

food



NEW MANUAL SETS FORTH STANDARDS FOR SENSORY TESTING

R54-29

In spite of the importance of flavor, it has received only minor attention in food specifications because of the lack of reliable tests to determine compliance either at the stage of the awarding of subsistence contracts or in checking compliance at time of delivery. Now, however, the new science of sensory evaluation has advanced to the point where such methods are a distinct possibility. This article discusses the development of a manual for this purpose.

By David R. Peryam

With each passing year food technologists are making greater use of sensory testing methods. Their interest includes both acceptance evaluation based on consumer preference and analytical testing to obtain information of a more specific kind. More people are applying such methods to a wider range of problems and are using them with greater confidence. But despite this growing popularity and the fact that sensory testing methodology, which has had a reputation for extreme variability, is beginning to converge toward generally accepted practices and procedures, as yet there exists no standard reference in the field. In fact, there is no reference at all which is worthy of the name. The technologist who has a flavor testing problem cannot just reach for a manual on his reference shelf as he might with a chemical problem. Manuals of sensory testing methods are conceptual, existing as yet only in the minds of those persons experienced in the field who occasionally enjoy the ambitious day-dream of writing the first authoritative work.

This situation represents no hardship for the expert, who has probably kept up with the literature and has had the opportunity to see most of the methods in action. When he has a new problem he can develop an approach in the light of practically all past work without even the trouble of getting up from his chair. But consider the plight of one who has only the flavor problem and the awareness that its solution requires sensory testing. He must work from the ground up, either ferreting his solution out of a highly dispersed and sadly incomplete literature or else discovering and contacting one or more of the peripatetic reference sources. Without both good motivation and perseverance he is likely either to forget the whole thing or else seize upon the first idea suggested. Consider, also, the problem of achieving uniformity of technique even among the experts who need no manuals. Even though they may be equally competent and the solutions they produce may be equally good, there is better than an even chance that two of them, taken

[1]

at random, would not handle a given problem in exactly the same way.

This lack of reference sources in large part merely represents the general status of the field. It is new and in a state of flux. Methods are continually being developed, their variations tried, and their results evaluated. Certainly the field is not ready for rigid standardization; perhaps the absence of attempts at standardization represents no more than proper scientific caution. However, certain methods have been successfully used for so long and rest on sound enough theoretical bases that it is unlikely future research will discard them. Consider, for example, the various discrimination test methods for measuring sensory differences or the use of rating scales for measuring subjective impressions. Everyone agrees on the value of many of the methods but it would be rare indeed to find full agreement on how to apply one of them to a practical problem. There usually are several variations of a method, and the data may depend critically upon some of them. At present, if two laboratories wish to duplicate an experiment it is necessary to draw up a set of working rules. This state of affairs may encourage effort toward methodology research, but it is definitely discouraging when the objective is utilization of methods to solve practical problems.

the specifications problem

This was the situation which confronted the Quartermaster Corps two years ago in attempting to provide for the control of palatability and other aspects of flavor in food specifications. Sensory testing was the only logical approach, but, since uniform results were of first importance, the problem of the non-uniformity of test methods loomed large.

Many of the food specifications have always recognized flavor as a quality factor in its own right, but for a long time this recognition was no more than a nodding acquaintance since neither criteria

nor test methods were provided. When palatability or some other aspect of flavor is deemed critical, a common practice is to require the submittal of pre-contract award samples for inspection and evaluation. The test methods problem first came into the foreground because of the necessity for reliable and uniform pre-award evaluations, although it is also pertinent to the inspection of deliveries under contracts since they must conform to the same requirements. Usually the specification required only "approval by a taste panel" without specifying either the nature of the panel or its mode of operation. A requirement so stated has little meaning either for the supplier who designs the product or for representatives of the Government who must determine its adequacy. Worse still, it may have whatever meaning anyone wants to assign to it. Specificity and uniformity were needed and the attempt has been made to attain them by incorporating into the specification for a given product descriptions of both the requirement and the test for determining compliance. But there must be a reasonable limit on the length of a specification. It is not the place for development of the complex detail which is sometimes advisable. The specification write-ups have proven less than completely satisfactory because they were disproportionately long but were still too short to present the material adequately.

record of progress

Throughout this period of growing emphasis on flavor control in the specifications, work was continuing on the development of techniques suitable for the type of flavor quality control needed in subsistence purchasing. Many of the methods were being put to use in actual pre-award evaluation. Eventually, early in 1952, it was decided that the time was ripe and that sufficient progress had been made to justify an attempt at codification of the results into a Military Standard, a publication with independent and official status which could be referenced in the individual specifications and would be available to Government and suppliers alike, to serve as a uniform standard of operation. A first draft of this publication was ready in August 1952. Following the normal channels of specifications review, it was sent out to representatives of the food industries and to other Service agencies for comment and coordination. Also included in the distribution were a number of laboratories and individuals known to be well qualified in the field of sensory testing.

