EFFECT OF TRAINING ON

THE RELIABILITY OF SENSORY EVALUATION

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

COLLEGE OF NUTRITION, TEXTILES,

AND HUMAN DEVELOPMENT

BY

CLAUDIA KERN CONKIN

DENTON, TEXAS

DECEMBER, 1981

The Graduate School

Texas Woman's University

Denton, Texas

November 06, 1981

We hereby recommend that the thesis prepared under our supervision by Claudia Kern Conkin entitled Effect of Training on the Reliability of Sensory Evaluation

be accepted as fulfilling this part of the requirements for the Degree of Master of Science.

Dissertation/Theses signature page is here.

To protect individuals we have covered their signatures.

ACKNOWLEDGMENTS

The author would like to thank Dr. William Hart, chairman of my committee, for his continual guidance throughout this project. I would also like to extend my appreciation to Dr. Shirley Baird and Mrs. Kathy Lawson for their input as committee members.

Gratitude is also extended to the Space Food Development Group for their participation as panelists in this study and to Mrs. Mildred S. Bass for her assistance in the typing of this paper.

Most of all, I would like to thank my husband, Johnny, for his constant support, input, and understanding throughout the course of my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In developing a menu for the Space Shuttle Food System, each food goes through considerable analysis and testing. The food system as a whole must meet certain criteria. First, it must meet the mechanical constraints, which include food weight and volume. The food system also must be able to withstand extremes in temperature, pressure, humidity and vibration. Second, the food must comply with the nutritional demands of the mission, especially the expected inflight energy demands. Third, the foods must adhere to the organoleptic constraints. The foods must be aesthetically acceptable to be a positive morale factor during the flight (1). Because of these constraints, only certain types of foods and processing methods can be utilized.

New foods constantly are being evaluated for their suitability for use in flight through the use of a laboratory taste panel. This taste panel is made up of various members of the Space Food Development Group. The panel ascertains the acceptability of the flight foods and evaluates new products for possible use in the space food program.

The method used for sensory evaluation is a hedonic scale, a nine point scale for preference. Panelists are asked to rate the characteristics of a food in each of the following areas: appearance, color, odor, flavor, texture, and overall acceptability (see appendix A). In order for a food item to be considered for use on a flight, it must receive an overall mean score of six or above (1).

Training in the Space Food Development Group in the past has consisted of a procedure in which each participant read a memorandum which established the guidelines for a taste panel (see appendix B). Variance in the scoring of food items, both that among individual panelists and that within the group as a whole, has in the past been larger than desired to insure reliable results. Current literature on this subject emphasizes the importance of using trained judges. Training has been found to improve the discriminatory ability of a panelist and thus makes it possible to obtain more reliable scores (2). However, training must be of a nature that allows the panelist to apply his new knowledge to all foods in a consistent manner.

PROBLEM STATEMENT

The questions dealt with in this study were: Will trained panelists be able to detect textural differences when complicated with the flavor attribute of sweetness? Will the same panelists be able to detect textural differences when complicated with saltiness?

HISTORICAL PERSPECTIVE

The key analytical tool in sensory evaluation is the taste panel. This panel can establish an initial approximation of overall acceptability, show how the product is rated on various dimensions, and ascertain which dimensions seem to interact most with overall acceptance. The value of the panel depends on the objectivity, precision, and reproducibility of the judgment of the panelists (3). This means that a test must possess validity, the assumption that what is measured is what is planned to be measured. It must also possess reliability, the ability to get the same results on repeated trials.

If sensory evaluation is viewed basically as an experiment in social or physiological psychology, then the many sources of invalidity which might jeopardize the meaningfulness of the interpretation of the test data readily may be conjectured. One area in which invalidity may occur is the design of the test. Sample size and uniformity may not be equal for all panelists. The temperature of the sample may be varied. Many times improper vehicles or carriers may be used when tasting such foods as condiments. The position of a sample in relation to all of the others to be tested may cause unplanned interactions to occur. The code which a sample receives may give clues to a panelist about the food, such as its source or order of intensity. Thus, when designing a test, care must be taken to assure that such factors do not affect the final results of the test (4).

Besides design flaws, invalidity also may be the result of interactions involving the panelists. Many factors may bias a panelist and therefore invalidate the response. One form of biasing is through prior information about a product. Crocker et al. (5) have suggested that a taster's knowledge or guess as to the source of a product may cause assumptions to be made concerning flavor and quality, which may influence his opinion.

Biasing may occur through group interactions. One study compared a round-table or discussion type setting with an individual test situation (6). Here it was noted that discussions can create flavor and preference differences where none exist. It also was shown how easily one person's response can be swayed by the response of another, especially by one in a leader's position.

Pangborn (7) showed that extreme hunger decreases an individual's ability to discriminate food differences. Thus, the time of day that a taste test is planned and the degree of hunger a judge possesses can bias the results of a test.

The ability to discriminate accurately also may be affected by substances eaten prior to a taste test. Berg (8) showed that alcohol enhances the perceived sweetness of a sucrose solution. Kamenitzky and Pilgrim (9) reported that added sucrose significantly decreased the intensity of caffeine perceived.

Another area of variability comes from the various backgrounds of the panelists. Because of different experiences and exposures to foods, people tend to have individualized definitions of terms and conceptions of what is proper in foods (10).

Thus, to best utilize a taste panel, proper training must exist. Panelists need to learn proper testing techniques to avoid as many biases as possible. These individuals also need to understand and feel comfortable with the terminology used in analysis in order to use it effectively. Panelists need to develop familiarity with the product and its characteristics and to develop a common language for expressing these characteristics. This would improve the individual panelist's ability and that of the panel as a whole in providing consistent judgments (11).

