PN-ABR-145
87179

SUSTAIN

TRAINING: SENSORY EVALUATION II
SHORT COURSE

NOVEMBER 4 - 15, 1993

GUATEMALA

S haring
U nited
S tates
T echnology to
A id in the
I mprovement of
N utrition

A U.S. Private Food Industry initiative
in collaboration with the U.S. Agency for International Development
through a Cooperative Agreement with the National Cooperative Business Association

Upgrading the Food Processing Industries in Developing Countries.

Why SUSTAIN?

SUSTAIN represents a successful collaborative effort between the U.S. food industry and the Agency for International Development (A.I.D.) to upgrade food processing in developing countries. It provides an excellent model for similar private-public sector joint ventures in health, agriculture and other areas of concern to developing countries.

Food processing is a major contributor to development. It serves multiple roles. Food processing can increase the available food supply by extending the life of perishable food products. It can improve the nutritional quality of the diet by making nutritious foods available the year round. It can lead to the growth of related enterprises in transportation, storage, distribution and marketing. And, it can produce much needed foreign exchange by creating value added products both for export and for internal substitution of imported processed foods.

The U.S. food industry has embraced the concept that freely sharing its expertise and knowledge is of mutual benefit to recipient and donor - to the recipient by improving current operations - to the donor by contributing to a healthier global future.

How SUSTAIN Works

A.I.D. missions and trade associations in developing countries publicize SUSTAIN's goals and activities. Executives of U.S. food companies with technical expertise and overall knowledge of the food industry serve as the SUSTAIN Steering Committee, providing guidance and overseeing activities.

Food related companies in developing countries submit their requests to SUSTAIN through the A.I.D. mission or a designated organization in their country. SUSTAIN screens all incoming requests and if necessary asks for additional information. Appropriate U.S. companies are then invited to respond.

Some problems can be readily resolved by providing information. Others require that consultants be sent. When a consultant is sent, the usual assignment is for one to three weeks. Upon completion of the assignment, the consultant prepares a report describing findings and making recommendations. Depending on need, some consultants may return for follow-up visits to ensure that recommendations have been appropriately implemented.

SUSTAIN Helps

Requests are diverse. Help may be needed to solve processing problems, to identify equipment needs and sources of new and used equipment, to train personnel in the use of new equipment and new technologies, to find new uses for indigenous commodities, to establish or improve quality assurance procedures, to control insects and rodents in food processing plants and to improve plant layouts and materials handling.

In the past, U.S. food companies, large and small, have provided technical assistance in the form of information, consultants and training to food processors in Africa, Asia, Latin American and the Caribbean.

SUSTAIN PROGRAM

TRAINING: SENSORY EVALUATION II SHORT COURSE

Guatemala

November 9 - 12, 1993

by

SUSTAIN Volunteer

Ms. Brenda Bravatty, Sensory Evaluation Consultant

This publication was made possible through support provided by the Office of Nutrition, Bureau for Research & Development, U.S. Agency for International Development, under Cooperative Agreement No. DAN-5120-A-00-1066-00.

NCBA/SUSTAIN Project 111.031

TABLE OF CONTENTS

I. INTRODUCTION

A. SUSTAIN

The program Sharing U.S. Technology to Aid in the Improvement of Nutrition (SUSTAIN) provides access to U.S. expertise in food processing to help improve nutrition in the developing world. Technical assistance is provided by volunteer professionals from U.S. food companies, universities, and other organizations who donate their time and expertise. In 1991, the Office of Nutrition of U.S. Agency for International Development awarded the National Cooperative Business Association a $2 million, five-year cooperative agreement to work with SUSTAIN's volunteer leadership to improve, expand, and manage the program.

The assistance SUSTAIN volunteers provide contributes to improved health and nutrition through improved food quality, safety, and availability. In many countries, sufficient food is produced but populations are underserved because much of it goes to waste due to inefficient processing and storage. Improper food handling presents a hazard to human health, and improper waste disposal can contribute to environmental degradation. Strengthening manufacturing practices not only captures scarce resources, but also improves food safety and elevates nutritional status.

B. Definition of Sensory Evaluation

Sensory evaluation is a measure of those attributes of food and materials as they are perceived by the senses of sight, smell, taste, touch, and hearing. A variety of standardized testing procedures has been established to evaluate these attributes, the results of which are used for product evaluation. These tests employ the use of either trained or untrained panelists, depending on the type of test used in the evaluation. There are two major classifications of sensory tests - analytical and affective. The analytical tests are used for laboratory evaluation of products, testing for differences or similarities among a particular product, using either discriminative or descriptive tests. Affective tests are used to evaluate preference and/or acceptance of products. Panelists are not trained and usually have used the product being evaluated.

There are a number of applications of sensory evaluation in product development. Since most new products are imitations or variations of some established standard, product developers need information on the sensory quality and relative acceptability of samples as input for marketability. Initially, analytical tests are used to determine product uniqueness, establish differences among samples, and establish acceptability requirements, that is, whether the samples are equal or better than the standard. Affective tests are then employed to evaluate consumer acceptance and preferences.

Sensory evaluation is an important tool in developing nutritious foods. A nutritious food product must have acceptable taste, texture, flavor, and visual appeal. Consumer acceptance tests can be used to determine whether consumers like the product and their willingness to purchase and consume the product.

C. The Second Sensory Evaluation Short Course in Guatemala

The second Central American Sensory Evaluation Short Course was offered at the Institute of Nutrition of Central America and Panama (INCAP) by SUSTAIN volunteer Brenda Bravatty from November 9 - 12, 1993. The short course was a structured training session for food professionals from the food industry in Central America, and was designed to train them in the correct use of sensory evaluation techniques and the application of these techniques to the processes of product development and quality control.

Twenty-one food professionals participated in the course. Course content included lectures on selected topics with emphasis on difference and acceptance tests, and laboratory practices demonstrating the use of techniques on food products.

Ms. Bravatty gave a test at the beginning and at the end of the course to evaluate knowledge in general concepts of sensory evaluation. At the end of the course, Ms. Bravatty also conducted a course evaluation session in which participants rated the overall quality of the course.

II. REPORT

The second Central American Sensory Evaluation Short Course was offered at INCAP by SUSTAIN volunteer Brenda Bravatty (MSc. Food and Nutrition) from November 9 - 12, 1993. The objective of the course was to train professionals from the food industry in Central America in the correct use of sensory evaluation techniques, with emphasis on difference and acceptance tests. Twenty-one food professionals from four Central American countries participated in the course (Appendix IV).

Ms. Bravatty established contact with INCAP professional two months prior to the course dates. During that time, she put together a laboratory manual and gathered a current bibliography. She also prepared audiovisual aids and a list of products to be evaluated. At INCAP, temporary booths for sensory evaluation were built (in addition to permanent ones).

Ms. Bravatty arrived in Guatemala on November 5th and held meetings with INCAP professionals. She reviewed the program and laboratory procedures for the course. Selection of some products to be evaluated was done in order to have samples of interest to the participants.

The course included lectures of selected topics and laboratory practices. She was assisted by INCAP professionals and two technicians. Two lecture periods and two formal lab practices were held daily. The course schedule is attached in Appendix III. For the statistics session, Dr. Ricardo Sibrian was the guest speaker. He reviewed the different statistical analyses used in the analysis of sensory evaluation tests. He also provided assistance with the use of SAS statistical program for data analysis.

Participants were given a lab manual (Appendix V) which included guidelines and questionnaires for each lab session and current bibliography related to the different topics. They also received the manual "Metodos Sensoriales Basicos para la Evaluacion de Alimentos," published by the International Development Research Center (IDRC) in Canada.

Approximately 40% of the total time was spent in lab practices. The lab work helped the participants to better understand the concepts and to use the methodologies in the evaluation of products of their interest. Participants were welcome to bring their company's products for testing. About 50% of participants brought samples. A wide variety of products was evaluated including orange juice, shrimp, cookies, nacho chips, butter spreads, margarine, chicken nuggets (from two different companies), oil, coffee, apples sauce, chocolate, and ham.

A demonstration session on how to prepare and serve samples and how to set trays was carried out. Samples of juice, ice cream, oil, carbonated drinks, and cookies were served. For each lab session the tasters were given the samples and ballots and were asked to evaluate the different samples. For the acceptance test session, the participants were divided into three groups. Each group developed a questionnaire for acceptance of a particular product (nacho chips, chicken nuggets, and margarine). The questionnaire was then reproduced and all the participants evaluated the three products. After each lab session, a detailed statistical analysis of results was carried out. A written report of the results for each test was written by the participants.

