = Low Choice; 13 = Medium Choice; 14 = High Choice.

IV. COOKING PROCEDURES

The ground beef patties were removed from the freezer the day before the test and refrigerated until the morning of the test when they were thawed at room temperature. The patties were broiled in a conventional oven in which the broiling unit had been preheated. Two patties of the same treatment were placed on a broiling pan eight inches from the unit. A copper-constantan thermocouple was inserted into the center of one patty of each treatment to determine the time of turning and removal from the oven. Preliminary investigation indicated that for medium-well done patties, they should be turned at 130-140°F. and removed from the oven at 180-190°F.

Roasts were thawed at room temperature on the day prior to cooking and then refrigerated overnight. Weights were taken to estimate cooking time. For cooking, roasts were placed rib side down in roasting pans in a rotary despatch oven that had been preheated to 325°F. The end point temperature of 154°F. was determined by mercury filled thermometers that had been inserted in the center of the longissimus dorsi muscle. Roasts were placed in the oven at different times in order that they reach the end point temperature at approximately the same time and about one hour before the scheduled panel session. The cooked roasts were put in a holding oven (150°F.) until the longissimus dorsi muscle was removed and sliced for the sensory evaluation.

V. SAMPLE PRESENTATION AND ENVIRONMENT CONTROL

All samples were presented with a three digit code preassigned from a table of random numbers. The samples were served on beige

plastic five-inch plates with one sample per plate. A knife and fork as tasting utensils and a glass of tap water at room temperature for rinsing the mouth between samples were provided. The judges were seated at individual tables that were supplied with individual fluorescent lamps. Plates of samples were placed on white enamel trays. In the triangle tests the odd sample was placed in different positions on the tray for each judge in order to equalize any positional bias. In the scoring test each judge received the samples in the same random order on any one day.

Cooking odors were removed by an exhaust fan and normal room temperature was maintained. Noise was controlled as much as possible during each sensory testing session. The only movement permitted was related to the presentation or removal of samples for the judges and to the judging of the samples.

VI. ANALYSIS OF DATA

Since the objective of the present study was to train and select new panelists who possess the ability to subjectively evaluate differences among beef products, methods to measure their abilities became an integral part of the study. The performance of each prospective panelist was evaluated by various statistical methods.

In the first phase, which was the preliminary screening with triangle tests, each potential panelist's performance was assessed by observing the percentage of correct responses.

Sensory data from the second phase, training with hamburger patties varying in fat level and source of lean, were evaluated by

analyses of variance (two-way design) for each of the panelists for each attribute scored. The hamburger patties were fabricated with known differences in composition which should have produced differences in sensory attributes. Juiciness is generally recognized to be a function of the amount of fat in meat and the domestic chuck was of higher quality than the ungraded foreign chuck. Therefore, differences in fat level and source of lean should have been the sources of variation in each panelist's sensory scores. F ratios from analysis of variance (two-way design) and coefficients of determination (R^2) were used as two criteria for ranking the panelists. Higher R^2 values were assumed to indicate the more discriminating and consistent judges. The analyses of variance were computed on an IBM 360 with a program developed by the author. An analysis of variance was also computed for the percentage of fat (ether extract) in the cooked patties. A simple correlation coefficient between the juiciness scores of each panelist and the percent of fat in the cooked patties was calculated. Panelists were then ranked according to the magnitude of correlation coefficients.

The second training phase utilized 6-7-8th rib roasts. In this phase, criteria for ranking the panelists were the magnitudes of simple correlation coefficients for each attribute scored versus the numerical USDA grades and simple correlation coefficients between the panelists' tenderness scores and the Warner-Bratzler shear values.

Since panelists were ranked by several criteria for each attribute evaluated, the mean for each panelist was calculated for their performance in scoring hamburger patties and in scoring roasts. Also, all possible rank correlations were computed.

CHAPTER IV

RESULTS AND DISCUSSION

The training and selection of panelists for the sensory evaluation of beef proceeded in three phases - - preliminary screening, training with hamburger patties varying in levels of fat and source of lean, and training with 6-7-8th rib roasts from steers varying in USDA grades. Panelist performance was evaluated by individual statistical analysis of the sensory data collected. Panelists were ranked according to each of the statistical criteria for each of the attributes scored.