The reaction was varied and seemed to reflect a good deal of the confusion, misunderstanding, and lack of agreement which has been an unfortunate but constant phenomenon in this area. Some comments were strongly negative, reflecting mainly the attitude that the whole approach was cumbersome and unnecessary. Then there were the people who were merely skeptical, and, finally, there were those who were enthusiastic. It was significant and gratifying that the last category included most of the companies, laboratories, and individuals who had been seriously concerned with sensory testing or flavor quality control work. Objections, comments, and suggestions were frequent, even among the enthusiastic. But the consensus seemed to be that the standard represented a definite contribution that would be good for the field even if it were not changed at all and whether or not it solved all of the flavor problems arising in the procurement of military subsistence.

The major task of rewriting was lightened by such encouragement. By February 1953 all comments had been reviewed and analyzed and a revision of the original manuscript completed which drew heavily on the advice and suggestions that had been offered. Since then there have been several more revisions, all of them minor in relation to the total structure but some of major importance in relation to particular content. This "Military Standard for the Sensory Evaluation of Subsistence Items" has not yet been made official. At the present time it is in the hands of the Department of the Army authority which has responsibility for reviewing and approving all specifications prior to publication.

rules for panel operation

No more than a brief summary of the content of the standard can be given here since it consists of about 6000 words of text in addition to statistical tables. The stated purpose is to set forth certain methods for the evaluation of palatability and other quality aspects of subsistence items by means of sensory tests using panels of human observers. Two general aspects of panel operation are discussed: (1) test conditions required to assure proper psychometric control, and (2) procedures for selecting and training sensory testing panels. Five specific test methods are described; also methods for setting up standards and determining acceptance limits. There was no intent to describe all methods considered suitable for use in quality control or even all those that may at some time be used in Military specifications. The 5 procedures selected for inclusion were

those which had shown most promise in previous pre-award testing at the Institute.

The standard was written with the small laboratory in mind. All of the methods are designed for use with relatively small, relatively constant panels such as may be assembled without too much difficulty in most laboratories or from adjacent factories and offices. The assumption is made that a certain amount of pre-selection of panel members will be possible and that, in conjunction with simple training, it will improve the panels' precision and reliability. A selection method is specified and a general approach to training indicated.

The section on "General Conditions for Panel Testing" is ambitious. It attempts to discuss all of the important controls in the course of only five single-spaced pages. Points covered include physical location of the testing room, presentation of samples, freedom from psychological disturbances, maintenance of physiological sensitivity, serving temperature, length of test session, number of samples in a session, control of time and position errors, the use of water rinses, and, finally, the problem of the panel members' motivation. In choosing these topics many others were omitted either as being of minor importance or on the basis of a decision that such test controls are already being applied with sufficient uniformity. Here is an example of the approach employed, "The following rules shall be observed in order to maintain the panel members' physiological sensitivity: (a) Panel members shall not test until at least one hour after meals, (b) shall avoid smoking, gum chewing, eating candy, drinking soft drinks, or use of other "between-meal" items for at least 10 minutes prior to testing, and (c) shall not test when ill or when suffering from the common cold."

methods and standards

The five test methods are set forth in detail but can be described only briefly here. The palatability rating test is a special application of the familiar rating scale. Its purpose is to provide an estimate of the palatability of a product lot, palatability being defined as "the resultant of sensory properties which tend to make a food pleasing and acceptable to the consumer." It is assumed that panel members' judgments will be guided both by their own preferences and by their knowledge of what represents good quality in the product. Twenty panel members rate the sample on the 9-point scale which is shown in Figure 1. The palatability rating for a product lot is the average of the 20 ratings when the numbers 1 to 9

PALATABILITY RATING TEST

Sample Code	Sample Code	Sample Code	Sample Code
Excellent	Excellent	Excellent	Excellent
Very Good	Very Good	Very Good	Very Good
Good	Good	Good	Good
Below Good	Below Good	Below Good	Below Good
Above Fair	Above Fair	Above Fair	Above Fair
Fair	Fair	Fair	Fair
Below Fair	Below Fair	Below Fair	Below Fair
Above Poor	Above Poor	Above Poor	Above Poor
Poor	Poor	Poor	Poor
Very Poor	Very Poor	Very Poor	Very Poor
Extremely Poor	Extremely Poor	Extremely Poor	Extremely Poor

Figure 1. Questionnaire for the palatability rating test.

are assigned to the scale points as shown.