There was good interaction between the participants and the instructor. The participants were motivated and very interested in any material available in the area. Most of them were interested in advanced sensory courses and some in technical assistance for specific needs in their companies.

At the beginning and at the end of the course, a test to evaluate knowledge in general concepts of sensory analysis was given. Using the same exam for testing before and after the course provides an indication of how much the participants learn from the course. The questionnaire and results are shown in Appendix VI. In the initial evaluation the mean score was 29 and in the final evaluation the score was 70. An increase in knowledge of sensory evaluation was achieved by all participants.

In a different questionnaire given at the end of the activity, the participants rated the organization, instruction quality, and laboratory practices of the course. The results are presented in Appendix VII. The coordination of the course and the organization of laboratory practices were rated as "excellent" by 55% of participants and "good" by 40%. The presentation of the lectures was rated as "excellent" by 65% and "good" by 35% of the participants. The tasks developed by the instructor were rated as "excellent" by 85% of the participants. The relationship between lectures and laboratory practices was rated as "excellent" by 75% of the participants. Forty-five percent of participants indicated that organization during lab practices was "excellent", 40% said it was "good". A suggestion for strengthening the course is to bring the instructor into the country several days before the course to allow for more interaction between the instructor and the technicians.

When asked about topics of their interest for future courses, most of the participants mentioned descriptive analysis, panel training, and/or specific techniques for different products such as oil, dairy, and meat products. They would like to learn more practical concepts of statistics. About 95% of them are interested in a more advanced course in sensory evaluation. This will be useful to solve problems related to product development and to acquire more knowledge in the area for their professional development and for the benefit of their companies and the food products of Central America.

APPENDIX I

SUSTAIN Description

The program **Sharing U.S. Technology to Aid in the Improvement of Nutrition (SUSTAIN)** provides access to U.S. expertise in food processing to help improve nutrition in the developing world. Technical assistance is provided by volunteer professionals from U.S. food companies, universities, and other organizations who donate their time and expertise.

SUSTAIN was granted a five-year renewal from the U.S. Agency for International Development (USAID) on September 30, 1991. The program is managed under a cooperative agreement with the National Cooperative Business Association (NCBA) and receives advice from a Steering Committee made up of private sector representatives.

NCBA was founded in 1916 and is a membership association representing America's 45,000 cooperative businesses. Known overseas as CLUSA, NCBA works overseas with its own member co-ops, USAID, World Bank, UNDP, and other donor agencies to promote development and joint ventures in the third world.

Many benefits can accrue to the developing world through improvements in food processing. From the standpoint of alleviating hunger and improving nutrition, food processing has much to offer. It helps meet food and nutritional requirements and reduce post-harvest food losses. From the economic standpoint, food processing provides a means for increasing foreign exchange earnings through exporting value-added processed foods rather than commodities. It helps generate employment and stimulates technological development and the growth of allied industries.

SUSTAIN helps improve food quality, expand production, and lower operating costs of locally grown and processed foods by providing technical assistance in post-harvest food systems, including: (a) food safety, quality, and sanitation (b) food preservation and storage (c) food processing (d) food fortification (e) packaging (f) marketing (g) weaning foods and (h) environmental technologies.

How the Program Works

SUSTAIN receives requests for assistance from individual food companies, research institutions, and USAID. Short-term technical assistance is provided by experienced U.S. professionals who donate their time and expertise to the project. Missions are typically one to three weeks in duration. SUSTAIN covers international travel costs. Companies or host organizations requesting SUSTAIN assistance are asked to contribute towards in-country expenses. Due to budget constraints, priority is given to requests that can demonstrate an ability to improve the nutritional quality, safety, and availability of food in the local community. To the extent possible, SUSTAIN coordinates its overseas activities through a local organization. This not only enhances opportunities for technology transfer, but also facilitates coordination of activities and contributes to long-term sustainable development.

SUSTAIN is able to solve many problems by providing information that exists either in technical literature or in the "memory" of a company. If the problem cannot be solved through correspondence, then SUSTAIN volunteers may be sent to provide short-term technical assistance. Workshops and seminars can also be organized to help address food technology issues. The program does not fund product or equipment acquisitions.

The program publishes a quarterly newsletter (*SUSTAIN Notes*) on food technology issues. It is provided gratis to over 1900 recipients in more than 50 countries.

For more information, please write to:

SUSTAIN Program
National Cooperative Business Association
1401 New York Avenue, NW, Suite 1100
Washington, DC 20005-2160
Phone: (202) 638-6222
Fax: (202) 628-6726

APPENDIX II

Biography of SUSTAIN Volunteer

BRENDA BRAVATTY, M.Sc. (Foods & Nutrition, University of Manitoba, 1988) is a sensory analyst and consultant. Ms. Bravatty works with NutraSweet to provide sensory evaluation support to specific products. Previously she worked with RQA, Inc. (a company specializing in quality evaluation of food and beverage products) to plan and organize their sensory evaluation department, including helping to design the facilities and selecting and training the panelists. She also worked for the Institute of Nutrition of Central America and Panama as the Manager of the Sensory Evaluation Department and coordinator and instructor of sensory evaluation workshops and consumer surveys. Ms. Bravatty has also translated textbooks and written workshop manuals and workbooks in the sensory evaluation field. As a SUSTAIN volunteer, she taught sensory evaluation short courses in Guatemala.

APPENDIX III

Short Course Curriculum

SEGUNDO CURSO CENTROAMERICANO DE ANALISIS SENSORIAL

PRESENTADO POR

INSTITUTO DE NUTRICION DE CENTRO AMERICA Y PANAMA (INCAP/OPS)

PROYECTO SHARING UNITED STATES TECHNOLOGY TO AID IN THE IMPROVEMENT OF NUTRITION (SUSTAIN)

ASOCIACION GUATEMALTECA DE TECNOLOGOS EN ALIMENTOS (AGTA)

Project SUSTAIN

AGTA

Guatemala, 9-12 de noviembre de 1,993.

SEGUNDO CURSO CENTROAMERICANO DE ANALISIS SENSORIAL

PROGRAMA

FECHA/HORA	ACTIVIDAD/TEMA	RESPONSABLE
Martes 9 de noviembre		
8:30-9:00	Inscripción Entrega de materiales	Srita. Claudia Pereira
9:00-9:15	Introducción	Dr. Luiz Elías
9:15-9:30	El Análisis Sensorial en la Industria de Alimentos Centroamericana	Ing. Ana Miriam Obregón
9:30-11:00	Análisis Sensorial: *Generalidades *Usos Fisiología de la percepción: *Olfato	Lic. Brenda Bravatty
11:00-12:15	Laboratorio 1A: Identificación de olores	Lic. Brenda Bravatty Técnicas de laboratorio
12:15-13:30	Almuerzo	
13:30-15:00	Fisiología del gusto: *Evaluación de textura *Umbrales de percepción	Lic. Brenda Bravatty
15:00-16:30	Laboratorios 1B, 1C y 1D	Lic. Brenda Bravatty Técnicas de laboratorio.
Miércoles 10 de noviembre		
8:30-10:30	Requisitos para llevar a cabo análisis sensorial	Lic. Brenda Bravatty
10:30-10:45	Coffee break	Srita. Claudia Pereira

10:45-11:45	Demostración de laboratorio: Preparación de muestras, bandejas, instrucciones y cuestionario.	Lic. Brenda Bravatty Técnicas de laboratorio
11:45-12:15	Lectura. Trabajo en grupo	Lic. Brenda Bravatty
12:15-13:30	Almuerzo	
13:30-14:45	Pruebas de Análisis Sensorial	Lic. Brenda Bravatty
14:45-16:00	Pruebas estadísticas	Dr. Ricardo Sibrián
Jueves 11 de noviembre		
8:30-10:30	Pruebas de diferencia	Lic. Brenda Bravatty
10:30-10:45	Coffe break	
10:45-12:15	Laboratorio 2	Lic. Brenda Bravatty
12:15-13:30	Almuerzo	
13:30-14:30	Resultados de pruebas de diferencia	Lic. Brenda Bravatty
14:30-16:30	Análisis Descriptivo: Teoría y laboratorio	Lic. Brenda Bravatty Técnicas de laboratorio
Viernes 12 de noviembre		
8:30-10:30	Pruebas de aceptabilidad	Lic. Brenda Bravatty
10:30-10:45	Coffe break	Srita. Claudia Pereira
10:45-12:15	Laboratorio Elaboración de cuestionario	Lic. Brenda Bravatty
13:30-15:00	Análisis de Datos Resultados Escribir reporte	Lic. Brenda Bravatty
15:00-16:00	Comentarios finales Discusión en grupo	Lic. Brenda Bravatty
16:00	Clausura	Dr. Hernán Delgado

Técnicas de laboratorio: Sra. Albertina de Cifuentes
Sra. Lucrecia de Polanco

EL INSTITUTO DE NUTRICION DE CENTRO AMÉRICA Y PANAMÁ (INCAP/OPS)

LA ASOCIACIÓN GUATEMALTECA DE TECNOLOGOS EN ALIMENTOS (AGTA)

Y EL PROYECTO SUSTAIN

Presentan el Segundo Seminario Centroamericano de:

ANÁLISIS SENSORIAL

Project SUSTAIN AGTA

INTRODUCCION

Con el objeto de apoyar a la industria alimentaria de Centroamérica y Panamá, el Instituto de Nutrición de Centro América y Panamá (INCAP), la Asociación Guatemalteca de Tecnólogos en Alimentos (AGTA) y el Proyecto Sustain organizan el Segundo Seminario Centroamericano de Análisis Sensorial

El análisis sensorial es una disciplina científica que se utiliza para medir las características de los alimentos a través de los sentidos. La industria alimentaria moderna aprovecha la información obtenida a través del análisis sensorial para el diseño y desarrollo de nuevos productos, así como para la determinación de la aceptabilidad, control de calidad de los productos ya existentes y como un apoyo al departamento de mercadeo.