It is necessary to establish a training program which will allow a panelist to perform in a consistent manner. Zook and Wessman (12) coordinated a four-step training program at Quaker Oats which is typical of training programs utilized by other food companies. The first step is the interview or prescreening. The importance of this step is to establish the interest, availability, and health of potential panelists. The influence of age on acuity is controversial. It was suggested, however, that the criterion of selection should be ability and not age (13).

The second step in a traditional training program is the screening. This usually is performed by the administration of triangle difference tests (12). In this type of test three samples are given of which two are duplicates. The panelists are asked to identify the odd sample. It is at this step that each panelist's affinity or discriminatory ability can be assessed. For this reason, usually twice as many panel candidates as necessary are screened to allow an adequate pool of individuals from which to choose. Those candidates who demonstrate comparable levels of ability to discriminate the attributes are chosen and proceed to the next step. Another important consideration at the screening step is to be sure the judges are representative of the group for which the product is designed (2). The panel should be representative of the population that will use the product.

The next step is the training itself. The purpose of this step is to familiarize the individuals with test procedures and to allow all judges to acquire an understanding and feel comfortable with the descriptive terms so that they can be used effectively in grading. This will put all members of the panel on an even footing so that all can make contributions to the general pool of knowledge about the sensory characteristics of a product (12). Training, therefore, will improve the individual's ability to recognize and identify sensory attributes. It also will improve an individual's sensitivity to and memory for test attributes so that sensory judgments can be precise and consistent (13).

Evaluation of the training program, step four, is initiated soon after training begins to help the panel leader identify problems among individual panelists (12). Generally, evaluation takes place after ten to twelve training sessions. The leader may determine here if more training is needed and, if so, in what areas. Panelists again are given triangle tests; the leader is now looking for replicated correct judgments. In industry, ten to twelve replications are the criteria for considering a panel reliable. Periodic testing of an established panel is also necessary, but this testing need not be as extensive. Four to six replications have been found to produce reliable data. Training is considered finished when replications of the tests are consistent.

This reliability phase of a training program emphasizes the importance of obtaining consistent data, not only directly after training but also at occasional periods of time throughout the use of the panel. There are two ways of establishing reliability. The first, replicate testing, is the retesting of individual judges at regular intervals with a set of "standard samples". New scores can be compared with earlier scores to show consistency of the judges. The second means of testing reliability is to have judges express the degree of difference perceived between samples. If the panel as a whole has found a significant difference, then an individual panelist's response can be evaluated by comparison to the overall results (2).

Not all trainees can be chosen to act as a panelist. The experimenter must be able to interpret the data obtained during testing and to choose only those judges who are both consistent in their judgments and sensitive to product differences. Many day-to-day problems exist that must be considered when conducting taste panels. Design plans need to minimize the interactive effects of judge, product, and time, rather than test for them.

It is the main objective of the training program to eliminate those factors which can influence or bias responses. Proper training has been shown to help panelists remain sensitive to a tested attribute in order to disallow the biasing effects of time that can be encountered on a long-term study. One study (14) looked at the consistency of scores in the long-term evaluation of frozen spinach and cauliflower. It showed that training allowed the judges to remain sensitive to the standard taste of these two vegetables for the test period of fourteen months. The training program for this study was similar to the Quaker Oats design described previously. Reliability was established by comparing the original test score with scores obtained throughout the study. It was found that all long-term scores closely approximated the original test score.

Training also has been shown to eliminate other internal factors of judge, product, or time. A meat science research laboratory developed a training program which sensitized panelists to three sensory attributes of meat quality (13). To test the validity of the program, i.e., to test the effect of outside factors on scores, four separate panels were trained. There was consistency in the scoring from session to session and among the various panels. Thus, the results showed that proper training helped alleviate the effect of session interaction.

The real test of any training program, however, is the panelists' ability to apply the learning. One area where training can be helpful is in the interaction of sensory attributes with each other.

Mackey (15) studied the validity of judgments for taste substances in water as compared with foods. The rationale for this study was that the selection of a judge for a panel is usually based on his ability to detect very small quantities of a substance in a water solution. This implies that a judge's ability to distinguish differences in water is highly correlated with his ability to discern differences in foods. Through a series of threshold tests, Mackey (15) found that the detection of taste substances in foods differed from the detection of the same substances in water. The difference was suggested to be due, first of all, to the presence of naturally occurring taste and flavor substances in foods and the interactions of these tastes with each other. Secondly, it was suggested that this difference could be attributed to the texture of the foods. This study showed that when a sensory attribute, flavor, was taken out of the neutral environment of water it became dependent on other sensory attributes of the foods for the overall flavor.

Taking this thought one step further, Mackey (16) conducted a second study which looked at the discernment of primary tastes in the presence of different food textures. This study investigated three food textures, i.e., liquid, gel, and foam and the effect of these on taste perceptibility of sweet, sour, bitter, and salty substances. The results showed that taste thresholds were lower when the substances were in water, rather than in the various textural forms. This indicates that the textural state of a product greatly influences the detectability of primary tastes.

Stone (17) studied the effect that viscosity had in the detection of relative sweetness. Both the coating ability of a food and the length of time it is in the mouth can affect the sensitivity an individual has for a taste, depending on the solvent medium. Mackey (18) previously had reported that the water solubility of a food was of greater importance than lipid solubility in contributing to the perception of a taste. Thus, increasing viscosity through the addition of selected gums, rather than oils, should increase taste perceptibility. The results of Stone's study (17) suggest that more viscous materials possessed greater relative sweetness. Aqueous samples resulted in faster responses to sweetness; whereas, the more viscuos a sample, the smoother and longer lasting the effect of sweetness.