The flavor identity test is a new one that has not been described elsewhere. The purpose is to determine the extent to which a product lot is identified as possessing a specified flavor. It was developed in connection with the problem of evaluating imitation maple sirups on the criterion of "trueness of flavor." Samples are presented singly and the panel member makes only the simple judgment of whether or not a sample possesses the flavor in question. A minimum of 16 trained test subjects is required. The test result, or flavor identity rating, is the percentage of positive identifications.

The familiar paired-comparison technique is included, the stated purpose of the test being to determine whether a lot sample is equal or superior to a standard sample on some defined criterion, such as palatability, flavor intensity, or flavor purity, when the two materials are available for direct comparison. Instructions for each test define the sensory property to be considered and are so stated that the panel member selects the sample which more nearly meets the specification requirements. For example, if the factor evaluated is palatability, the panel member selects the "better" of the two samples, if it is the absence of off-flavor, he selects the one with the least off-flavor, if it is strength of maple flavor, he selects the one with the stronger maple flavor. All panel members participating in a particular test must know both the criterion of judgment and how they are to answer.

Other procedural details are closely specified in the interests of uniformity. A minimum of 16 panel members participate and the test result is the number of times a sample is selected.

Another familiar item is the triangle test. It was selected for inclusion over other discrimination tests because it is more widely known and seems to be easier to run. It has the usual purpose of determining whether there is a flavor difference of any kind between a lot sample and a standard sample. The normal controls for a test of this type are described. Where there has been disagreement about points of procedure, arbitrary selection of one of the alternatives has been made. Sixteen judgments from at least eight panel members is stated as the minimum amount of data to be obtained on each sample. The number of "correct" discriminations is the test result.

The last of the five methods is the Dilution Number test. Its purpose is to determine whether the flavor of a lot sample is such that it can be mixed with a specified standard material at a pre-established percentage without changing the flavor of the standard material. The Dilution Number is defined as that amount of a sample, expressed as a percentage of the mixture, which may be mixed with standard material without making the mixture noticeably different from the standard itself. The method assumes that the standard material and the Dilution Number will have been predetermined and will be stated in the specification. It requires that the specified

Table I. Acceptance Limits for the Triangle Test According to Panel Size or Number of Judgments and Required Inspection Level

(Given as highest acceptable number of choices of the lot sample)

Panel Size (No. of judgments) ¹	Acceptance limits at various inspection levels			Panel Size (No. of judgments) ¹	Acceptance limits at various inspection levels		
	A	B	C		A	B	C
16	7	8	9	29	12	13	15
17	8	9	10	30	13	14	16
18	8	9	10				
19	8	9	11	31	13	14	16
20	9	10	11	32	14	15	17
				33	14	15	17
21	9	10	12	34	15	16	18
22	10	11	12	35	15	16	18
23	10	11	13				
24	10	11	13	36	15	16	18
25	11	12	13	37	15	17	19
				38	16	17	19
26	11	12	14	39	16	17	19
27	12	13	14	40	17	18	20
28	12	13	15				

¹ When one or more panel members repeat a test the total number of judgments (individual tests) is considered as the panel size.

dilution be tested against the standard by the triangle test.

This Dilution Number standard is one of three kinds that are discussed. A second type is represented by physical reference standards such as those required for the triangle and paired-comparison tests. These standards are prepared and maintained by the responsible Government agency, but the problems involved are not discussed in the present document. The third type of standard is applicable to the palatability rating and flavor identity tests. They are derived by testing representative lots of the product type involved by the same method and with the same panel which will later do the control testing. The palatability standard rating is the grand mean rating of at least six such representative products. The flavor identity standard rating is the over-all percentage of positive identifications from the pooled results on at least six representative samples.

The last section of the standard is devoted to a detailed explanation of acceptance limits for each type of test and how to use them to determine rejection or acceptance in particular cases. Where applicable there are statistical tables giving the limits directly in terms of number of choices for given panel sizes or number of judgments. The table for the triangle test has been reproduced as Table I. Limits are given for three different inspection levels. Level A is the most strict and B and C are progressively easier. The level to be used in

any particular case is to be designated in the specification according to how tight a control is judged permissible or desirable on the basis of past experience with the product.

milepost

As yet it is too early to tell whether or not the present document can serve the purpose for which it was intended in the original planning, i.e., as an official reference for use in food specifications. Its feasibility from an administrative standpoint still has to be determined. No matter what the final decision, however, the effort expended on this project will in no sense have been wasted. This attempt at codification of sensory testing methods, which is more extensive than anything of its kind which has been done before, represents definite progress. It has excited the interest of a considerable number of those people who are concerned with the use of sensory testing methods for evaluating and describing foods and is believed to represent a milestone on the way to eventual standardization.

David R. Peryam
Chief, Acceptance Division