DIRIGIDO A

Profesionales que trabajan en el área de producción, control de calidad y mercadeo en la industria centroamericana de alimentos, que tengan poca o ninguna experiencia en el uso de las pruebas de análisis sensorial.

OBJETIVOS

1. Describir los métodos sensoriales básicos, sus ventajas, usos y el análisis de sus resultados
2. Dar a conocer la importancia de la evaluación sensorial dentro del proceso de producción, mercadeo y consumo de alimentos.
3. Capacitar al personal de la industria de alimentos y docentes universitarios en el uso adecuado de las metodologías utilizadas en la evaluación de características sensoriales y aceptabilidad de alimentos

METODOLOGIA

La exposición del contenido será por medio de intervenciones magistrales, discusiones de grupo, revisiones bibliográficas y prácticas diarias de laboratorio.

CONTENIDO

1 Introducción a la evaluación sensorial.

2 Descripción detallada de pruebas de diferencia y aceptabilidad, usos, ventajas, análisis de datos e interpretación de resultados

3 Breve descripción de análisis descriptivo, sus usos.

4 Selección y entrenamiento de jueces para pruebas sensoriales y como motivar y monitorizar su rendimiento.

5 Uso de las diferentes metodologías para resolver problemas específicos en la industria de alimentos

6 Cómo preparar informes de pruebas sensoriales.

Los participantes podrán proporcionar muestras de uno de sus productos para ser evaluados durante el curso (favor indicarlo al momento de inscribirse)

SEDE Y FECHA

La sede del curso será el Instituto de Nutrición de Centro América y Panamá (INCAP), Carretera Roosevelt, zona 11 Ciudad de Guatemala, Guatemala.

El curso tendrá una duración de cuatro días comprendidos del 9 al 12 de noviembre de 1993.

COSTO E INSCRIPCION

El curso tendrá un costo de US$ 300 00 (Q1,740.00) por participante. Este valor incluye el derecho al uso de las instalaciones y laboratorio, material bibliográfico y equipo a utilizarse en el laboratorio, almuerzos durante cuatro días y transporte local. Los gastos de hotel no están incluidos.

La inscripción para el curso se cierra el 26 de octubre de 1,993. Inscripciones después de esa fecha tendrán un costo adicional de US$ 20 00. Si cancela con cheque, favor hacerlo pagadero a INCAP (favor llevar cheque personalmente o enviarlo vía courier).

Inscripción e información comunicarse con:

Programa de Mercadeo y Servicios Técnicos
INCAP
c/o Ing. Ana Miriam Obregón
Teléfono PBX (502-2) 719912
Facsímil (502-2) 736529
Correo electrónico
RFLORES@UCRVM2.BITNET

Para inscribirse incluir. nombre y apellido, institución o empresa donde labora, profesión, cargo que desempeña, dirección, teléfono, fax. Para participantes extranjeros, por favor incluir la fecha de llegada a Guatemala, la línea aérea y el vuelo

INSTRUCTORA DEL CURSO

Lic. Brenda Bravatty
MSC. en Alimentos y Nutrición
Consultora de Análisis Sensorial, EEUU

La Lic. Bravatty obtuvo su postgrado en Análisis Sensorial de Alimentos en la Universidad de Manitoba, Winnipeg, Canadá. Actualmente trabaja como consultora para compañías como Nutra Sweet, Tropicana y otras, además de colaborar en proyectos de investigación en la Universidad de Illinois. Su experiencia en el uso de diferentes metodologías de evaluación sensorial en una amplia variedad de productos será compartida en este curso.

PERFIL DEL EGRESADO

Al finalizar el curso los participantes serán capaces de

1. Definir el análisis sensorial de alimentos
2. Comprender y explicar los procesos fisiológicos de percepción en el ser humano
3. Identificar los requisitos mínimos necesarios para llevar a cabo pruebas de análisis sensorial
4. Describir las pruebas sensoriales más importantes que se utilizan en la evaluación de muestras
5. Establecer la importancia de los métodos de análisis sensorial dentro de los procesos llevados a cabo por la industria de alimentos
6. Seleccionar a los jueces más apropiados para la evaluación de las muestras en las diferentes pruebas sensoriales
7. Analizar e interpretar adecuadamente, a través de métodos estadísticos, los datos obtenidos en pruebas sensoriales.
8. Conocer la literatura y bibliografía más reciente en el área de análisis sensorial
9. Planificar, organizar y realizar, a nivel de laboratorio, la mayoría de las pruebas sensoriales aprendidas

APPENDIX IV

List of Participants

Jeanny E. Zimeri V.
Empacadora Toledo, S.A.
Avicola Billalobos
Guatemala
Ph: 773344/773608

Ana Lucrecia Urizar H.
Alimentos Centroamericanos S.A. (ALCASA)
1a Calle 36-07 Zona 11
Guatemala
Ph: 912167

Liza Margarita Pira
Pollo Campero
50 calle 16-03 Zona 12
Guatemala
Ph: 761050/70

Orland Buitrago Rodriguez
Compania Industrial Lido-Pozuelo
cd. San Francisco, calle al Country Club
Comayaguela, Honduras
Ph: 336636
Fax: 340762

Paulina Wittkowski
Central Distribuidera
3 Ave. Final Norte
Finca el Zapote Zona 2
Cludad, Guatemala
Ph: 512777

Javier Fernandex Schwartz
Osmosis
19 calle 12-S2 Zona 10
Guatemala
Ph: 335335
Fax: 335336

Liz Murillo Gomez
Numar
B Cuba San Jose Costa Rica
Apartado postal 3657-1000
San Jose, Costa Rica
Ph: 239666
Fax: 506-552892

Nelson Omar Gonzalez Aguilera
Alimentos del Valle
Palmerola, Comayagua
Honduras Apdo #121
Ph: 721382-84
Fax: 720204

Lisbeth Villaroto
Salveu S.A.
33 Calle 0-73 Zona 3
Guatemala

Victoria Monje de la Cerda
Alimentos, S.A.
Km 15 Carretera a San Jose Pinula
Guatemala City, Guatemala
Ph: 0341262
Fax: 0341816

Marissa Ivonne Umana Velasquez
Alimentos, S.A.
Km 15 Carretera a San Jose Pinula
Guatemala City, Guatemala
Ph: 0341262 y 66
Fax: 0341816

Indra Mara Flores Zanboni de Reyes
Fabrica Venus
4a Calee 10-65 Zona 1
Fabrica bAV 4-90 Zona 1
Gautemala City, Guatemala
Ph: 24192

Pedro Antonio Rodgriquez Zavala
Alimentos del Valle
Palmerola, Comaygua
Honduras Apdo #121
Ph: 721382-84
Fax: 720204

Martha Lita Caceres Flores
Alimentos del Valle
Palmerola, Comayagua
Honduras Apdo #121
Ph: 721382-84
Fax: 720204

Patricia Guadelupe Bernal
Pollo Campero S.A. de C.V.
Avenida Olivpica y Antiqua
Calle a Santa Ceda #8544
San Salvador, El Salvador
Ph & Fax: 985166

Sonia del Carmen Maradiago de Sarminto
Granja Marinas San Bernardo
B[a] Cabanas Cholutica
Honduras
Ph:820822

Carlos Gustavo Reyes
DEMAHSA
Antiqua Locol de Chico Club
Tequicgalpa, Honduras
Ph: 721355
Fax: 720970

Iris Janeth Lopez de Quan
Quimicas Magna
Boulevard Suyapa Frente Comisanato Popular
Tegucigalpa, Honduras
Ph: 31077478 328070 al 73
Fax: 326127

Nidia Noemu Fajardo Del Gid
Gragas Y Aceites S.A.
2a Av. 5-16 Zona 9
Guatemala
Ph: 362671, 348337
Fax: 362671, 348337

Rosalina Lopez
Bimbo de Centro America
Quintas Las Victorias Lote 265
Chimaltenango, Guatemala
Ph: 0391592
Fax: 0391569

Frank R. Tosta
Alimentos del Volle
Palmerola, Comayagua
Honduras Apdo #121
Ph: 721382-84
Fax: 720204

APPENDIX V

Laboratory Manual

SEGUNDO CURSO CENTROAMERICANO DE ANALISIS SENSORIAL

Instituto de Nutrición de Centro América y Panamá
(INCAP)

Guatemala, 9 al 12 de noviembre de 1993

Instructora General:

Licda. Brenda Ríos de Bravatty, Analista Sensorial

Consultora E.E.U.U.