I. PRELIMINARY SCREENING

Preliminary screening with triangle tests was planned to provide a method of identifying persons who have the ability to detect differences in test products with known differences in composition. Persons were asked to select odd samples in tests in which hamburger patties contained varying fat level differences. When the triangle test is used as a screening device, panelists with less than 60% correct selections should be eliminated, as recommended by the ASTM (1968).

Performance of the prospective panelists in the triangle tests resulted in 40 to 90% correct responses (Table 4). Although two of the eight people (I and M) should have been eliminated, all persons were carried through the entire study in order to evaluate the use of the triangle test as a screening device. Panelists were ranked according to their percentage of correct responses.

Table 4. Potential panelists performance in ten triangle tests in the preliminary screening.

Potential panelist	Percent correct	Rank
F	70	4
G	80	2
H	90	1
I	50	7
J	70	4
K	60	6
L	70	4
M	40	8

II. TRAINING TO SCORE USING HAMBURGER PATTIES VARYING IN LEVELS OF FAT AND SOURCE OF LEAN

An analysis of variance (two-way design) for each individual panelist, both experienced and potential, for each attribute evaluated provided the criteria to rank panelists. The degree to which a person discriminated between samples and was consistent in his replicate judgment was reflected in higher F ratios and R^2 values. Rank orders were determined separately for the experienced and "training" panels by the magnitude of the F ratios and R^2 values (Table 5). It should be pointed out that the experienced panel had little or no previous experience in scoring ground beef, but had scored beef roasts from two to almost 20 years.

Five of the eight prospective panelists (F, G, H, I, and M) exceeded or approximately equaled the experienced panel mean for R^2 . Four of the prospective panelists (F, G, H, and M) and two of the experienced panelists (A and E) had significant F ratios. When the prospective panelists had different ranks for the R^2 values and F ratios, the change in rank order could be related to the use of F ratios from the analysis of variance in which the variation was due to the fat level and the use of R^2 values which included variation associated with level of fat, source of lean, and interaction between these two.

To rank panelists' performance on juiciness of hamburger patties, the criteria were R^2 for juiciness, F ratios with the fat level as source of variation, and correlation coefficients between panelists' juiciness scores and percent of fat in the cooked patties (Table 6).

Table 5. Rank of panelists by R^2 values and F ratios in estimating flavor in hamburger patties.^a

Panelists	R ² (percent)	F ^b	Rank	
			R ²	F
Experienced				
A	41.2	9.73**	1	1
B	17.9	1.16	5	5
C	26.1	2.62	4	3
D	28.6	1.90	3	4
E	32.9	6.83**	2	2
Mean	29.3			
Prospective				
F	43.1	9.52**	2	2
G	47.7	10.01**	1	1
H	29.0	6.33**	5	3
I	31.6	1.74	3	7
J	16.6	1.84	6	5
K	14.8	1.80	8	6
L	15.2	0.57	7	8
M	29.1	4.71**	4	4

^aExperienced and prospective panelist ranked separately.

^b F ratio from analysis of variance with level of fat as the source of variation.

**p < 0.01.

Table 6. Rank of panelists by R^2 values and F ratios in estimating juiciness in hamburger patties and by simple correlation coefficients.

Panelists	R ² (percent)	F ^a	r ^b	Rank		
				R ²	F	r
Experienced						
A	51.5	15.68**	0.97**	2	1	1
B	53.9	13.80**	0.92**	1	2	2
C	20.0	2.82*	0.90**	4	4	3
D	39.7	8.52**	0.89**	3	3	4
E	19.8	1.58	0.61	5	5	5
Mean	36.7					
Prospective						
F	10.0	1.43	0.35	8	8	8
G	21.9	3.94*	0.86**	6	3	4
H	28.2	2.32	0.65	3	7	7
I	23.7	2.69	0.85**	5	5	5
J	27.8	3.31*	0.66	4	4	6
K	18.7	2.60	0.95**	7	6	1
L	32.0	5.84**	0.88**	1	1	3
M	29.2	5.79**	0.91**	2	2	2

^aF ratios from analysis of variance with level of fat as the source of variation.

^bJuiciness score versus percent of fat in cooked patties.

*p < 0.05.