In summary, various interrelationships have been found to exist between flavor and other sensory attributes, particularly texture. However, some testing situations require that a food be evaluated by looking at each of the sensory attributes in order to be judged on overall acceptability. A panelist must be able to discern between attributes to effectively score on the various dimensions of the product. To do this requires training. As is commonly done in flavor and texture profiles, training consists of evaluating a standard set of samples for the given attribute. This enables a panelist to evaluate consistently the standard foods over a period of time. But more importantly, it allows the panelist to apply his memory for the learned attribute in day-to-day sensory evaluations. Thus, with proper design and training, panelists should be able to perceive similar differences in test foods as are perceived in the standard samples used in training. Panelists should be able to discriminate these differences, regardless of the effect other attributes have on a food. Thus, reliable and consistent data should be able to be obtained in any test situation with the use of good technique and a trained panel.

HYPOTHESIS

The training of panelists has been shown to improve the reliability of data obtained. However, this training needs to go one step further. To be used, reliable data also must be obtained outside the training environment. When another sensory attribute is added to the standard textural training samples, panelists should still be able to rely on their training to detect textural differences and to disallow other factors to distort their textural perception.

The null hypotheses of this study are: 1) varying levels of sweetness will not affect a panelist's ability to differentiate textural properties; 2) varying levels of saltiness will not affect a panelist's ability to differentiate textural properties.

DEFINITIONS*

Hardness:	The force required to penetrate a solid or semisolid substance with molar teeth.
Brittleness:	The ease or force with which a sample crumbles, cracks, or shatters.
Chewiness:	The length of time in seconds to masticate a sample at the rate of one chew per second in order to reduce it to the consistency for swallowing.
Guminess:	Denseness that persists throughout mastication.
Adhesiveness:	The force required to remove the material that adheres to the mouth (generally to the palate) during normal eating.
Viscosity:	The force required to draw a liquid from a spoon over the tongue.

*Szczesniak, A.S., Brandt, M.A., and Friedman, H.H. Development of standard rating scales for mechanical parameters of texture and correlation between objective and the sensory methods of texture evaluation. J Food Science, 1963, 28, 397.

METHODS AND PROCEDURES

Sampling

The panel of judges used in this study included those who normally participate in the sensory evaluation in the Space Food Development Group, National Aeronautics and Space Administration, Johnson Space Center. Eleven panelists were used, ranging in experience for sensory evaluation from the project leader, with eighteen years food development experience, to newly hired support personnel with no foods background. The panelists were in a normal state of health and none were on any medication that might affect sensory discriminatory ability.

Training

Because of the complexity of sensory attributes of food, this study trained and evaluated panelists only for texture. Since one of the objectives in this study was to evaluate the effectiveness of training on individual and group scores, a baseline measurement was established. Panelists were asked to rank six textural characteristics of food products (see appendix C). This testing took place over three days and consisted of ranking six sets of foods. Each set exhibited degrees of one of the following characteristics: 1) hardness, 2) brittleness, 3) chewiness, 4) gumminess, 5) viscosity, or 6) adhesiveness. Panelists received a score sheet on which they ranked samples using a point structured scale (see appendix D). For example, hardness had nine degrees, with "one" equal to "soft" and "nine" equal to "extremely hard". Individual scores then were compared to the set standards (19).

After a baseline had been established, the training sessions were conducted. These training sessions were modeled after one established by the Meat Science Research Board. The textural scale used was designed by Szczesniak et al. (19).

Training consisted of seven sessions, conducted in the early morning in a test kitchen setting to insure the availability of all panelists and to allow adequate room for the group evaluation. The first session was reserved for the establishment of a professional taste panel technique. Panelists were familiarized with proper testing procedure and general terminology for attributes other than texture. Using Szczesniak's scales of textural characteristics (see appendix C), each of the six remaining sessions dwelled on a respective characteristic. The purpose of this training was to improve panelists' ability to recognize and identify textural attributes and also to improve memory and sensitivity for these sensory attributes so that judgments would be consistent.

During the training sessions, panelists tasted the products in each of the scales and were asked to rank the food as to the degree of the characteristic depicted. Ranking was used since Szczesniak et al. (19) considered that these scales were exemplified by a graded increase for each characteristic. Time intervals between sampling were standardized for one minute to permit recovery from flavor buildup, but not so long as to lose discriminatory ability. Individual scores were compared to the set standards, and discussion was encouraged until there was total group agreement with the established standards.

As stated by Bressman (2), reliability can be tested two ways, i.e., either by retesting with set standards or by comparing individual scores with overall group scores. Both of these methods were used in the course of analysis in this study. After training, evaluation was performed through a series of two posttests. The first posttest was conducted one week after the conclusion of the training sessions. Panelists were asked to rank the standard products in order of a textural characteristic. An individual testing environment was achieved by having panelists evaluate the foods at their desks. Two sets of foods were evaluated daily, the times dependent on the availability of each panelist. Care was taken, however, to insure consistency in sample size, temperature, and integrity. A second posttest, similar to the first, was performed one week later. If training were effective, not only would individual judgments between the two posttests be comparable, but the ratings by close, if not equal to, the set standard.

Testing

Up until this phase, panelists learned the degrees of differences in characteristics through the set standards. Once reliability had been established, a second test was implemented. Because of its blandness, the textural characteristic of gumminess was selected for the test. On the first day, three sets of flour mixtures were prepared with 3.42g, 5.13g, and 6.84g of sugar per 100g of sample, respectively. On the second day, three levels of salt were added to the three sets of flour mixtures as follows: .05g, .15g, and .25g of salt per 100g of sample (see appendix E). Each day the panelists were asked to rank each set of samples for gumminess. The samples were randomly presented to the panelists to prevent any biasing effect the order might have on the evaluation. Results were compared to the set standards by use of Friedman's two-way analysis of variance to ascertain the panel's ability to apply to all other foods its training in the determination of gumminess.