Técnicos:

Sra. Albertina de Cifuentes, INCAP

Sra. Lucrecia de Polanco, INCAP

Ing. Ana Miriam Obregón, Ing. Química. USAC

Sensory Evaluation Guide for Testing Food and Beverage Products

by the Sensory Evaluation Division of the Institute of Food Technologists

□ SENSORY EVALUATION has been defined as "a scientific discipline used to evoke, measure, analyze, and interpret reactions to those characteristics of foods and materials as they are perceived by the senses of sight, smell, taste, touch, and hearing" (IFT, 1975). Key questions of methodology confront anyone who undertakes the sensory evaluation of foods. Which test method is most appropriate? What experimental design is most efficient? What statistical treatment is most applicable?

To develop standardized testing procedures, the Committee on Sensory Evaluation of the Institute of Food Technologists in 1964 published a "Sensory Testing Guide for Panel Evaluation of Foods and Beverages" (IFT, 1964). Prell (1976) reviewed and revised that information to include current procedures, test methods, applications, and references. The information presented in this guide represents a further revision and expansion of these versions. This guide is designed to serve as a reference for individuals working in the field of sensory evaluation and to promote standardization and consistency of test procedures and results.

For supplementary information on physical conditions of testing, e.g., testing area, sample preparation, and sample presentation, see Amerine et al. (1965), ASTM Committee E-18 (1968; 1973), Stahl and Einstein (1973), and Larmond (1977). For information on experimental designs for sensory tests, refer to Cochran and Cox (1957), Amerine et al. (1965), Winer (1971), Harrison (1972), and Sidel and Stone (1976). For a glossary of standard definitions of terms relating to sensory evaluation, see ASTM Committee E-18 (1978).

The IFT Sensory Evaluation Division has also published recommended guidelines for the preparation and review of papers reporting sensory evaluation data (IFT, 1981)—a revision of the IFT (1971) author guidelines.

This guide was prepared by a committee of the IFT Sensory Evaluation Division consisting of A.E. Dethmers (chairman), G.V. Civille, J.M. Eggert, J.P. Erhardt, R.C. Hootman, K. Jehle, R. A. Kluter, P. Low, H.R. Moskowitz, R.M. Pangborn, D.R. Peryam, J. Powers, P.A. Prell, K. Preziosi, J.B. Reaume, N.C. Rodriguez, E.Z. Skinner, J. Speight, H. Stone, S. Snyder, L.S. Tanno, C. Tassan, and S.W. Whitlock.

Each member of the IFT Sensory Evaluation Division will automatically receive one reprint of this guide from the Division free of charge. Nonmembers of the Division may obtain reprints for $1.00 each, prepaid—make the check payable to "Institute of Food Technologists" and send it with a request for a reprint of "Sensory Evaluation Guide for Testing Food and Beverage Products" to IFT Reprint Dept., Lock Box 91332, Chicago, IL 60690.

Types of Applications

It is the responsibility of the sensory scientist to select and implement proper testing methodologies. A thorough understanding of the experimenter's objective(s) is essential. The most commonly occurring industrial applications are as follows:

• **New Product Development.** Some new products are unique (i.e., there is no prototype), but most new products are imitations or variations of some established standard. In either case, product developers need information on the sensory quality and relative acceptability of experimental prototype samples as input for marketability.

Sensory evaluation of a new product may involve the following test sequence:

1. Characterization of product prototype samples to determine uniqueness or a "point of differentiation" from related established products.
2. Evaluation of the experimental prototype samples to establish whether differences exist among them (or between the prototype samples and a standard).
3. Determination of whether the prototype samples meet the acceptability requirements established for the product (e.g., whether they are equal to or better than the standard).

• **Product Matching.** Duplicating a standard (a prototype, a competitor's product, etc.) requires a sensory testing sequence similar to that for new product development. The objective of product matching, however, is to verify that there is no difference between the standard and the experimental product, and that the two have equivalent acceptability.

• **Product Improvement.** Real improvement of a product can be measured in a number of ways. The following is a logical testing sequence:

1. Difference tests to determine whether the experimental product is different from the control (if it is not different, it cannot be better).
2. Affective tests, if products differ, to establish whether the experimental product is liked more than the control (i.e., represents an improvement).

• **Process Change.** A process change should maintain or improve the product. The testing sequence is similar to that for product improvement:

1. Difference tests to determine whether the experimental product is different from the control (if it is not different, it cannot be poorer.)
2. Affective tests, if products differ, to establish whether the experimental product is liked as well as or more than the control.

• **Cost Reduction and/or Selection of a New Source of Supply.** A successful cost-reduction program based upon lower-priced ingredients, a lower-cost process, or production in a different location must yield an end product comparable to the product formerly produced. Change to a new supplier of raw materials should also result in an end product comparable to the standard or control. The sequence of sensory evaluation in these situations usually is as follows:

1. Difference tests to establish whether the experimental product is different from the control (if it is not different, it must be as good as the control).
2. Affective tests, if products differ, to establish whether the experimental product is liked as well as or more than the control.

• **Quality Control.** Quality control procedures are used during production, distribution, and marketing to ensure that the end product is as good as the standard. Representative samples are usually evaluated as follows:

1. Difference tests to determine whether the sample is different from the standard (if it is not different, it must be as good as the standard.)
2. Descriptive tests, if the sample is found to be different, to indicate how the sample differs from the standard. Results of these tests may be used to guide remedial action, such as changes in processing procedures.

• **Storage Stability.** Product stability during transportation, warehousing, and retailing and during storage in the home is essential to consumer satisfaction. To establish information on product shelf life, representative samples are obtained, evaluated initially, and subjected to controlled storage conditions for subsequent tests. At specific time intervals, storage samples are withdrawn and evaluated, generally in comparison with a control. The control must be of the same production lot or batch as the test samples and must be held under conditions known to maintain the original quality. Sensory tests to determine product storage stability may include the following:

1. Difference tests to determine whether the storage samples are different from the control (if no significant difference is found, product stability is assumed).
2. Descriptive tests, used alone or in conjunction with difference tests, to characterize and/or quantify the changes that may have occurred during storage. Descriptive analysis is frequently used in situations where maintenance of a control is unrealistic.
3. Acceptance tests to determine the relative acceptance of stored products.

• **Product Grading or Rating.** Product grading or rating requires an accurate classification of the samples according to grade standards defined for the product, as well as an evaluation of samples in relation to each other. Although grading may be done by chemical or instrumental analyses, discussion in this guide will be limited to grading by sensory evaluation. Category scoring or ratio scaling based on the presence and intensity of selected characteristics may be used to measure samples against standard specifications set for the product.

—Continued on next page

• Consumer Acceptance and/or Opinions. After laboratory screening, it may be desirable to submit product to a central-location or home-placement test to obtain consumer reaction. Acceptance tests will indicate whether the current product can be marketed or improvement is needed.

• Consumer Preference. Preference tests among consumers of the product can be used to determine which sample is preferred. Preference screening tests are often conducted with employee panelists. Although employee tests may not represent a random sampling of the target population, directional information may be gained for designing subsequent consumer tests.

• Panelist Selection and Training. The selection of appropriate individuals for participation in trained (analytical) panels is essential to effective panel performance. Initial training can usually be accomplished during the selection process. The methods most frequently used to select and train panelists in the laboratory are as follows:

1. Sensitivity tests to determine recognition of basic tastes. Although there is no evidence that sensitivity to sweet, sour, salt, and bitter stimuli is related to performance on sensory panels, the panelists (judges) should be able to differentiate unequivocally among these sensations to avoid confusion of taste terminology.

