The ABC's of Sensory Evaluation and Consumer Testing

Jennifer Janz
PhD Candidate
University of Alberta

So you're doing some meat research, or maybe you're into product development, or maybe you're just a keener for reading the literature in this area. Do you know the basics of sensory evaluation and consumer testing? I have had skeptical undergraduate students discount sensory evaluation and consumer testing reports, citing the use of untrained panellists ("they couldn't have know what they were doing") or retail surveys ("isn't purchase decision just personal opinion anyway?") as evidence of poorly conducted research. Do you know the difference between trained and untrained panellists, and when and how to use them? Do you know what questions you want to answer with your research and what test procedures will elicit them? Did you know that consumer testing is a science with protocols as rigorous as any laboratory research?

Even if you're not directly involved in sensory evaluation and consumer testing it may be as well to know a few of the basics, so you can be an informed reader of meat quality literature. If you are involved in research, being familiar with these essential aspects will also help you to make informed decisions about testing options provided by a colleague or an outside agency consulting on or providing testing services for your project.

The following discussion will provide just a few ideas to get you thinking, and the references listed at the end are an introduction to the vast body of sensory evaluation and consumer testing literature that you may want to consult for further information. In this short space all aspects of sensory evaluation and consumer testing cannot be covered, but keep in mind or make sure to gather information on the essentials of such things as experimental design, sample presentation design, sample preparation and serving, and the physical testing environment

Sensory Evaluation:

Test Types and Data Handling

There are three basic types of quantitative sensory evaluation tests. Each type is designed to answer a different sort of question. **Difference or discrimination testing** is used to determine if there are differences between products or samples, using tests such as paired comparison, triangle, or duo-trio. The results are determined as frequencies or proportions of correct choices of the test sample from a set of similar or control products. **Descriptive testing** is used to quantify perceived intensities of product characteristics, using a trained panel that agrees upon the attributes to be evaluated and then rates the intensity of each attribute. Common methods include Quantitative Descriptive Analysis (QDA), the Spectrum Method, and the Flavour Profile Method. **Affective or hedonic** testing is used to quantify the degree of liking or preference of a product, by having panellists choose amongst alternatives or rate their degree of liking on a scale which may be structured or unstructured, bipolar or unipolar, and with varying numbers of anchors. The nine-point hedonic scale is commonly used although other systems including purchase intent scaling and preference ranking are available.

Other unique scaling systems exist such as time-intensity scaling, labelled affective magnitude scaling, and rank-rating and can be used to answer more specific questions. With any scaling system, however, it is essential that the scale be constructed on a clear attribute dimension and that all participants understand the meaning of the attribute of the product.

So, the nature of the test type dictates the type of panellists that may be used. In all cases, recruitment is followed by screening for qualities such as preference for the product category (you probably wouldn't invite panellists to evaluate a product they absolutely detest!), frequency of use, sensory acuity and ability to discriminate basic tastes,

availability, and willingness to participate. Detailed instruction in the testing procedures is always provided, however, training in the art and science of discrimination and description is where panellist types differ.

Untrained panellists are used in affective testing to gather hedonic or opinion data. Is the product or certain aspects of it liked? How well is it liked? An untrained panel is meant to capture the opinions of average consumers and is unable to provide detailed descriptive evaluations. That is the role of trained panellists. After months of extensive training this type of sensory panel is able to provide descriptive evaluations of qualities such as taste, odour, flavour (the combination of taste and odour), and texture. These evaluations can be minutely descriptive, for example, they can detail the precise order and intensity of appearance of dozens of individual flavour notes in a single sample. A highly trained sensory panel functions much like a piece of quantitative laboratory equipment and, as such, is not able to provide information reflecting average consumer opinions. Conversely, it is inappropriate to ask untrained panellists to provide descriptive information since they have not been trained and calibrated for this type of assessment.

For completeness, it is worth mentioning the use of expert evaluators. For products such as wine, coffee, and perfume, expert graders or evaluators are used almost exclusively, much as in beef grading to assign value to a commodity. Be aware, however, that product quality assignment in this manner represents an expert evaluation which does not necessarily reflect the needs and wants of the typical consumer population .

Depending on the type of sensory testing employed, different forms of data can be captured. **Nominal data** appear when numbers are used as category designators , and are not associated with a specific numerical value (e.g. 1 = small, 2 = medium, and 3 = large). Mode (the number appearing most frequently) is appropriate for comparison amongst categories. **Ordinal data** are gathered when numbers are used as ranks. In this case the median and percentiles are used for comparisons and non-parametric statistical analysis is conducted on ranked data. **Interval data** are created from scales where the distance between numbers is at equal intervals, although no true zero point exists. Arithmetic means and standard deviations are used for comparisons amongst groups, and parametric statistical procedures such as ANOVA or GLM are used to analyse data unless panellists are considered a fixed effect, in which case a MIXED model would be appropriate (but this topic is still up for debate!). Finally, **ratio data** are generated when scale numbers are placed at equal intervals following a true zero point. Geometric means are used for comparisons with parametric statistical analysis of normalized data .