RESULTS AND DISCUSSION

To enhance the value of a sensory evaluation panel, personal preference and individual ideologies must be kept at a minimum and objectivity on the part of the panelist must be increased. This establishes the need for training; but the value of this panel for use as an analytical tool depends on objectivity, precision, and reproducibility in the judgments by the panelists (20). Reliability in a panel comes when these three factors are established.

This study was designed to determine the effect that textural training has on a panelist's ability to evaluate foods complicated with the sensory attribute of flavor. To conduct this study, a reliable taste panel was mandatory. Since one measure of reliability is consistency of scores on tests using set standards, a training program was conducted to instruct panelists in the standard ranks of the foods for six textural characteristics. To determine if the training improved the panel, a pretest preceded the training to establish a baseline for comparison to the posttests.

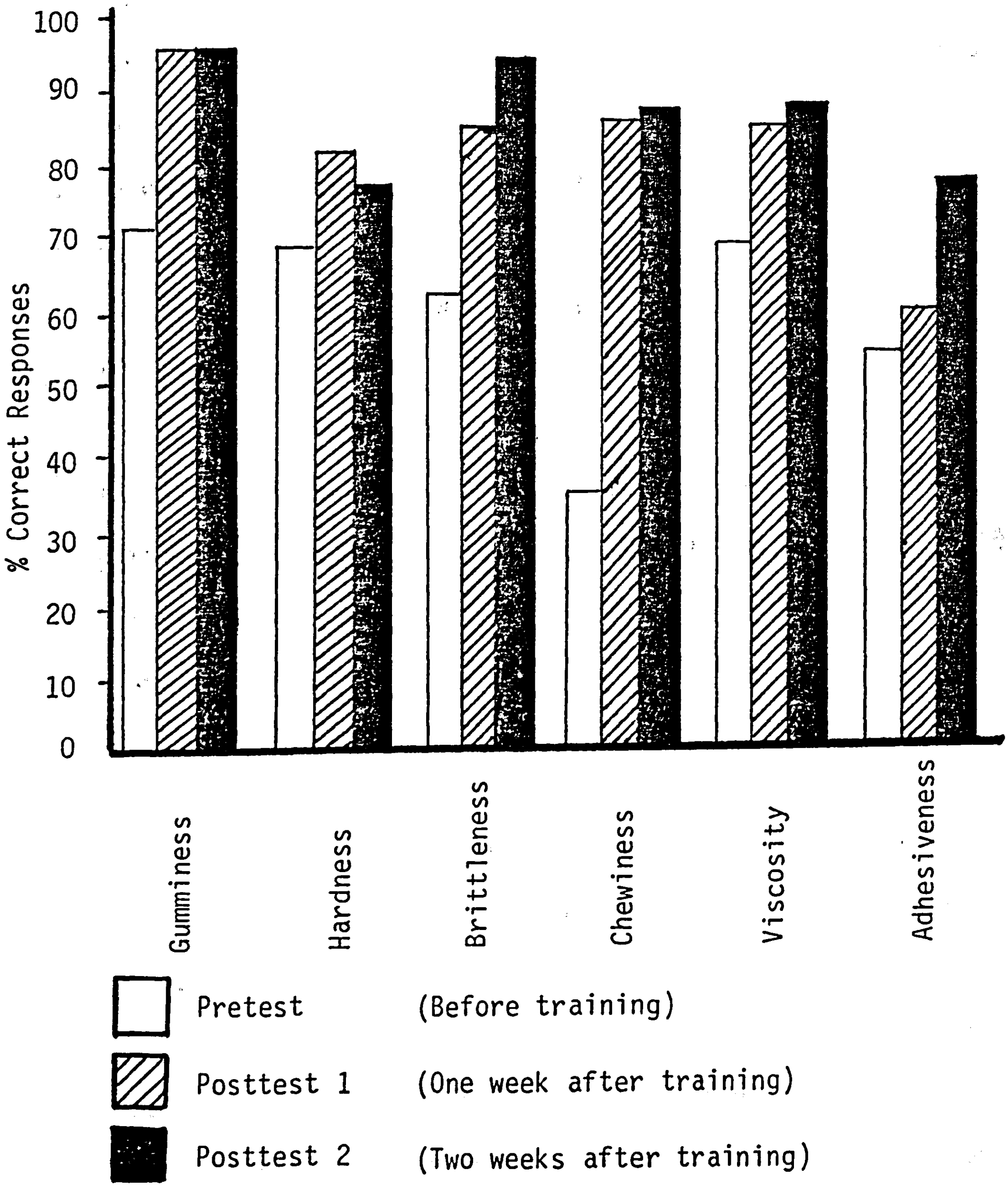
Figure 1 gives a graphic overview of the results of the three sets of tests in this study. Calculation of the percentage of correct answers for each test was performed as follows:

$$\frac{\text{\# correct answers of all judges}}{\text{total \# responses possible}} \times 100 = \text{\% correct answers}$$

It can be seen from the graph that the percent of correct responses

FIGURE 1

PERCENT CORRECT RESPONSE OVER THE THREE TEST PERIODS



increased from the pretest to the first posttest for each characteristic. The most drastic change was for the characteristic, chewiness, for which there was an increase of 50.6%. This indicates that a change, in the ability to rank this characteristic, occurred as a result of the training. Also, this graph suggests that there was a retention of the training for these learned attributes as the percent of correct responses remained constant or increased in five out of six cases. Thus, this figure gives an indication that learning took place in the training sessions.