2. Difference tests to determine ability to detect specific variations of the test product and to generate reproducible results. The product variations illustrated during the selection process should be similar to those which may be encountered during the actual operation of the panel.

3. Descriptive tests to determine ability to measure differences and to generate reproducible results. As with the difference tests, the product variations illustrated during the selection process should be similar to those which may be encountered during the actual operation of the panel, and the rating scale used should be the same as that which will be used during the actual panel operation.

• Correlation of Sensory with Chemical and Physical Measurements. Because sensory response to a product is of concern to the developer, it is essential to know how any chemical or physical method used compares with the human senses, i.e., the panel's ability to detect and quantify sensory characteristics. Descriptive test methods with trained panelists are generally used for this purpose.

Types of Tests

There are two major classifications of sensory tests—analytical and affective (Table 1):

Analytical tests are used for laboratory evaluation of products in terms of differences or similarities and for identification and quantification of sensory characteristics. There are two major types of analytical tests—discriminative and descriptive. Both employ experienced and/or trained panelists. Potential panelists are screened for selected personal traits, interest, and ability to discriminate differences and generate reproducible results. Training further familiarizes the panelists with test procedures and increases their ability to recognize, identify, and recall sensory characteristics.

Affective tests are used to evaluate preference and/or acceptance of products. Generally, a large number of respondents is required for such evaluations. These panelists are not trained, but are selected at large to represent target or potential target populations. Panel members are selected in accordance with a number of criteria, which frequently include: (1) previous use of the product, (2) size of family or age of specific family members, (3) occupation of head of household, (4) economic or social level, and (5) geographic area.

The various analytical-discriminative, analytical-descriptive, and affective tests are described below and in Table 2, and the specific tests for each type of application discussed in the previous section are described in Table 3.

ANALYTICAL-DISCRIMINATIVE TESTS

There are two types of discriminative tests—difference and sensitivity. Difference tests measure whether samples can be differentiated at some predetermined level of statistical probability, e.g., $p < 0.05$. Sensitivity tests measure the ability of individuals to detect sensory characteristics.

• Difference Tests. There are several types of difference tests:

1. *Paired-Comparison* (Test 1 in Table 2). Two coded samples are evaluated simultaneously or sequentially in a balanced order of presentation. There are two variations of this test:

a. Simple Difference. The judge indicates whether there is a difference between the samples. The judge is told beforehand that the samples in each trial set to be tested may be identical or different. Complete randomness of presentation is essential so that the panelist responds to each trial set independently.

b. Directional Difference. The judge chooses the sample within each pair that has the greater amount of a specified characteristic. A forced choice (no indeterminate answers) is required.

The number of pairs presented at one session is limited by fatigue or adaptation induced by extensive sensory testing. The chance probability of selecting one sample over another within individual pairs is one-half.

BEST AVAILABLE

Table 2—INFORMATION GUIDE FOR SENSORY EVALUATION METHODS

Method	No. of samples per test	Analysis of data	Selected references for analysis of data[a]
1. Paired-comparison (or paired-preference)	2	Binomial distribution	Gridgeman (1955; 1961); Peryam (1958); Roessler et al. (1978); Scheffe' (1952)
2. Duo-trio	3 (2 identical, 1 different)	Binomial distribution	Gridgeman (1955); Lockhart (1951); Peryam (1958); Roessler et al. (1978)
3. Triangle	3 (2 identical, 1 different)	Binomial distribution	Byer and Abrams (1953); Gridgeman (1955); Kramer and Twigg (1962); Peryam (1958); Roessler et al. (1978)
4. Ranking	2-7	Rank analysis Analysis of variance	Kahn et al. (1973), Kramer (1960; 1963); Snedecor and Cochran (1967); Tompkins and Pratt (1959)
5. Rating difference/scalar difference from control	1-18 (the larger number only if mild-flavored or rated for texture only)	Analysis of variance Rank analysis	Mahoney et al. (1957); Peryam (1958); Snedecor and Cochran (1967); Tukey (1949); Wiley et al. (1957)
6. Threshold	5-15	Sequential analysis	ASTM Committee E-18 (1968); Green and Swets (1971); Gregson (1962); Guilford (1954); Pilgrim et al. (1955); Wald (1947).
7. Dilution	5-15	Sequential analysis	Bohren and Jordan (1953); Tilgner (1962a; b); Wald (1947)
8. Attribute rating (category scaling; and ratio scaling or magnitude estimation)	1-18 (the larger number only if mild-flavored or rated for texture only)	Analysis of variance Rank analysis Regression analysis Factor analysis Graphic presentation	ASTM Committee E-18 (1972a); Carlin et al. (1956); Duncan (1955); Kramer and Twigg (1962); Mahoney et al. (1957); Snedecor and Cochran (1967); Wiley et al. (1957); Moskowitz (1974; 1975; 1978); Moskowitz and Barbe (1976); Moskowitz and Wehrly (1972); Moskowitz and Sidel (1971); Stevens (1962); Winer (1971)
9. Flavor profile analysis	1	Graphic presentation Principal components and multivariate analysis of variance	Caul (1957); Little (1958); Sjöström and Cairncross (1954); Sjöström et al. (1957); Kendall and Stuart (1968); Morrison (1976)
10. Texture profile analysis	1-5	Graphic presentation Principal components and multivariate analysis of variance	Brandt et al. (1963), Civille and Szczesniak (1973); Szczesniak et al. (1963), Kendall and Stuart (1968); Morrison (1976)
11. Quantitative descriptive analysis	1-5	Analysis of variance Factor analysis Regression analysis Graphic presentation	Stone et al. (1974)
12. Hedonic (verbal or facial) scale rating	1-18 (the larger number only if mild-flavored or rated for texture only)	Analysis of variance Rank analysis	ASTM Committee E-18 (1972b), Ellis (1966), Gridgeman (1961); Hopkins (1950); Peryam and Pilgrim (1957)
13. Food action scale rating	1-18 (the larger number only if mild-flavored or rated for texture only)	Analysis of variance Rank analysis	Schutz (1965)

[a]Additional general references: Amerine et al. (1965); ASTM Committee E-18 (1968); Drake and Johansson (1969; 1974); Ellis (1967), IFT (1964); Larmond (1977), Sidel and Stone (1976)

Table 1 — CLASSIFICATION OF SENSORY EVALUATION METHODS AND PANELS[a]

Classification of methods by function	Appropriate methods[b]	Type and No. of panelists[c]
ANALYTICAL: Evaluates differences or similarity, quality and/or quantity of sensory characteristics of a product		• Screened for interest, ability to discriminate differences and reproduce results • Trained to function as a human analytical instrument
1. Discriminative:		• Normal sensory acuity • Periodic requalification • Panel size depends on product variability and judgment reproducibility • No recommended "magic number"—a number often used is 10; a recommended minimum number is generally 5, since any fewer could represent too much dependence upon one individual's responses
a. *Difference:* Measures simply whether samples are different	Paired-comparison Duo-trio Triangle Ranking Rating difference/scalar difference from control	
b. *Sensitivity:* Measures ability of individuals to detect sensory characteristic(s)	Threshold Dilution	
2. Descriptive: Measures qualitative and/or quantitative characteristic(s)	Attribute rating Category scaling Ratio scaling (magnitude estimation) Descriptive analysis Flavor profile analysis Texture profile analysis Quantitative descriptive analysis	
AFFECTIVE. Evaluates preference and/or acceptance and/or opinions of product	Paired-preference Ranking Rating Hedonic (verbal or facial) scale Food action scale	• Randomly selected • Untrained • Representative of target population • Consumers of test product • No recommended "magic number"—minimum is generally 24 panelists, which is sometimes considered rough product screening; 50-100 panelists usually considered adequate

[a]Suggested references: Abbott (1973), Amerine et al. (1965), AMSA (1978); ASTM Committee E-18 (1968; 1979; 1981); Brandt et al. (1963), Cave (1957), Dawson et al. (1963); Ellis (1966, 1967; 1970); Hash (1974, 1975b), IFT (1964; 1979); Larmond (1977), Martin (1973), *Pilgrim and Peryam* (1958), Schutz (1971); Sjöström et al. (1957), Stahl and Einstein (1973); Stone et al. (1974)

[b]See Table 2

[c]Ignoring these considerations increases the possibility of bias and error in the results

2. *Duo-Trio* (Test 2). This test employs three samples, two identical and one different. One sample is identified as the standard and presented first, followed by two coded samples, one of which is identical to the standard. The judge is required to identify the sample which matches the standard. The sample used as the standard may be constant or alternated. As with the paired-comparison and triangle tests, a forced choice is required and statistics may be applied to determine significance. The chance probability of selecting the matching sample is one-half.