**p < 0.01.

The analysis of variance for percentage of fat in the cooked patties indicated that fat content varied with level of fat in the uncooked patties and source of lean ($p < 0.01$). The R^2 value of each potential panelist was lower than the experienced panel mean. The juiciness scores of four prospective panelists (G, J, L, and M) and four experienced panelists (A, B, C, and D) were related to fat level as shown by significant F ratios. Four experienced panelists and five prospective panelists had significant correlation coefficients between their juiciness scores and percent of fat in the cooked hamburger patties. Again, the rank orders were inconsistent; only three potential panelists (F, I, and M) achieved the same rank according to all three criteria.

The criteria to rank panelists on estimating tenderness of hamburger patties were R^2 values and F ratios associated with variation in the source of lean (Table 7). Five of the potential panelists had higher R^2 values than the experienced panel mean. The F ratios were significant for potential panelists F, I, and J and experienced judges A, B, and E. Only two of the potential panelists (F and I) achieved the same rank by each criterion used.

According to the recommendation of the ASTM (1968), panelists should be selected from the high ranking candidates whose F ratios are significant at the 5% level. No prospective panelist achieved significant F ratios in scoring all three attributes of the patties. Panelist F had significant F ratios in the scoring of flavor and tenderness, but not juiciness. G and M had significant F ratios in the scoring of flavor and juiciness. J had significant F ratios for juiciness and

Table 7. Rank of panelists by R^2 values and F ratios in estimating tenderness of hamburger patties.

Panelists	R ² (percent)	F ^a	Rank	
			R ²	F
Experienced				
A	22.6	10.17**	4	2
B	16.9	4.82*	5	3
C	23.1	1.21	3	5
D	24.5	3.68	2	4
E	33.6	11.50**	1	1
Mean	24.1			
Prospective				
F	37.6	23.76**	1	1
G	14.2	0.27	7	8
H	30.4	1.17	3	7
I	36.2	9.41**	2	2
J	19.9	4.84*	6	3
K	26.2	3.74	5	4
L	10.6	1.83	8	6
M	27.5	2.93	4	5

^aF ratios from analysis of variance with source of lean as the source of variation.

*p < 0.05.

**p < 0.01.

tenderness. Some panelists had a significant F ratio in only one attribute: H for flavor; L for juiciness; and I for tenderness. One panelist (K) had no significant F ratios.

III. TRAINING TO SCORE USING 6-7-8TH RIB ROASTS FROM STEERS VARYING IN USDA GRADES

To train potential panelists in the third phase, 6-7-8th rib roasts were scored. Correlation coefficients between the individual scores for flavor and juiciness with the numerical USDA grades of the roasts and correlation of tenderness scores with the Warner-Bratzler shear values were criteria for ranking panelists (Table 8). In this phase the test product did not have a known difference in composition other than grade. It must be recognized that the assignment of a USDA numerical grade to an animal carcass is based on subjective evaluation of its conformation, finish, and quality of flesh and fat. Two of the potential panelists (F and J) had significant negative correlations and one experienced panelist (A) had a significant positive correlation between the numerical USDA grade and flavor scores. Thus, there did not seem to be a clear relationship between USDA grade and flavor. Therefore, no attempt was made to rank the panelists in scoring of this attribute. Three of the experienced panelists (A, D, and E) and three prospective panelists (I, L, and M) had significant correlation coefficients between juiciness scores and numerical USDA grades. Tenderness scores and numerical USDA grades correlations were eliminated from consideration because there was not a significant relationship between numerical USDA grades and Warner-Bratzler shear values ($r =$

Table 8. Rank of panelists by simple correlation coefficients on scoring 6-7-8th rib roasts.

Panelists	Correlation coefficients			Rank	
	Flavor versus USDA grade	Juiciness versus USDA grade	Tenderness versus W. B. shear	Juiciness	Tenderness
Experienced					
A	0.80*	0.95**	-0.84**	1	1
B	-0.64	0.69	0.02	5	5
C	0.63	0.70	-0.62	4	2.5
D	0.50	0.79*	-0.62	3	2.5
E	-0.55	0.88**	-0.41	2	4
Prospective					
F	-0.78*	0.54	-0.01	5	8
G	-0.15	0.69	-0.05	4	7
H	-0.14	0.39	-0.74*	7	3
I	0.34	0.90**	-0.90**	2	1
J	-0.84*	0.04	-0.37	8	6
K	-0.39	0.45	-0.50	6	4
L	0.56	0.76*	-0.41	3	5
M	0.36	0.97**	-0.86**	1	2

*p < 0.05.