These percentages then were analyzed using Friedman's two-way analysis of variance to see if the observed changes were significant. Table 1 lists the percentages for each test and the F-value computed from the comparison of the three tests. The increase in correct answers was significant for each characteristic except hardness at the $p \leq .05$ level. Although hardness increased from the pretest to the posttest, the change was not large enough to be significant.

Once a significant difference was detected with analysis of variance, the data was subjected to posttest analysis using the Student-Newman-Keuls multiple range test. The test looked at the significant difference between each possible pair of groups. In each case, the difference was between the pretest and the two posttests.

Learning can also be seen in the improvement of the panelist's ranking ability from test to test. Table 2 shows the mean ranks for each characteristic and each test. The mean rank is the average difference a panelist's rank was from the ideal rank. The group as a whole

TABLE 1

FRIEDMAN'S TWO-WAY ANALYSIS OF VARIANCE FOR PERCENT CORRECT RESPONSES OVER THE THREE TEST PERIODS

Characteristic	% Correct Responses			F-Value
	Pretest	Posttest 1	Posttest 2	
Gumminess	70.9%	96.4%	96.4%	5.57*
Hardness	69.7	81.8	77.8	1.93
Brittleness	62.1	84.8	93.9	20.08*
Chewiness	35.1	85.7	87.0	20.71*
Viscosity	68.2	86.4	87.5	4.16*
Adhesiveness	54.5	59.1	77.3	3.62*

*Significant at $p \leq .05$

mproved over the series of tests. Since a perfect mean rank would be equal to zero, the decrease in the panel's mean ranks from pretest to posttest means their ranks are closer to the ideal rank.

Brittleness, chewiness and adhesiveness showed improvement each time the test was given. In most cases panelists were eager to see how well they ranked the foods. This lead to learning taking place between the two posttests, thus lowering the mean rank even more by the second posttest.

In the case of gumminess, the ranks did not vary between the two posttests. The panelists had to rely on the mouthfeel of the flour pastes used and on visual cues. There could be no memorization of the order as was possible with the other foods since samples were coded only with random numbers. Thus, the only learning that took place occurred between the pretest and the first posttest. It is also worth noting, however, that the memory for gumminess was retained since the mean ranks for the two posttests remained constant.

Hardness was the only characteristic for which the mean rank of the second posttest was higher than the first posttest. Although only a small difference existed, this was attributed to a few panelists still having problems determining degrees of hardness with some samples at the second posttest.

Viscosity posed another problem. It would seem the panelists were fairly good at determining degrees of viscosity before training. At the time of the first posttest, panelists seemed confused. Many comments suggested that their perception of the order differed from the training

TABLE 2

ANALYSIS OF PANELIST'S RANKING ABILITY

Characteristic	Mean Ranks		
	Pretest	Posttest 1	Posttest 2
Gumminess	2.75	1.63	1.63
Hardness	2.22	1.83	1.94
Brittleness	2.67	2.00	1.33
Chewiness	2.71	1.71	1.57
Viscosity	1.50	3.00	1.50
Adhesiveness	2.50	2.00	1.50
Overall	2.32	2.08	1.61

order, and at this point they couldn't remember which order was correct. Each panelist was again asked to review the list of standard foods and the definition of viscosity. when the second posttest was performed, much of the confusion had been alleviated.

The amount of learning that had taken place can also be measured by the degree of variance among the panelists' scores. Each panelist's rank was compared, form one test to another, for each of the characteristics in order to determine the variation between these ranks. This variation illustrates how scattered the data is about a mean. The amount of variation in the way the panelists perceived the order of the foods was then analyzed.

Table 3 shows the variation the panelists exhibited between the tests. The variance found between the first posttest and the pretest suggest a high degree of change for the perceived order of the foods. The variance between the second posttest and the pretest was larger than that observed for the first posttest for five out of six characteristics. Again, it is suggested that learning continued to take place.

Reliability has been defined as the ability to score foods consistently. Comparison of the amount of variation between the ranks of the two posttests can be used as one measure of consistency. In five out of six characteristics, the amount of variance was greatly reduced. Adhesiveness, however, showed greater variance during comparison of the two posttests than did the other five characteristics in relation to its other tests. This variance was found to be the result of a greater increase in correct responses between the two posttest than between the

TABLE 3

COMPARISON OF THE VARIATION BETWEEN TESTS

	Comparison*		
	I	II	III
Characteristic	Post 1 with Pre	Post 2 with Pre	Post 1 with Post 2
Gumminess	.544	.609	.272
Hardness	.742	.795	.553
Brittleness	.977	1.023	.392
Chewiness	1.147	1.147	.487
Viscosity	.758	.678	.429
Adhesiveness	.374	.571	.528

* Friedman's two-way analysis of variance between these three comparisons results in F=4.80, a significant difference at the $p < .05$. The significant difference was found between comparison I and II and comparison III.

pretest and either posttest. In this case, learning continued between the administration of the posttests, thereby further imporving the panelist's ranking ability.