3. *Triangle* (Test 3). This test employs three coded samples, two identical and one different, presented simultaneously. None of the samples is identified as the standard. Control and experimental treatments are systematically varied so that each is presented in odd and identical sample positions an equal number of times. The judge must determine which of the three samples presented differs from the other two. A forced choice is required. Statistical analysis is used to determine whether a significant difference between treatments exists. The probability of choosing the different or odd sample by chance alone is one-third.

4. *Ranking* (Test 4). This test is used to make simultaneous comparisons of several samples on the basis of a single characteristic. A control need not be identified; all test samples are coded. Samples (which may include a control or standard) are presented simultaneously and ranked according to intensity of the characteristic designated; no ties are allowed. Rank totals or average ranks are obtained for each sample; differences are interpreted through statistical analysis of the data.

5. *Rating Difference/Scalar Difference from Control* (Test 5). This test may be used when a control sample is available for comparison with one or more experimental samples. Judges receive all samples simultaneously, an identified control and coded experimental treatments. The control may be introduced as an unknown sample. Category scales ranging from "No difference from control" to "Very large difference from control" are typical. Statistical analysis of the data is used to show whether the degree of difference from the control is significant.

• Sensitivity Tests. There are two types of sensitivity tests:

1. *Threshold* (Test 6). Thresholds are usually

—*Text continued on page 55*

expressed as absolute, which is the minimum detectable level of concentration of a substance. (Thresholds should be distinguished from the "just noticeable difference," which is the least amount of change in concentration of a given stimulus necessary to produce a noticeable change in perception of that stimulus.) Criteria of response in determining these thresholds include *detection threshold* (awareness of change from some neutral background) and *recognition threshold* (point at which the stimulus becomes identifiable). Samples for threshold testing are generally prepared as a log series of concentrations of a chemically pure stimulus material in a neutral substrate.

Difference tests (normally paired-comparison, duo-trio, and triangle) are appropriate for determining thresholds, provided fatigue, carry-over, and/or other constraints do not eliminate a particular method. In a paired-comparison test, the point in the series at which 75% of a judge's responses are correct is usually designated as the absolute threshold.

An ascending and descending series of concentrations may also be used for determining the absolute threshold. Samples are presented one at a time in order of physical concentration, and the judge must indicate whether the designated stimulus is detected. Sample presentation within the series is continued until the same judgment occurs for two successive presentations. Ascending and descending series are given alternately; the threshold represents the average of values obtained.

Regardless of the sensory method employed, threshold should be expressed only by procedures which are statistically sound.

2. *Dilution* (Test 7): The dilution technique determines the smallest amount of test material that can be detected when it is mixed with a standard material. The technique may provide information on relative intensities of treatments at comparable dilution levels. Diluting food samples may reveal components that are masked in the composite. Dilution testing is limited to food products that can be made homogeneous without affecting the factor being tested. Various difference test methods such as paired-comparison, triangle, and ranking may be used to measure the ability of the panelist to detect changes in the concentration of the test food product. Data obtained may be statistically analyzed and graphically illustrated.

ANALYTICAL-DESCRIPTIVE TESTS

Descriptive tests attempt to identify sensory characteristics and quantify them. Panelists are selected on their ability to perceive differences between test products and verbalize perceptions. Special training is required to perform profile and quantitative descriptive analyses.

There are two general types of descriptive tests—attribute rating and descriptive analysis:

• Attribute Rating (Test 8). There are, in turn, two types of rating tests:

1. *Category Scaling* (structured and unstructured). Coded samples are presented simultaneously or sequentially in a balanced order which differs among the individual panel members. A single-sample product evaluation is seldom employed; most perceptual judgments are relative. Category scales consisting of a series of word phrases (adverbial or adjectival modifiers) structured in ascending or descending order of intensity are used to measure the specified attribute (e.g., sweetness, off-flavor, etc.) An alternate scaling procedure is an unstructured vertical or horizontal line with verbal anchors at each end to describe or limit the attribute. For analysis purposes, successive digits are later assigned to each point represented on the scale, usually beginning at the end representing zero intensity. This follows the convention of having higher numbers represent a greater magnitude or more of a given quality. A statistical analysis (e.g., analysis of variance) of the mean intensity scores for each sample is used to determine significant differences among the mean scores for the samples represented.

2. *Ratio Scaling (Magnitude Estimation)*. This test is used to estimate the relationship between physical intensity and sensory magnitude. It can also be used for comparative ratings on specific attributes among two or more products. This method permits the participant to use a wide range of numbers of his or her own choice with the property that ratios or proportions among the numerical assignments reflect ratios of sensory intensities. Participants must be given a brief orientation in the method of ratio scaling. Samples are presented successively, in a balanced order. If a reference is used, it must be presented first, and it may be reintroduced later if desired. The number

Table 3— RECOMMENDED SENSORY TEST METHODS *for specific types of applications*[a]

Type of application	Appropriate test methods listed in Table 2
New product development	1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 13
Product matching	1, 2, 3, 5, 8, 9, 10, 11, 12, 13
Product improvement	1, 2, 3, 5, 8, 9, 10, 11, 12, 13
Process change	1, 2, 3, 5, 8, 9, 10, 11, 12, 13
Cost reduction and/or selection of a new source of supply	1, 2, 3, 5, 8, 9, 10, 11, 12, 13
Quality control	1, 2, 3, 5, 8, 9, 10, 11
Storage stability	1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 13
Product grading or rating	8
Consumer acceptance and/or opinions	12, 13
Consumer preference	1, 4, 12
Panelist selection and training	1, 2, 3, 4, 5, 6, 7, 8
Correlation of sensory with chemical and physical measurements	5, 8, 9, 10, 11

[a]Suggested general references: Abbott (1973), Amerine et al. (1965), ASTM (1978); ASTM Committee E-18 (1968; 1979; 1981); Ellis (1967); Hesh (1974; 1975a; b), IFT (1964; 1979, 1980), Larmond (1977), Stahl and Einstein (1973)

