Conducting Consumer Sensory Evaluation Studies

Edgar Chambers IV *

Introduction

Consumers want meat products that "taste good," are affordable and meet various health needs (National Research Council, 1988). Each of those product properties, as well as price and convenience, are important purchasing decision criteria. However, many researchers have concluded that the sensory properties must be right or the product will almost always fail. For example, the meat industry has attempted to sell low-fat products for several years, yet products such as low-fat frankfurters have had poor sales (Rejtman, 1993). The main reason for the lack of success with those franks has been because of a lack of sensory quality (Young, 1993). Kuntz (1994) points to the developing trend toward use of "super-premium" products that may be a back-lash reaction to a decade of health-consciousness. These products are expected by the consumer to be high quality, with the sensory attributes being at or near "perfect."

Consumer sensory evaluation is critical to understanding whether products meet consumer expectations and, if not, how they can be improved. Consumer testing is used for a variety of functions, including product maintenance, product optimization, new product development and evaluation of market potential (Meilgaard et al., 1991). The data collected is some of the most important because the consumer is the ultimate user of the product and only the consumer knows what s/he wants (Muñoz and Chambers, 1993). Affective testing, i.e. testing for acceptance or preference, should only be done with consumers except in rare circumstances.

Overall, few studies have reported on the use of consumer evaluation for meat research (see summary by Chambers and Bowers, 1993), but consumer studies are being conducted continually by industry and occasionally by the academic community. The traditional attributes of importance have been color, tenderness, juiciness and flavor. Courington (1992) discussed the general reasons for conducting market research/consumer tests and gives some information on the types of tests conducted for marketing research purposes. The tests are similar to those consumer tests conducted when the purpose is consumer sensory evaluation.

*E. Chambers IV, The Sensory Analysis Center, Justin Hall, Kansas State University, Manhattan, KS 66502.

Reciprocal Meat Conference Proceedings, Volume 47, 1994.

Types of Consumer Tests

Consumer tests generally fall into either qualitative or quantitative tests. Qualitative tests, often focus groups, are designed to elicit information about consumer desires and to probe consumers' minds for clues to purchase behavior or further understanding of acceptance measures (Courington, 1992; Chambers and Smith, 1991). Conversely, quantitative tests, e.g. preference or acceptance tests, are designed to give data that indicate how many people prefer a product, how much they like the product or what they perceive about specific attributes (Meilgaard et al, 1991).

Qualitative Tests

Qualitative tests are used less frequently for consumer sensory evaluation than for marketing research, but they have application to sensory product guidance. Generally, these focus groups or one-on-one interviews are used by marketing research to help determine what consumers want and what needs they have that are not already being met. They are used in consumer sensory analysis studies to determine what attributes might be tested with consumers and to determine whether questionnaires are easy to read and complete. In some circumstances, they can be used after quantitative tests to obtain more detailed information on the product. These qualitative tests tend to answer the questions how, what and why (Chambers and Smith, 1991).

Qualitative studies often seem deceptively simple to the novice. It appears there could be few things easier than asking people questions and letting them talk. The keys to successful qualitative research are to ask the right consumers the right questions, not to lead the consumers, and to keep the consumers on track. Unfortunately, that is more difficult than it seems.

Successful qualitative studies revolve around several important parts. The first, as with most studies, is to understand the objectives before the study. Wanting to "find out what consumers think about..." usually is far too general. Although the objective may start there, it needs to be focused and made more specific so that the consumer interviews can begin to answer the question. A practiced interviewer or moderator is essential to the success of the qualitative study. The interviewer must be, or at least appear to the consumer to be, completely neutral. The moderator must allow the consumers to do the talking, maintain the flow of the discussion, but also keep the discussion on the topic of interest rather than letting it wander. If products or information sheets are used in the study, they must be prepared properly and served at the proper temperature in order for consumers to discuss the product as it really is. Under no circumstances should the researcher "count hands." Asking for a "vote" in a focus group almost

assures a biased answer because other people are watching and many consumers like to be on the "winning side." Also the number of consumers is too small in most qualitative studies to provide appropriate information.