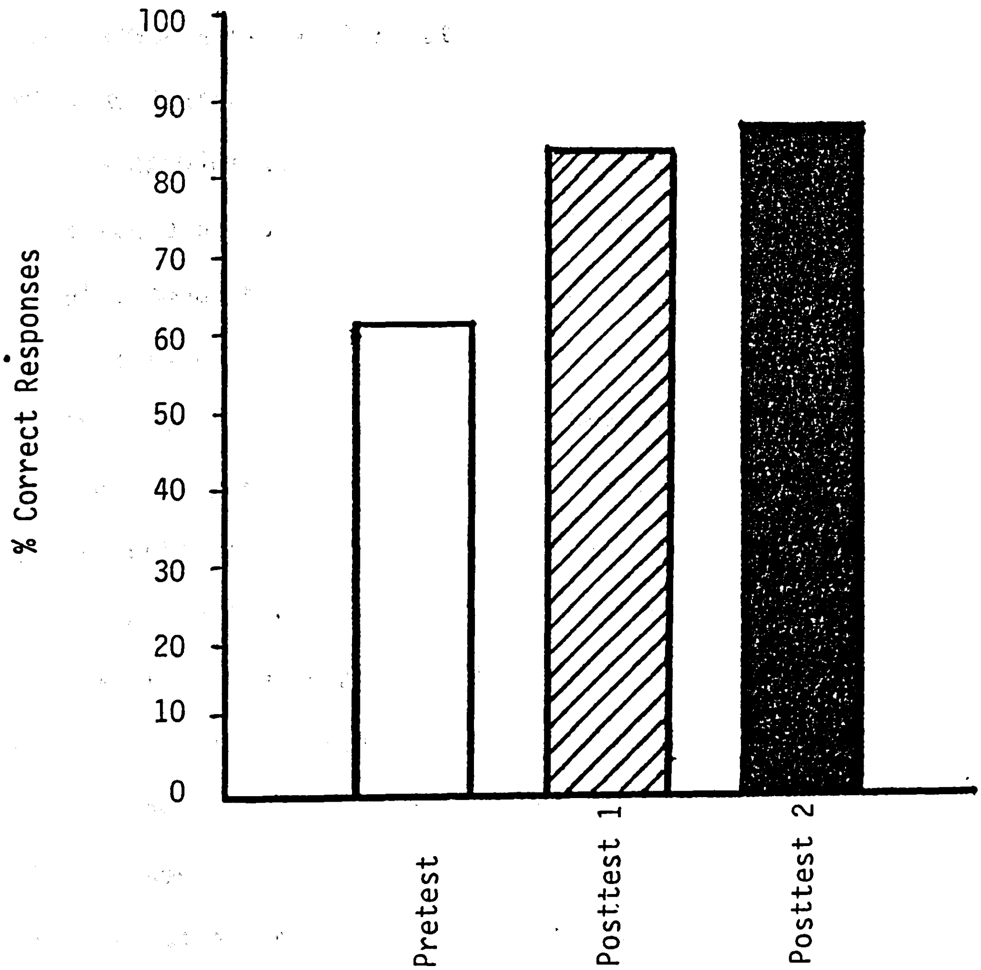
When the three comparisons were analyzed by Friedman's two-way analysis of variance, a significant difference was found between the levels of variance for the tests. By Student-Newman-Keuls Multiple Comparisons, a significant difference was found to exist between the third comparison (two posttest) and the first two comparisons (each posttest with the pretest). Thus, the variance between the two posttest was significantly lower than the variance between the posttests and the pretest. This means that there was more consistency in the scoring between these two posttests then between the other tests.

Analysis of the data showed that learning occurred as a result of the training sessions, allowing the panelists to make more correct responses with each successive test. Figure 2 shows that overall, the number of correct responses increased significantly from 60.6% to 86.2%. There was a significant difference between the pretest and the posttests, but not between the two posttests. It has been established through comparison of the mean ranks that panelists significantly improved their ability to rank foods according to the standards. Also, panelists were able to significantly decrease the variance between tests after the training. Thus, since reliability is the ability to reproduce scores, it was assumed that this panel was reliable.

With the establishment of a reliable panel, the second series of tests began. Mackey (16) previously had reported that various food attributes, such as texture, could affect the perception of flavor.

FIGURE 2

OVERALL PERCENTAGE OF CORRECT RESPONSES



Stone (17) also showed that viscosity could influence the detection of sweetness. However, training has been also shown to sensitize panelists to certain attributes for specific testing as in the determination of meat quality (13) or in texture profile determinations.

Using the gumminess scale because of its blandness, various levels of sugar and salt were added to the flour mixtures. The purpose was to see if the training was effective enough to enable panelists to disregard other sensory attributes such as sweetness and saltiness, when ranking texture.

A problem was encountered at this point in the testing. Some panelists would no longer tolerate the taste of the flour mixtures and as a result, about half the panelists in each test ranked the gumminess of their samples on another sensory characteristic, appearance, rather than taste. Since this study was designed to determine the effect flavor had on textural perception, data from these panelists was disregarded. The study continued with the remaining six panelists.

Table 4 looks at the group's mean ranks for the varying levels of the two flavor attributes. The mean rank decreased as the level of sugar intensity increased. In other words, the more sugar that was added, the more able were the panelists to discriminate different levels of gumminess. However, when the differences in these ranks were analyzed using the chi-square statistic, no significant difference between the three levels of sugar was found. Thus, the level of sugar added had no effect on the panelist's ability to rank the samples for gumminess. When salt was added, the data appeared a bit different. In

TABLE 4

ANALYSIS OF PANELIST'S RANKING ABILITY OF GUMMINESS WITH ADDED FLAVOR ATTRIBUTE

	Mean Ranks			Chi-Square
	Level 1	Level 2	Level 3	
Gumminess with Sugar	2.14	2.01	1.84	1.59
Gumminess with Salt	1.78	2.08	2.11	1.86

* No significant difference was found between the mean ranks of gumminess when three levels of sugar or salt were added

this case the mean ranks increased as the level of salt was increased. This means that the higher the level of salt, the further the scores were from the ideal rank. When chi-square analysis was performed however, no significant difference between the three levels was found.