numerical scoring in subjective evaluation of foods. Food Res. 21: 273.
Caul, J.F. 1957. The profile method of flavor analysis. Adv. in Food Res. 7: 1.
Civille, G.V. and Szczesniak, A.S. 1973. Guidelines for training a texture profile panel. J. Text. Studies 4(2): 204.
Cochran, W.G. and Cox, G.M. 1957. "Experimental Design," 2nd ed. John Wiley and Sons, New York.
Dawson, E.H., Brogdon, J.L., and McManus, S. 1963. Sensory testing of differences in taste. I. Methods. II. Selection of panel members. Food Technol. 17(9). 45; 17(10): 39.
Drake, B. and Johansson, B. 1969. "Sensory Evaluation of Food." Annotated bibliography. Vol. 1, physiology, psychology; Vol. 2, methods applications, index. SIK-Rapport Nr. 255. Svenska Institutet för Konserveringsforskning, Göteborg, Sweden
Drake, B. and Johansson, B. 1974. "Sensory Evaluation of Food." Annotated bibliography, supplement 1968-1973. Vol. 1, physiology, psychology; Vol. 2, methods, applications, index SIK-Rapport Nr. 350. Svenska Institutet för Konserveringsforskning, Goteborg, Sweden.
Duncan, D.B. 1955. Multiple range and multiple F tests. Biometrics 11(1): 1.
Ellis, B.H. 1966. "Guide Book for Sensory Testing." Continental Can Co., Inc., Chicago, Ill.
Ellis, B.H. 1967. Efficient use of sensory evaluation methods. Food Prod. Dev. 1(5): 18.
Ellis, B.H. 1970. Sensory methodology for product development. Food Prod. Dev. 4(5). 80; 4(6): 16.
Green, D. and Swets, J. 1971. In "Signal Detection Theory in Psychophysics," p. 79. Robert E. Grieger Pub. Co., Inc., Huntington, N.Y.
Gregson, R.A.M. 1962. A rating-scale method for determining absolute taste thresholds. J. Food Sci. 27(4): 376.
Gridgeman, N.T. 1955. Taste comparisons: Two samples or three? Food Technol. 9: 148.
Gridgeman, N.T. 1961. A comparison of some taste-test methods. J. Food Sci. 26(2): 171.
Guilford, J.P. 1954. "Psychometric Methods." McGraw-Hill, New York.
H[illegible]rison, D.L. 1972. Planning and conducting experiments. In "Food Theory and Applications," p. 701. John Wiley and Sons, New York.
Hirsh, N. 1974. Getting fullest value from sensory testing. Part I: Use and misuse of testing methods. Food Prod. Dev. 8(10): 33.
Hirsh, N. 1975a. Getting fullest value from sensory testing. Part II: Considering the test objectives. Food Prod. Dev. 9(1): 10.
Hirsh, N. 1975b. Getting fullest value from sensory testing. Part III: Use and misuse of test panels. Food Prod. Dev. 9(2): 78.
Hopkins, J.W. 1950. A procedure for quantifying subjective appraisals of odor, flavor, and texture of foodstuffs. Biometrics 6(1). 1.
IFT. 1964. Sensory testing guide for panel evaluation of foods and beverages. Committee on Sensory Evaluation, Inst. of Food Technologists. Food Technol. 15(8): 25.
IFT. 1971. Author guidelines for IFT research papers reporting sensory evaluation data. Committee on Sensory Evaluation, Inst. of Food Technologists. Food Technol. 25(3): 53.
IFT. 1975. Minutes of Sensory Evaluation Div. business meeting at 35th Ann. Meet., Inst. of Food Technologists, Chicago, June 10.
IFT. 1979. "The Role of Sensory Evaluation in Product Quality Assurance." Inst. of Food Technologists' Sensory Evaluation Div. symposium. Food Technol. 33(9): 37.
IFT. 1980. "The Wide Scope of Sensory Evaluation." Inst. of Food Technologists' Sensory Evaluation Div. symposium. Food Technol. 34(11): 56.
IFT. 1981. Guidelines for the preparation and review of papers reporting sensory evaluation data. Sensory Evaluation Div., Inst. of Food Technologists. Food Technol. 35(4): 16.
Kahan, G., Cooper, D., Papavasiliou, A., and Kramer, A. 1973. Expanded tables for determining significance of differences for ranked data. Food Technol. 27(5). 61.
Kendall, M.G. and Stuart, A. 1968. "Advanced Theory of Statistics," Vol. 3, 2nd ed. John Wiley & Sons, New York.
Kramer, A. 1960. A rapid method for determining significance of differences from rank sums. Food Technol. 14: 576.
Kramer, A. 1963. Revised tables for determining significance of differences. Food Technol. 17(12): 124.
Kramer, A. and Twigg, B.A. 1962. "Fundamentals of Quality Control for the Food Industry." Avi Pub. Co., Westport, Conn.
Larmond, E. 1977. "Laboratory Methods for Sensory Evaluation of Food," Pub. 1637, rev. ed. Food Res. Inst., Canada Dept. of Agriculture, Ottawa, Ont.
Little, A.D., Inc. 1958. "Flavor Research and Food Acceptance." Reinhold Pub. Corp., New York.
Lockhart, E.E. 1951. Binomial systems and organoleptic analysis. Food Technol. 5(10). 428.
Mahoney, C.H., Stier, H.L., and Crosby, E.A. 1957. Evaluating flavor differences in canned food. I. Genesis of a simplified procedure for making flavor difference tests. II. Fundamentals of the simplified procedure. Food Technol. 11(9): Insert 29, 37.
Martin, S.L. 1973. Selection and training of sensory judges. Food Technol. 28(11): 16.
Morrison, D.F. 1976. "Multivariate Statistical Methods." McGraw-Hill, New York.
Moskowitz, H.R. 1974. Sensory evaluation by magnitude estimation. Food Technol. 28(11): 16.
Moskowitz, H.R. 1975. Applications of sensory measurement to food evaluations. II. Methods of ratio scaling. Lebensmittel-Wissenschaft + Technologie 8(6): 249.
Moskowitz, H.R. 1978. Magnitude Estimation: Notes on how, what, where and why to use it. J. Food Qual. 1: 195.
Moskowitz, H.R. and Barbe, C. 1976. Psychometric analysis of food aromas by profiling and multidimensional scaling. J. Food Sci. 41(3): 567.
Moskowitz, H.R. and Sidel, J.L. 1971. Magnitude and hedonic scales for food acceptability. J. Food Sci. 36: 677.
Moskowitz, H.R. and Wehrly, T. 1972. Economic applications of sweetness scales. J. Food Sci. 37(3): 411.
Peryam, D.R. 1958. Sensory difference tests. In "Flavor Research and Food Acceptance," A.D. Little, Inc., p. 47. Reinhold Pub. Co., New York.
Peryam, D.R. and Pilgrim, F.J. 1957. Hedonic scale method of measuring food preferences. Food Technol. 11(9): Insert 9.
Pilgrim, F.J., Schutz, H.G., and Peryam, D.R. 1955. Influence of monosodium glutamate on taste perception. Food Res. 20: 310.
Prell, P.A. 1976. Preparation of reports and manuscripts which include sensory evaluation data. Food Technol. 30(11). 40.
Roessler, E., Pangborn, R., Sidel, J., and Stone, H. 1978. Expanded statistical tables for estimating significance in paired-preference, paired difference, duo-trio, and triangle tests. J. Food Sci. 43(3): 940.
Scheffe, H. 1952. An analysis of variance for paired comparisons. J. Am. Statistical Assn. 47(259): 381.
Schutz, H.G. 1965. A food action rating scale for measuring food acceptance. J. Food Sci. 30(2): 365.
Schutz, H.G. 1971. Sources of invalidity in the sensory evaluation of foods. Food Technol. 25(3). 53.
Sidel, J.L. and Stone, H. 1976. Experimental design and analysis of sensory tests. Food Technol. 30(11): 32.
Sjöström, L.B. and Cairncross, S.E. 1954. The descriptive analysis of flavor. In "Food Acceptance Testing Methodology," symposium sponsored by Quartermaster Food and Container Inst. for the Armed Forces, Chicago, Ill.
Sjoström, L.B., Cairncross, S.E., and Caul, J.F. 1957. Methodology of the flavor profile. Food Technol. 11(9): Insert 20.
Snedecor, G.W. and Cochran, W.G. 1967. "Statistical Methods," 6th ed. Iowa State Univ. Press, Ames.
Stahl, W.H. and Einstein, M.A. 1973. Sensory testing methods. In "Encyclopedia of Industrial Chemical Analysis," Vol. 17, p. 608. John Wiley and Sons, New York.
Stevens, S.S. 1962. The surprising simplicity of sensory metrics. Am. Psychologist. 17: 29.
Stone, H., Sidel, J., Oliver, S., Woolsey, A., and Singleton, R.C. 1974. Sensory evaluation by quantitative descriptive analysis. Food Technol. 28(11): 24.
Szczesniak, A.S., Brandt, M.A., and Friedman, H.H. 1963. Development of standard rating scales for mechanical parameters of texture and correlation between the objective and the sensory methods of texture evaluation. J. Food Sci. 28(4): 397.
Tilgner, D.J. 1962a. Dilution tests for odor and flavor analysis. Food Technol. 16(2): 26.
Tilgner, D.J. 1962b. Anchored sensory evaluation tests—A status report. Food Technol. 16(1): 47.
Tompkins, M.D. and Pratt, G.B. 1959. Comparison of flavor evaluation methods for frozen citrus concentrates. Food Technol. 13: 149.
Tukey, J.W. 1949. Comparing individual means in the analysis of variance. Biometrics 5(2): 99.
Wald, A. 1947. "Sequential Analysis," John Wiley & Sons, New York.
Wiley, R.C., Briant, A.M., Fagerson, I.S., Murphy, E.F., and Sabry, J.H. 1957. The Northeast regional approach to collaborative panel testing. Food Technol. 11(9): Insert 43.
Winer, B.J. 1971. "Statistical Principles in Experimental Design," 2nd ed. McGraw-Hill, New York.

the subject must indicate a preference for one sample over another. Results are obtained in terms of the relative frequencies of choice of the two samples represented, generally as accumulated for all participants.

• Ranking Test (Test 4). This is, in effect, an extension of the paired-preference approach. Three or more coded samples are presented simultaneously, sufficient in amount so that the subject can check back on his or her first impression. The total number of samples tested is dependent upon the subject's span of attention and memory, as well as physiological considerations. The subject is asked to assign an order to the samples according to his or her preference. As with the paired-comparison (paired-preference) method, rank order evaluates samples only in relation to one another. The amount of liking (or disliking) for individual samples cannot be adequately determined by this method.

• Rating Tests. Scale ratings reflect respondents' perceived intensity of a specified attribute under a given set of conditions. Various rating scales have been developed and used:

1. *Hedonic Rating Scale* (Test 12). This test is used to measure the level of liking for food products by a population. It may be applied in testing for preference or acceptance (i.e., preference is inferred from hedonic ratings). The method relies on test subjects' capacities to report, directly and reliably, their feelings of like and dislike.