**p < 0.01.

-0.70, $p > 0.05$) indicating that tenderness was not a clear cut function of grade in this study. The correlation coefficients for tenderness scores versus Warner-Bratzler shear values varied greatly for both the experienced and potential panelists. Only one experienced panelist (A) and three potential panelists (H, I, and M) had significant correlations. Rank orders indicated all potential panelists varied in their proficiency in scoring the attributes of juiciness and tenderness.

IV. RANK AND RANK CORRELATIONS FOR PROSPECTIVE PANELISTS BY STATISTICAL CRITERIA USED

As seen in Table 9, no one panelist ranked consistently high for all criteria applied. Rank of four potential panelists (F, G, K, and L) varied from first to last place. I varied from first to seventh place. The range of panelist M was from first to fifth. Amerine et al. (1965) and Girardot et al. (1952) have previously reported that panelists are not always equally proficient in evaluating different attributes of a test product.

The mean ranks and their standard deviations, excluding the triangle tests, are presented in Table 10. Comparison of the mean ranks indicated that knowledge and skill acquired in scoring one beef product is not necessarily transferred to another beef product. Three of the potential panelists (F, G, and J) ranked lower on scoring the roasts than on scoring beef patties, while four panelists (I, K, L, and M) improved their mean ranks as the study progressed.

Table 9. Rank order of the prospective panelists by statistical criteria.

Hamburger patties										6-7-8th rib roasts	
Panelists											
	R^2 for flavor	F value (fat level) for flavor	R^2 for juiciness	F value (fat level) for juiciness	"r" for juiciness versus percent fat in cooked patties	R^2 for tenderness	F value (lean) for tenderness	"r" for juiciness score versus numerical grade	"r" for tenderness score versus shear value		
F	2	2	8	8	8	1	1	5	8		
G	1	1	6	3	4	7	8	4	7		
H	5	3	3	7	7	3	7	7	3		
I	3	7	5	5	5	2	2	2	1		
J	6	5	4	4	6	6	3	8	6		
K	8	6	7	6	1	5	4	6	4		
L	7	8	1	1	3	8	6	3	5		
M	4	4	2	2	2	4	5	1	2		

Table 10. Mean rank of prospective panelists on scoring hamburger patties and 6-7-8th rib roasts.

Panelists	Hamburger patties ^a	Roasts ^a
F	4.3 ± 3.5	6.5 ± 2.1
G	4.3 ± 2.8	5.5 ± 2.1
H	5.0 ± 2.0	5.0 ± 2.8
I	4.1 ± 1.9	1.5 ± 0.7
J	4.8 ± 1.2	7.0 ± 1.4
K	5.3 ± 2.3	5.0 ± 1.4
L	4.8 ± 3.1	4.0 ± 1.4
M	3.3 ± 1.2	1.5 ± 0.7

^aMean ± standard deviation.

All possible rank correlations were computed and are presented in Table 11. These analyses indicate only one significant rank correlation, the R^2 rank for tenderness and the F ratio rank for juiciness (fat level) of the hamburger patties. Other rank correlations which are approaching significance are F ratio and R^2 ranks for flavor in hamburger patties and F ratio and R^2 ranks for juiciness in hamburger patties. These correlations indicate that performance on triangle tests was not related to subsequent performance in the two training phases. Therefore, it is concluded that the triangle test could not adequately be used as a screening device.

V. SELECTION OF PANELISTS

In scoring the hamburger patties, the panelists with a mean rank less than 5.0 and a standard deviation of less than 2.0 were I, J, and M (Table 10). In scoring roasts, panelists with a mean rank less than 5.0 and a standard deviation of less than 2.0 were I, L, and M. The only two panelists who were in the top of both training phases and who had significant correlation coefficients between subjective and objective measures, juiciness scores and percent fat in cooked patties (Table 6, p. 26) and tenderness scores versus Warner-Bratzler shear values (Table 8, p. 30) were I and M. These two panelists are considered capable of serving on the regular panel. Other panelists who are recommended for additional training based on the same criteria are G, H, K, and L.

Table 11. Rank correlations for prospective panelists.