Quantitative Tests

Quantitative consumer tests are the most frequently used tests for consumer sensory evaluation. Ultimately, most people want to know which product is liked best or which sample is preferred. These tests also are the most frequently abused consumer tests because researchers often "just add one quick question" on liking or preference to other studies they conduct. Usually, the number of consumers is too small or the population is biased (e.g. "trained" panels or food science students) in those cases where preference or liking is an add-on. The question of acceptance is too important to be relegated to add-on status; it deserves its own well-planned study.

Quantitative studies can be conducted in several ways, depending on the objectives. They may either be conducted in a central location where consumers come to the research or they may be in-home studies.

"Intercept" Studies

Studies that are organized to run quickly, where consumers are asked few questions, often just liking or acceptance, can be run in "intercept" situations. In that type of study, consumers are "intercepted" and asked to participate in an immediate test that usually takes five minutes or less. This type of test works well when products can be prepared so that they are ready when the consumers are ready to evaluate and where a steady supply of unbiased consumers is available. This often is done in shopping malls, state fairs or other locations where consumers congregate. These studies are popular because they usually are inexpensive to conduct.

"Pre-recruited" Studies

For studies that require more of the consumers' time or where meat, such as steaks or roasts, must be prepared to be served at certain times, "pre-recruited" consumers are necessary. In this case, consumers are contacted ahead of time and asked to be at a specified location at a specific time. These tests can be conducted in any location that has ample equipment for preparing the sample and appropriate, ample space for the consumer to test the products without distraction.

"Home Use" Tests

In some cases, consumers need to test the products in "home use" tests. These often are necessary when the consumer must prepare the product and serve it to a family, store the product in a home situation, use it a second day, or a myriad of other reasons that preclude testing the product at a central test site. Home use tests are the least controlled of all consumer studies. Once the product is given to the consumer, there is no guarantee what will happen to it. The instructions for these tests are critical, they must be simple and easily understood by the consumer, especially if a certain type of preparation or storage is necessary.

Conducting Quantitative Tests

All quantitative studies are similar in certain respects. The consumers must be representative of the population that the product is intended for and there must be an adequate number. Users, or potential users, of products are the target population and should be tested. There is no magic formula for determining the number of consumers who need to be tested. The intent of the test must be considered. If the test is for an initial evaluation to determine if the product has promise, fewer people are necessary than for the final consumer tests of the marketability of the product. For preliminary testing, 75 to 100 consumers usually are adequate to determine if further testing is warranted. Later testing may require 100 to 200 people to determine if the product is continuing to meet expectations. Final testing of the product often requires many more people. If acceptance or preference claims are to be made about the product, the main television networks and the American Society for Testing and Materials suggest that, depending on the claim, 300 to 800 people are required for the consumer test.

One important cautionary note is that it takes more consumers to truthfully show that a product is as good as another product than it does to show that one product is better than another product. Although this may seem odd, the higher number is necessary because it is statistically harder to "prove" equality than it is to "prove" superiority.

In addition to appropriate consumers, all types of quantitative testing require that a questionnaire be developed that is simple to understand and use (Meilgaard et al., 1991; Stone and Sidel, 1993). Also, the tests must be properly controlled from the standpoint of test design and product handling. The tests are too important and cost too much in time and money to inadequately plan the procedures and testing protocols. Meilgaard et al. (1991) provides worksheets for making sure that samples are handled and served properly.

One of the most important, and sometimes the most difficult, aspect of quantitative testing is developing the questionnaire for the test. The simple solution to designing questionnaires is to ask questions that are actionable. Determine in advance what specifically you will do with the information you obtain from the question. For example, if you learn that a specific characteristic of the product is not scoring the way you hoped, what will you do — specifically? Do not just assume you will "change the product somehow;" what specific things will you do? Ultimately, this helps develop the questionnaire by eliminating questions that have no action. Finding that the "texture" is not right does nothing to help you change the texture because it is unknown whether tenderness, juiciness or some other factor was the problem. Specific attributes, rather than general attributes, are necessary. Sometimes problems with attributes can be fixed by adding additional information to the term; e.g. "flavor" can be actionable if you ask "flavor intensity."