Despite their reputation as sophisticated or perplexing methods, multivariate statistical analysis techniques are really quite simple to use and can be incredibly informative regarding the relationship amongst sensory characteristics, products, or even panellists. Multivariate techniques provide simultaneous interpretation of many variables to provide an integrated view of the data, whereas univariate techniques examine one influential variable at a time. Being most familiar with SAS software, I have found that this package provides such tools as principal components analysis (PCA) and cluster analysis, which are no more difficult to implement than basic GLM or MIXED models. Principal components analysis chooses "slices" through the data cloud that account for the greatest proportion of variability. This information is related as two-dimensional biplots on which products and/or their attributes are located for examination of the degree of their relationship based on correlation or covariance matrices. Cluster analysis nicely complements PCA, by providing statistical groupings of products or attributes that appeared closely associated in the biplot. Other useful multivariate tools include canonical variates analysis, generalized Procrustes analysis, multidimensional scaling, and response surface methodology.

Consumer Testing:
Some General Concepts

Consumer testing is a logical companion to sensory evaluation and can be applied in situations such as product maintenance, product improvement, new product development, and shelf life testing. In general, consumer testing makes use of untrained individuals, representative of the population of end-product users, to measure such features as liking, preference, purchase intent, and consumption, using methods that help to understand product acceptance and consumer behaviour. While consumer sensory tests focus mainly on sensory acceptability of "blind" products, this type of testing may be taken a step further, into market research that includes product, consumer, and context variables, with use of mathematical modeling to predict consumer purchase behaviour, based on large respondent groups evaluating branded products

Qualitative consumer testing may involve focus groups, in which concepts or products are discussed in a group setting to elicit consumer expectations, feelings, and ideas; in-depth individual interviews; projective techniques, in which feelings and attitudes are expressed in abstract and creative ways such as free association or collaging; or ethnography, with direct observation of consumer behaviour in the consuming environment. Rather than generating hard data, this type of testing provides key observations and insights into consumer behaviour.

Quantitative techniques such as concept or product testing, usage and attitude surveys, simulated test markets, and packaging and advertising tests provide numbers relating to overall product liking and comparisons to existing products, purchase intent, buying and usage habits, attitudes, motivations, and expectations of the product category or specific brands, unmet needs in the marketplace, and packaging preferences and "noticeability" on the store shelf. Participants in consumer testing may be company employees, local residents, or a nationwide consumer sample. Recruitment can be conducted by ways that include intercepts at malls or grocery stores, advertisements, selection from membership lists from community organizations, random telephone solicitation, or established participant database. Screening helps to meet demographic and product usage targets according to the experimental design.

Testing can take place in a variety of locations. Surveys may be conducted from a distance by mail or telephone. A fixed sensory laboratory may be used, or it may be taken on the road as a mobile lab in a bus or van. These types of settings generally increase control over product preparation and consumption, and while this may be good scientific practice, it may hinder the reality of the consumer test. Central location testing is another option and complements recruitment by intercepts in a public place. Once a potential participant has agreed to sample the product they are ushered into a separate area, maybe a vacant mall space set up for sample rating, to complete the survey. Home use testing exerts the least control over the testing environment, but allows participants to prepare and use the product in natural consumption conditions. The objectives of the project will be helpful in determining where along this control/reality gradient testing should be conducted. When, how, and by whom a survey is completed cannot always be assured in home use testing, but this must be balanced against the need for an appropriate sampling context. Laboratory testing can ensure the greatest adherence to sampling protocol, but not everyone wants to sample beer or eat ice cream in a laboratory panel booth first thing in the morning. Other context factors such as convenience, effort, social pressures, and advertising should also be considered.

Consumer sensory testing generally involves some kind of ranking or the use of a scale to evaluate products and provide hedonic ratings. A wide variety of scale types are available including the trusty nine point hedonic, labelled affective magnitude, just about right, food action rating scale, agree/disagree, importance, frequency of use, or price scale questionnaires. The key is to design the ballot specifically for the intended re-

spondents, for example, happy/sad faces for children vs. numerical for adults. It is also important to stick to "need to know" questions and not to be tempted to gather extra tidbits of information just because "the study is going on anyway". This practice risks diverting the focus of the test and tends to add extra time to the survey, which may influence respondent attitudes.

A particularly useful way to analyze consumer testing data is to conduct preference mapping, to explore the relationships between consumer acceptance and sensory evaluation or analytical data from laboratory analysis. Regression analysis, PCA, and response surface methodology are all tools that may be used. For example, internal preference mapping is a specific type of PCA which describes consumer preference ratings across a set of products and indicates the direction and strength of each consumer's preference. Again, cluster analysis of preferences is a complementary technique. Taken one step further, external preference mapping could be conducted by projecting analytical or sensory descriptive data onto the preference map to complete the preference picture. This helps to define how sensory attributes and physical or biochemical characteristics drive product acceptance.