		Hamburger patties							Roasts		
		Triangle test	R ² for flavor	F value (fat level) for flavor	R ² for juiciness	F value (fat level) for juiciness	"r" for juiciness versus percent fat in cooked patties	R ² for tenderness	F value (lean) for tenderness	"r" for juiciness score versus numerical grade	"r" for tenderness score versus shear value
Hamburger patties	Triangle tests	1.00	0.15	0.49	-0.05	-0.24	-0.54	-0.22	-0.51	-0.61	-0.54
	R ² for flavor		1.00	0.69	-0.36	-0.14	-0.45	0.38	0.02	0.33	-0.24
	F value (fat level) for flavor			1.00	-0.43	-0.36	-0.43	0.24	-0.24	-0.21	-0.50
	R ² for juiciness				1.00	0.69	0.24	-0.43	-0.45	0.29	0.45
	F value (fat level) for juiciness					1.00	0.60	-0.76*	-0.45	0.52	0.14
	"r" for juiciness versus percent fat in cooked patties						1.00	-0.52	-0.31	0.43	0.36
	R ² for tenderness							1.00	0.62	0.05	-0.24
	F value (lean) for tenderness								1.00	-0.02	-0.02
Roasts	"r" for juiciness score versus numerical grade									1.00	0.43
	"r" for tenderness score versus shear value										1.00

* = p < 0.05.

VI. RECOMMENDATIONS FOR FUTURE WORK

Some suggestions for future studies in the area of training and selection of panelists are: the use of a greater number of triangle tests with sufficient replications in preliminary screening; repeating triangle tests at the end of the training period to determine if panelists can detect a difference even though they may not be able to quantitate it; the use of replicate cuts from the same animals rather than having each test roast from a different animal (permitting better evaluation of consistency of performance); the use of objective criteria in addition to USDA grades to categorize meat samples; and the use of similar statistical criteria throughout the entire study.

CHAPTER V

SUMMARY

I. SCOPE OF STUDY

The purpose of this study was to train and select panelists for the sensory evaluation of beef. Training of potential panelists involved practice scoring of hamburger patties which varied in fat level contents (16, 25, 30, and 35%) and sources of lean (domestic and foreign) and practice scoring of 6-7-8th rib roasts from steers varying in USDA grades (High Standard to High Choice). Various statistical methods were used to evaluate panelists' performance on scoring flavor, juiciness, and tenderness of the hamburger patties and roasts. The study proceeded in three phases. In the first, a preliminary screening was done by triangle tests. In the second phase, training with hamburger patties, panelists were ranked by coefficients of determination (R^2) and F ratios for flavor, juiciness, and tenderness scores. A correlation coefficient of juiciness score versus percent of fat in the cooked patties was also used as a criterion to rank the panelists. For the final training phase, scoring of 6-7-8th rib roasts, the criteria for ranking panelists were correlation coefficients of panelists' scores for juiciness versus numerical USDA grades and correlation coefficients of panelists' scores for tenderness versus Warner-Bratzler shear values. Mean rank and standard deviations were calculated for each panelist according to the different statistical criteria for scoring hamburger patties and for scoring roasts. Rank correlations

were computed on the rank order of panelists for statistical criteria used in this study.

II. PRINCIPAL FINDINGS

Panelists varied in their ability to score the different attributes of a single test product and similar attributes of the two test products. Some panelists, who did poorly on the triangle tests of the preliminary screening phase and would have been dropped from the study if ASTM (1968) recommendations had been applied, developed a discriminatory ability that probably resulted from training through practice and discussions. In the present study triangle tests were not a good device for predicting panelists' performance in scoring tests of beef. Most of the panelists were able to significantly associate differences in juiciness with differences in fat levels of cooked patties as indicated by significant correlation coefficients. Only three prospective panelists had a significant correlation coefficient between their tenderness scores for roasts and Warner-Bratzler shear values. Based on the mean ranks and standard deviations, two panelists were considered capable of serving on a regular panel for scoring beef after the two training sessions. Four panelists were recommended for further training.