A rare but important problem with some consumer questionnaires is that of using attributes that consumers do not understand or understand differently from the researcher. Careful thought must be given to questionnaires to determine that this is not the case. Published accounts of having consumers measure "connective tissue" or "grind" are suspect because consumers may or may not have any idea what those words mean or may interpret them differently from the con-

sumer. For example, chewiness of frankfurters often is considered in the literature as the difference in first and second compressions on a texture instrument; a determination that is heavily dependent on the firmness of the internal meat of the frankfurter. Conversely, consumers may evaluate the chewiness more as it relates to skin strength and awareness of skin in the mouth because that is what they have to chew. The key is to pre-test the questionnaire to determine what, if any, problems are encountered with it before the test is conducted.

Relating Consumer Studies to Other Research

In many cases, consumer studies are undertaken with part of the goal to relate those studies to sensory descriptive tests or to physical and chemical measures. A variety of univariate statistics, such as plots, correlation and simple regression, may be used either to relate liking or individual attributes to other data. More complicated relationships may be studied with multivariate statistics such as plots, multiple regression, cluster analysis, principal components or partial least squares (often simply called PLS) (see Muñoz and Chambers, 1993).

An important consideration for attempting to develop these relationships is that the effects should logically have some suspected relationship to begin with. There is no reason to believe that a single gas chromatographic peak for some flavor measurement will "predict" overall liking of meat so there is little need to conduct an analysis to determine that. Similarly, there is no reason to believe that a compression measurement on a texture instrument will predict "firmness" when consumers bite through a frankfurter with their incisors. Therefore, if no relationship is found, consumers cannot be faulted for not acting like the instrument; the measurements were different and probably should not have been related.

Conclusions

The ultimate users of meat are consumers and what they have to tell us about our products is an important and necessary piece of information for developing new products. The specific consumer tests that are needed in various projects depend on the objectives of the test. Regardless of the specific method, good science must prevail by using adequate numbers of appropriate consumers and by thoughtful preparation of questionnaires and product samples. Poorly conducted consumer tests are worse than no tests at all because poorly conducted studies may give information that leads to the wrong decisions.

References

Chambers, E. IV; Bowers, J. R. 1993. Consumer perception of the sensory qualities in muscle foods. Food Technol. 47(11):116, 118-120.

Chambers, E. IV; Smith, E.A. 1991. The uses of qualitative research in product research and development, In "Sensory Science Theory and Applications in Foods" (eds. Lawless, H. and Klein, B). Marcel Dekker, New York, USA p. 395-412.

Courington, S. M. 1992. Consumer market research tools. Proc. 45th Reciprocal Meat Conf. Fort Collins, CO, USA. pg. 133-136.

Kuntz, L. A. 1994. Wrapped in richness: Creating super-premium foods. Food Product Design 4(2): 85-102.

Meilgaard, M.; Civille, G.V.; Carr, B.T. 1991. Sensory Evaluation Techniques, 2nd edition. CRC Press, Boca Raton, FL, USA. pp. 201-235.

Muñoz, A.M; Chambers, E. IV. 1993. Relating sensory measurements to consumer acceptance of meat products. Food Technol. 47 (11): 128-131, 134.

National Research Council. 1988. Consumer concerns and animal product options. In "Designing Foods," pg. 63-97. National Research Council, National Academy Press, Washington, D.C. USA

Rejtman, J. 1993. Hold the mustard; hot-dog sales drop. Wall Street Journal. July 8, Section B: pg 1.

Stone, H.; Sidel, J.L. 1993. Sensory Evaluation Practices, 2nd edition, Academic Press, New York, USA pp 234-270.

Young, B.J. 1993. Leaning on fat. The National Provisioner. 207(1): 44.

Summary of Discussion

Robert W. Rogers
Chairman, Sensory Evaluation Program Committee

Eight questions were asked and answered during the discussion period which followed the three formal presentations by the panel. The first question related to the ability of panelists to distinguish between myofibular and connective tissue tenderness and then combine those scores into one overall tenderness score. Gail Civille stated that she doubted that in meat where tenderness is attributed to myofibular and connective tissue, as is generally the case in whole-muscle cuts, that panelists could distinguish between the two types of tenderness. She also said that an overall tenderness score would be more reproducible than trying to relate to the two subtopics of tenderness.