Whereas the mean rank and chi-square statistics measured the ranking ability of panelists when three levels of the flavor attribute were added, these statistics did not directly address the effect flavor had on a panelist's ability to rank foods specifically for textural qualities. Table 5 shows the percent of correct responses when no flavor was added and when three intensities of sugar or salt were added. The zero level used was the percent correct answers on the second posttest. Because of the consistency between the ranking ability between the two posttests, this was assumed to be a reliable score.

These scores were analyzed using Friedman's two-way analysis of variance. Even though the percentage of correct responses varied between the tests for each added flavor, there was no significant difference between these levels. Thus, the addition of a flavor attribute to a gumminess scale did not affect how a panelist was able to discriminate between different levels of gumminess. Therefore, the null hypothesis that sweetness or saltiness would not affect a panelist's ability to differentiate textural properties was not rejected.

TABLE 5

FRIEDMAN'S TWO-WAY ANALYSIS OF VARIANCE FOR PERCENT CORRECT ANSWERS WITH ADDED FLAVOR

	Percent Correct Answers for Levels of Added Flavor				F-Value*
	0	1	2	3	
Gumminess with Levels of Sugar	94.3	74.3	82.9	94.3	1.73
Gumminess with Levels of Salt	92.0	84.0	60.0	64.0	1.37

* No significant difference was found between the four levels of sugar or salt.

CONCLUSIONS AND IMPLICATIONS FOR FURTHER RESEARCH

Standardization of laboratory sensory testing procedures has been found to be necessary for consistent performance of a given panel. The selection and use of panelists that can discriminate between levels of several sensory characteristics, and reproduce these judgments, requires training of the panelists with standards foods and sensory attributes necessary for the test situation (20).

Because of its beneficial effect on the reliability of panelists used in sensory testing, training was incorporated as a necessary step in the selection of judges. Although presumably experienced sensory panelists were used, training still statistically improved the performance and consistency of this panel's ability to judge textural characteristics.

Once reliability has been extablished, a panel can be used with confidence. Panelists can judge foods for which they were trained, and also utilize the training for analysis and ranking of new foods. Results from this study substantiate that training indeed helped panelists retain memory for the learned characteristic of gumminess and allowed panelists to avoid the interactive effects flavor can have on texture, such effects having been noted in studies by Mackey (16, 18) and Stone (17).

Thus, when three intensities of salt or sugar were added to the various flour mixtures, panelists who tasted the mixtures were still able to rank the samples with acceptable accuracy for the characteristic gumminess. This leads to the conclusion that, because of training, panelists were able to detect textural differences even when complicated with a flavor attribute.

This study showed how training allowed panelists to retain a memory for an attribute even under new test conditions. It also brought to light two areas where perceptibility and precision were affected. Panelists showed fatigue from the number of foods being evaluated and the number of repeated tests. They also displayed lack of enthusiasm with some of the tests because of the nature of the foods. For example, the flour mixtures used were good examples for gumminess; however, they were not well tolerated by the panelists. A more palatable mixture would allow repetitious sampling to be more acceptable to panelists. Thus, research needs to be conducted to see how these psychological and physical aspects of testing can be avoided. The value of this tool depends on its precision and reproducibility which is only attained if all possible outside factors are eliminated.

APPENDIX A

SENSORY EVALUATION

NAME ______________________________ DATE ______________________________

ITEM						
APPEARANCE						
COLOR						
ODOR						
FLAVOR						
TEXTURE						
OVERALL						

COMMENTS:

9-Like Extremely
8-Like Very Much
7-Like Moderately
6-Like Slightly
5- Neither Like nor Dislike
4-Dislike Slightly
3-Dislike Moderately
2-Dislike Very Much
1-Dislike Extremely

APPENDIX B

GUIDELINES FOR TECHNICAL TASTE PANELS

The object of the Technical Taste Panel evaluation is to determine the quality and acceptability of the product. Quality should receive more importance than acceptability since acceptance is a personal preference and our panel is too small to apply much creditability to personal preference.

The following guidelines should be adhered to when rating a product on the Technical Taste Panel:

1. Place emphasis on quality of the product rather than your personal preference.
2. If you absolutely dislike the product because of personal preference, don't rate it.
3. If a product is rated below 7 for any parameter, note the reason. Be specific.
4. The overall rating should be consistent with the other parameters. For example, in most cases, the overall score should not be lower than the lowest individual parameter rating.
5. Remember that an overall rating of 6 or below generally means the product does not qualify for further consideration.

APPENDIX C

STANDARD FOODS USED IN TEXTURAL TRAINING*

Standard Hardness Scale

Panel Rating	Product	Brand or Type
1	Cream Cheese	Philadelphia
2	Egg white	Hard cooked, 5 minutes
3	Frankfurters	Large, uncooked, skinless
4	Cheese	Yellow, American
5	Olives	Exquisite giant size,stuffed
6	Peanuts	Cocktail type in vacuum tin
7	Carrots	Uncooked, fresh
8	Peanut brittle	Candy part
9	Mint Candy	**

Standard Brittleness Scale

Panel Rating	Product	Brand or Type
1	Corn muffin	Jiffy**
2	Cream puffs	
3	Graham crackers	Nabisco
4	Melba toast	Inside piece
5	Ginger snaps	Nabisco
6	Peanut brittle	Candy part

---

* Reference 19.