REFERENCES

REFERENCES

- American Society for Testing and Materials Committee E-18. 1968. "Manual on Sensory Testing Methods." STP 434. American Society for Testing and Materials, Philadelphia.
- Amerine, M. A., Pangborn, R. M., and Roessler, E. B. 1965. "Principles of Sensory Evaluation of Food." Academic Press, New York.
- Bengtsson, K. and Helm, E. 1946. Principles of taste testing. Wallerstein Laboratories Communications 9(28), 171-180.
- Bennett, G., Spahr, B. M., and Dodds, M. L. 1956. The value of training a sensory test panel. Food Technol. 10, 205-208.
- Boggs, M. M. and Hanson, H. L. 1949. Analysis of foods by sensory difference tests. Advances in Food Research 2, 219-258.
- Byer, A. J. and Abrams, D. 1953. A comparison of triangle and two sample taste test methods. Food Technol. 7, 185-187.
- Byer, A. J. and Gray, P. P. 1953. Some consideration in applying systematic taste testing to beer. Wallerstein Laboratories Communications 16(55), 303-314.
- Dawson, E. H., Brogdon, J. L., and McManus, S. 1963. Sensory testing of differences in tastes. I. Methods. Food Technol. 17, 1125-1131.
- Girardot, N. F., Peryam, D. R., and Shapiro, R. 1952. Selection of sensory testing panels. Food Technol. 6, 140-143.
- Gray, P. P. 1962. Modern taste testing from the viewpoint of the master brewer. Wallerstein Laboratories Communications 25(87), 228-241.
- Helm, E. and Trolle, B. 1946. Selection of a taste panel. Wallerstein Laboratories Communications 9(28), 181-194.
- Liebmann, A. J. and Panettiere, B. R. 1957. Quality control and consumer testing distilled beverages. Wallerstein Laboratories Communications 20(68), 27-39.
- Lowe, B. and Stewart, G. F. 1947. Subjective and objective tests as food research tools with special reference to poultry meat. Food Technol. 1, 30-38.
- Mackey, A. O. and Jones, P. 1954. Selection of members of a food testing panel: Discernment of primary tastes in water solutions compared with judging ability for foods. Food Technol. 8, 527-530.

- Mahoney, C. H., Stier, H. L., and Crosby, E. A. 1957. Evaluating flavor differences in canned foods. II. Fundamentals of the simplified procedure. Food Technol. 11(9), 37-43.
- Mitchell, J. W. 1957a. Problems in taste difference testing. I. Test Environment. Food Technol. 11, 476-477.
- 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, 477-479.
- Moser, H. A., Dutton, H. J., Evans, C. D., and Cowan, J. C. 1950. Conducting a taste panel for the evaluation of edible oils. Food Technol. 4, 105-109.
- Ough, C. S. and Amerine, M. A. 1967. Rosé wine color preference and preference stability by an experienced and inexperienced panel. J. Food Sci. 32, 706-711.
- Overman, A. and Li, J. C. R. 1948. Dependability of food judges as indicated by an analysis of scores of a food tasting panel. Food Res. 13, 441-449.
- Tarver, M. and Ellis, B. H. 1961. Selection of flavor panels for complex differences. Industrial Quality Control 17(12), 22-26.
- Ward, A. C. and Boggs, M. M. 1951. Comparison of scoring results for two and four samples of corn per taste session. Food Technol. 5, 219-220.

APPENDIX

GRADING CHART FOR MEAT

Date _____ Name _____

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

Scale for scoring tenderness:

- 10 - Extremely tender
- 9 - Very tender
- 8 - Tender
- 7 - Moderately tender
- 6 - Slightly tender
- 5 - Slightly tough
- 4 - Moderately tough
- 3 - Tough
- 2 - Very tough
- 1 - Extremely tough
- 0 - Reject sample

Scale for scoring flavor and juiciness:

- 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:

VITA

Constance Maria Hudgins was born in Montgomery, Alabama, on January 26, 1947. From 1953 to 1961 she attended elementary and junior high in a number of schools in North Carolina and Georgia. In 1961 she moved with her family to Princess Anne, Maryland, where she attended and graduated from Washington High School. In September, 1965 she entered Auburn University, Auburn, Alabama, where she received a Bachelor of Science in Nutrition and Foods in June, 1969. From the fall of 1969 until the present she held a teaching assistantship in the Department of Food Science and Institution Administration and has been working toward a Master of Science degree in Food Science.

She is a member of Omicron Nu, the Institute of Food Technologists, and the American Home Economics Association.

The author is the daughter of Mr. and Mrs. William O. Hudgins of Seaford, Delaware.