The next question related to asking panelists to evaluate tenderness relative to overall acceptability, in addition to the actual rating or ratings for tenderness. Gail said that her advice would be not to do that. Edgar Chambers also made some comments about the two previous questions. He stated that with trained panelists, there are two problems. One is the bias issue from the view of the panelists who have an idea of what is going on in the test so they are likely to tell you what you want to hear about acceptability. Another problem, more significant than the first, is that trained panelists focus on things that are not necessarily the same things consumers consider. He said that he would not state that one could not get good

correlations, but that he did not trust the data to be truly representative of what consumers want because those people, trained panelists, are not really consumers any more. They are trained panelists who are doing specific work on specific products, focusing on attributes that other people may or may not consider. Civille concluded comments on these questions by saying that one would not ask a GC or Instron device to tell which sample they preferred so don't ask a trained panelist to do that either, because in this case they are instruments, not consumers.

The next question was to Chambers about a statement he had made concerning the issue that it takes more observations to prove that a product is equal to another than it does to prove that one product is different from another. He said that it is easier for one person to tell the difference when that difference is obvious than to tell that two things are the same. For example, it's much more believable, if one has an apple and an orange or a steak and a lamb chop, to get a person to say which one they like better (and they can do it repeatedly), than to get them to tell which steak they like the best if the steaks are very similar. When the products are similar, the person is not so sure and the data are not as consistent; so it takes many more observations to arrive at a significant difference. Likewise, when one proves statistically that one product is better ($P<.05$) than another, the testing can stop, but when do you stop testing to say that two products are the same? If two people liked one product and two liked another product, what can you say? You have power in the test. However, if 2,000 people liked one product and 2,000 preferred the other, one has an enormous power to say they are the same. The decision relates to how much alpha and beta risk you are willing to accept from a statistical analysis. Jean Guinard also stated that one may also include the alpha or beta type I, type II errors of accepting the null hypothesis when it is actually false or rejecting it when it is actually true.

The fourth question also related to consumer tests where one wants to test a product on a national level rather than on a regional or ethnic basis. Chambers stated that many products are only going to be successful with certain groups or segments of the population; so you need to test those products in those areas with those people, not the general population. But if the product is to be truly a national product and be sold in all areas and in all ethnic populations, you have to test a broad base of people in several areas and you can't do that with just a few people. Guinard also stated that if you are able to properly define the characteristics of your consumer population (target market), the number of subjects needed to get statistical significance will be greatly reduced. Civille also responded by saying that if you want to sell a product on a national basis, you may just want to sample a population that represents the different groups in the U.S. and not have large groups of consumers within each of the possible sub-test markets. If you just want to have overall general product acceptability, you only have to test populations that represent the U.S. population as you would do in sampling for an election.

The fifth question related to the *Journal of Food Science* editors requiring specific sensory methods to be used in order for them to publish research papers involving sensory analysis. Civille stated that she has not seen anything about that and that she had not been asked to comment about or contribute to those guidelines. Guinard said that there were some old guidelines for submitting papers to the *Journal of Food Science* that include sensory data, but nothing new is required. He also stated that they (IFT) are presently planning to publish some manuals to teach sensory methods and that Hildagon Hayman has been very active in this project.

The next question related to how much one tells potential panelists about the products to be tested. Civille stated that she tells and shows them everything about the products. She said that the key issue is to make sure that the panelists understand that showing differences is important, and being able to describe differences among samples is very important. She also stated that being able to call the same sample when it is duplicated within a set as identical is equally important. She said that the best taster is one who will look at two identical samples and say, "I can't tell any difference between those two samples and if I can't see the difference, there probably isn't one." She also emphasized that the panelists should know that there will not necessarily be a "high fat" control, a blind control, a rancid sample, etc. at each panel session. This tends to eliminate bias and prejudice from the panelists. She concluded her answer by saying that the answer is in the sample, not in what you are told ahead of time or even what is listed on the score sheet. To be a good panelist, one must get into the sample and look at all of the factors, not try to have preconceived ideas about one or two little things. Be a good diagnostician.

The next question was to clarify a previous response about whether one can have a trained consumer panel. The response from Gail Civille was that once you have a trained group, they are no longer truly consumers because they look at things consumers generally do not consider.

The final question related to the cost of properly conducting a consumer panel. Chambers responded by saying that a very simple preference study, including the cost of consumers, products, analysis, product, professional time, etc., depending on the number of consumers used, could cost between $3,000 and $15,000. He also stated that a more complex study of several products in four or five markets could cost many thousands of dollars and that's why studies of this type have to be done properly.