Take Home Message

To answer sensory project objectives with validity the following rules MUST be adhered to:

- The appropriate sensory test method (difference, descriptive, affective) must be chosen to match the project objectives.
- The type of panellist (trained vs. untrained) must be appropriate to the type of testing.
- The sensory task must be appropriate for the type of panellists employed (descriptive for trained, and hedonic for untrained)

If a project budget is limited, and only a single test can be completed, the experimenter must closely examine the objectives of the project and select the method which will deliver the most informative data. Sensory evaluation and consumer testing are valuable tools, but to make the most of these methods they must be applied correctly and paired with appropriate data handling and informative display of the results.

Acknowledgements

Much of the information presented here was gathered during the completion of the comprehensive on-line certificate program in Applied Sensory Evaluation and Consumer Testing available from the University of California at Davis. For more information on the program, check out the web site at:

http://extension.ucdavis.edu/agriculture/sensory

Extended Reading

Giovanni, M. 1983. Response surface methodology and product optimization. Food Technology. 37: 41-45, 83.

Guinard, J.-X., Uotani, B. and Schlich, P. 2001. Internal and external mapping of preferences for commercial lager beers: comparison of hedonic rating by consumers blind versus with knowledge of brand and price. Food Quality and Preference. 12: 243-255.

Kapsalis, J. G. 1987. Objective methods in food quality assessment. CRC Press, Inc. Boca Raton, FL.

Larmond, E. 1977. Laboratory methods for sensory evaluation of food. Agriculture Canada, Research Branch.

Lawless, H. T. and Klein, B. P. 1989. Academic vs. industrial perspectives on sensory evaluation. Journal of Sensory Studies. 3: 205-216.

Meilgaard, M., Civille, G. V. and Carr, B. T. 1987. Sensory evaluation techniques. CRC Press, Inc. Boca Raton, FL.

Moskowitz, H. R. 1983. Product testing and sensory evaluations of foods: Marketing and R&D approaches. Food and Nutrition Press, Inc. Westport, CT.

O'Mahony, M. 1986. Sensory evaluation of food: Statistical methods and procedures. Marcel Dekker, Inc. New York, NY.

Pangborn, R. M., Guinard, J.-X. and Davis, R. G. 1988. Regional aroma preferences. Food Quality and Preference. 1: 11-19.

Schutz, H. G. 1998. Evolution of the sensory science discipline. Food Technology. 52: 42-46.

Schutz, H. G. and Cardello, A. V. 2001. A labeled affective magnitude (LAM) scale for assessing food liking/disliking. Journal of Sensory Studies. 16: 117-159.

Sidel, J. L., Stone, H. and Bloomquist, J. 1981. Use and misuse of sensory evaluation in research and quality control. Journal of Dairy Science. 62: 2296-2302.

Yackinous, C., Wee, C. and Guinard, J.-X. 1999. Internal preference mapping of hedonic ratings for Ranch salad dressings varying in fat and garlic flavor. Food Quality and Preference. 10: 401-409.

Our Students:

Graduate Students in Meat and Meat-related studies in Canada

Student	Degree Level	Estimated Graduation	Institution	Study area	Supervisor
Alex Gill	PhD	2005+	University of Manitoba	Mechanisms of natural antimicrobial action	Dr. Rick Holley
Anas Al Nalbusi	PhD	2005+	University of Manitoba	Lactoferrin action against bacteria	Dr. Rick Holley
Ghandeer Mehyar	PhD	2005+	University of Manitoba	Better poultry carcass washing	Dr. J. H. Han
Parthiban Muthukumarasamy	PhD	2005+	University of Manitoba	*E. coli* O157:H7 survival in dry sausage	Dr. Rick Holley
Michael Peirson	MSc	Graduated	University of Manitoba	Meat Microbiology	Dr. Rick Holley
Pedro Chacon	MSc	2005+	University of Manitoba	*E. coli* O157:H7 survival in dry sausage	Dr. Rick Holley
Haihong Wang	PhD		University of Saskatchewan	Biochemical and physiochemical characteristics of natural actomyosin isolated from pale, soft and exudative and normal pork	Dr. Phyllis Shand
Jennifer Janz	PhD	2004	University of Alberta	3-D modeling of tenderness in semitendinosus and longissimus dorsi	Dr. Mick Price
Tineke Jones	PhD		University of Alberta	Behaviour of *E. coli* around minimum growth temperatures	Dr Lynn McMullen

Please contact Bethany Uttaro (Newsletter Chair) at *uttarob@agr.gc.ca* to contribute to this list.