** Substitution has been made from the original list due to nonavailability of some products.

Standard Chewiness Scale

Panel Rating	Product	Brand or Type
1	Rye bread	Fresh, center cut
2	Frankfurters	Large, uncooked, skinless
3	Gum drops	Gold Crest**
4	Steak	Round, 1/2" thick, broiled on each side 10 minutes
5	Milk Duds	
6	Tootsie Rolls	Midget size
7	Marathon Bar	1/4 of bar**

Standard Gumminess Scale

Panel Rating	Product	Brand or Type
1	40.0% flour paste	Gold Medal
2	42.5% flour paste	Gold Medal
3	45.0% flour paste	Gold Medal
4	47.5% flour paste	Gold Medal
5	50.0% flour paste	Gold Medal

* Reference 19.

** Substitution has been made from the original list due to unavailability of some products.

Standard Adhesiveness Scale

Panel Rating	Product	Brand or Type
1	Hydrogenated vegetable oil	Crisco
2	Biscuit, dough	Buttermilk
3	Cream cheese	Philadelphia
4	Peanut butter	Skippy, smooth

Standard Viscosity Scale

Panel Rating	Product	Brand or Type
1	Water	Tap
2	Light cream	Dairy Pure**
3	Heavy cream	Foremost Whipping Cream**
4	Evaporated milk	Pet Milk
5	Maple syrup	Cary's Pure Maple Syrup**
6	Chocolate syrup	Hershey's
7	Mixture:	
	1/2 cup mayonnaise	Hellman's
	2 tbsp heavy cream	
8	Condensed milk	Eagle Brand, sweetened**

* Reference 19.

** Substitution has been made from the original list due to unavailability of some products.

APPENDIX D

NAME____DATE____________________

RANKING SCALE FOR TEXTURAL CHARACTERISTICS

TEST FOR_______________

These samples have been shuffled. Taste these samples in the order presented, from left to right. Arrange the samples in the order of intensity from_____________ to_____________ based on the parameter listed above.

SAMPLE

____________________ ____________________

____

____

____

____

____

____________________ ____________________

COMMENTS:

APPENDIX E

STANDARD GUMMINESS WITH LEVELS OF SWEETNESS OR SALTINESS ADDED

Product	Grams Sugar Added Per 100 Grams of Product	Grams Salt Added Per 100 Grams of Product
40 % flour paste	3.42	.05
	5.13	.15
	6.84	.25
42.5 % flour paste	3.42	.05
	5.13	.15
	6.84	.25
45 % flour paste	3.42	.05
	5.13	.15
	6.84	.25
47.5 % flour paste	3.42	.05
	5.13	.15
	6.84	.25
50 % flour paste	3.42	.05
	5.13	.15
	6.84	.25

REFERENCES

(1) Stadler, C. R., Sanford, D. D., Reid, J. M., & Heidelbaugh, N. D. Skylab menu development. Journal of the American Dietetic Association, 1973, 62, 390.

(2) Bressman, L. D., & Behling, R. W. The selection and training of judges for discrimination testing. Food Technology, 1977, 31, 62.

(3) Blair, J. R. Interface of marketing and sensory evaluation in product development. Food Technology, 1978, 33, 61.

(4) Schutz, H. G. Sources of invalidity in the sensory evaluation of foods. Food Technology, 1971, 25, 53.

(5) Petit, L. A. Informational bias in flavor preference testing. Food Technology, 1958, 12, 12.

(6) Foster, D., Pratt, C., & Schwartz, N. Variation in flavor judgments in a group situation. Food Research, 1955, 20, 539.

(7) Pangborn, R. M. Influence of hunger on sweetness preferences and taste thresholds. American Journal of Clinical Nutrition, 1959, 6, 280.

(8) Berg, H. W., Filipello, F., Hinreiner, E., & Webb, A. D. Evaluation of thresholds and minimum difference concentrations for various constituents of wine. II. Sweetness: The effect of ethyl alcohol, organic acids and tannin. Food Technology, 1955, 9, 138.

(9) Kamenetzky, J., & Pilgrim, F. J. The effect of sucrose upon perceived intensity of salt and bitter. American Psychologist, 1958, 13, 420.

(10) Szczesniak, A. S. Classification of Mouthful Characteristics of Beverages. In: P. Sherman (ed), Food Texture and Rheology. London: Academic Press, 1979.

(11) Rainey, B. A. Selection and training of panelists for sensory testing. In: M. R. Johnston (Chair), Sensory evaluation methods for the practicing food technologist. IFT Short Course, St. Louis, 1979.

(12) Zook, K., & Wessman, C. The selection and use of judges for descriptive panels. Food Technology, 1977, 31, 66.

(13) Cross, H. R., Moer, R., & Stanfield, M. S. Training and testing of judges for sensory analysis of meat quality. Food Technology, 1968, 32, 48.

(14) Boggs, M. M., Venstrom, D. W., Harris, J. G., & Shinoda, S. Performance of flavor judges in long-term studies with frozen cauliflower and spinach. Food Technology, 1960, 14, 366.

(15) Mackey, A. O., & Jones, P. Selection of members of a food tasting panel: Discernment of primary tastes in water solution compared with judging ability for foods. Food Technology, 1954, 8, 527.

(16) Mackey, A. O., & Valassi, K. The discernment of primary tastes in the presence of different food textures. Food Technology, 1956, 10, 238.

(17) Stone, H., & Oliver, S. Effect of viscosity on the detection of relative sweetness intensity of sucrose solutions. Journal of Food Science, 1966, 31, 129.

(18) Mackey, A. Discernment of taste substances as affected by solvent medium. Food Research, 1958, 23, 580.

(19) Szczesniak, A. S., Brandt, M. A., & Friedman, H. H. Development of standard rating scales for mechanical parameters of texture and correlation between objective and the sensory methods of texture evaluation. J. Food Science, 1963, 28, 397.

(20) Bennett, G., Spahr, B. M., & Dodds, M. C. The value of training a sensory test panel. Food Technology, 1956, 10, 205.