Several variations of the traditional nine-point word hedonic scale have been used effectively. These include: (1) a reduced number of rating categories, although not fewer than five is recommended; (2) a greater number of "like" rating categories than "dislike"; (3) omission of the neutral rating category, (4) substitution of the verbal categories by caricatures representing degrees of pleasure and displeasure (facial hedonic scale); and (5) use of a nonstructured, nonnumerical line scale anchored with "like" and "dislike" on opposite ends.

Samples may be tested monadically (single sample), paired against a standard, or in combinations dictated by statistical design. The monadic test method is appropriate for determining the acceptability of a new or unusual food product where there are no similar products for comparison. In a paired test, the experimental product is directly compared to the standard. Multiple-sample comparisons may also be used for preference screenings. In paired or multiple-sample comparisons, samples are presented simultaneously or successively in a balanced order. The test subject is asked to evaluate each sample and mark the scales accordingly. Instructions must not influence the subject's response.

Hedonic scale ratings are converted to numerical scores, and statistical analysis is applied to determine difference in degree of liking between or among samples. A hedonic rating test can yield both absolute and relative information about the test samples. Absolute information is derived from the degree of liking (or disliking) indicated for each sample, and relative information is derived from the direction and degree of difference between or among the sample scores.

2. *Food Action Rating Scale* (Test 13). This test may be used to measure the level of acceptance of food products by a population. The scale is not applicable for rating specific characteristics; rather it is a measure of general attitude toward a food product. This rating scale includes action as well as affective-type statements. Nine successive rating categories ranging from "I would eat (buy, use, etc.) this every opportunity I had" to "I would eat this only if I were forced to" are represented. One or more samples may be tested. Samples are presented sequentially in a balanced order, and the test subject is told to decide which of the statements on the scale best represents his or her attitude. Subjects are allowed to make their own inferences about the meaning of the scale categories. The scale ratings are converted to numerical scores to facilitate statistical analysis of data.

References & Suggested Reading

REFERENCES

Abbott, J.A. 1973. Sensory assessment of textural attributes of foods. In "Texture Measurements of Foods," p. 17. D. Reidel Pub. Co., Washington, D.C.

Amerine, M.A., Pangborn, R.M., and Roessler, E.B. 1965. "Principles of Sensory Evaluation of Food." Academic Press, New York.

AMSA. 1978. "Guidelines for Cookery and Sensory Evaluation of Meat." Am. Meat Sci. Assn. and Natl. Live Stock and Meat Board, Chicago, Ill.

ASTM Committee E-18. 1968. "Manual on Sensory Testing Methods," STP 434. Am. Soc. for Testing and Materials, Philadelphia, Pa.

ASTM Committee E-18. 1972a. Standard recommended practice for determining effect of packaging on food and beverage products during storage, E460. In "1975 Annual Book of ASTM Standards," Part 46, p. 257. Am. Soc. for Testing and Materials, Philadelphia, Pa.

ASTM Committee E-18. 1972b. Standard recommended practice for sensory evaluation of industrial and institutional food purchases, E461. In "1975 Annual Book of ASTM Standards," Part 46, p. 260. Am. Soc. for Testing and Materials, Philadelphia, Pa.

ASTM Committee E-18. 1973. Standard recommended practice for establishing conditions for laboratory sensory evaluation of foods and beverages, E480. In "1975 Book of ASTM Standards," Part 46, p. 266. Am. Soc. for Testing and Materials, Philadelphia, Pa.

ASTM Committee E-18. 1978. Standard definitions of terms relating to sensory evaluation of materials and products, E253. Am. Soc. for Testing and Materials, Philadelphia, Pa.

ASTM Commitee E-18. 1979. Standard recommended procedure to establish guidelines for open dating food and beverage products, E687 79. Am. Soc. for Testing and Materials, Philadelphia, Pa.

ASTM Committee E-18. 1981. Guidelines for the selection and training of sensory panel members, STP 758. Am. Soc. for Testing and Materials, Philadelphia, Pa.

Bohren, B.B. and Jordan, R. 1953. A technique for detecting flavor changes in stored dried eggs. Food Res. 18: 583.

Brandt, M.A., Skinner, E.Z., and Coleman, J.A. 1963. Texture profile method. J. Food Sci. 28(4): 404.

Byer, A.J. and Abrams, D. 1953. A comparison of the triangular and two sample taste-test methods. Food Technol. 7: 185.

Carlin, A.F., Kempthorne, O., and Gordon, J. 1956. Some aspects of

RECONOCIMIENTO DE OLORES BASICOS

Panalista No.									% Correcto
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									

APÉNDICE 2 continuación

Boleta para la prueba de reconocimiento de olores básicos

Nombre: ______________________

Fecha: ________________________

Reconocimiento de olores básicos

Los frascos cubiertos contienen substancias olorosas que se encuentran comúnmente en el hogar o el lugar de trabajo. Acerque el frasco a su nariz, saque la tapa, husmee brevemente 3 veces y trate de identificar el olor. Si no se le viene a la memoria el nombre exacto de la substancia, trate de describir alguna cosa con la que usted asocie ese olor.

Código	Olor

Remembering Odors and Their Names

Trygg Engen

It is generally believed that odor memory is excellent, but its strengths and weaknesses compared with visual or auditory memory are not well appreciated. What odor perception does best, the basis for its reputation, is to recreate significant past episodes in a person's life. The French novelist Marcel Proust has provided the classic example, describing at the beginning of his masterpiece *Remembrance of Things Past* how the aroma and flavor of a morsel of petite madeleine soaked in a spoonful of tea brought back long-forgotten memories of his aunt's bedroom in the family's house at Combray: "When from a long-distant past nothing subsists, after the people are dead, after the things are broken and scattered, still alone, more fragile, but with more vitality, more unsubstantial, more persistent, more faithful, the smell and taste of things . . . bear unfaltering, in the tiny and almost impalpable drop of their essence, the vast structure of recollection."

It is not that odor memory provides storage for a large number of items. Rather, it involves significant episodes, such as that described by Proust, and odors of foods (including those to which one has an aversion because of their association with illness), persons, and places. These are not, of course, the odors one can study in the laboratory, where the excellence of long-term recognition memory of odors has been documented with typical stimuli (Engen and Ross 1973). Figure 1 compares these results with the Proustian phenomenon. The strength of memory varies with the special involvement a person has with the odor. Time's potential interference with memory is negligible in the case of either laboratory or episodic odors. In this regard visual and olfactory recognition memory are different. There certainly are such memorable events in visual memory, but long-term recognition memory is a special attribute of olfaction, which shows an almost flat forgetting curve in both the short and the long term (Engen et al. 1973).

We can recognize odors virtually forever, though we have difficulty conjuring up a particular smell and remembering its name

But there is an important distinction between memory by recall and by recognition. Vladimir Nabokov, another among the many writers who have provided illustrations of odor perception, calls attention in his early novel *Mary* to this distinction: "Memory can restore to life everything except smells, although nothing revives the past so completely as a smell that was once associated with it." At best, the ability to recall an odor is limited compared to one's ability to recall sounds and sights. There are two aspects to this. One is the inability to conjure up the odor sensation, to recall, for example, lemon, as one can recall its color or shape. The other, of primary concern here, is associative memory, the ability, for example, to use odors along with names or pictures for mnemonic purposes.

To begin with, one must make a distinction between being able to remember the name of an odor and being able to recognize the odor. Recognition memory refers to judgments regarding which one of a set of odors has been encountered before; it does not require special naming. Any means identification by a number or other symbol would be as valid as a name, the only requirement being that each odor have a unique label. In our early experiments, one odor was presented at random from a predetermined set at each trial to be identified without the help of comparison stimuli or a list of responses from which to choose (Miller 1956). The size of the set from which the test odor was selected was varied in order to determine the maximum number of stimuli a subject could handle without confusing their identities. For odor quality this number turned out to be about 16 in experiments with college students as well as with trained chemists and perfumers (Engen and Pfaffmann 1960; Jones 1968). Such a level of performance, known as channel capacity, is similar to that observed with Munsell chips varying in color, brightness, and size (Eriksen and Hake 1955). Contrary to earlier opinion, then, the human ability to identify a single odor without any other cue is not outstanding but is similar to that observed with comparable stimuli in vision and audition.

Other researchers have argued that subjects can do better if trained to name odors (Desor and Beauchamp 1974; Cain 1979). However, in evaluating the effect of training it must be borne in mind that in our experiments

Trygg Engen graduated from the Oslo Katedralskole in Norway in 1946 and received his Ph.D. from the University of Nebraska in 1954. Since then he has been engaged in teaching and research at Brown University and at the University of Stockholm. The research has been in psychological measurement, with a special interest in olfaction. Address: Walter S. Hunter Laboratory of Psychology, Brown University, Providence, RI 02912.

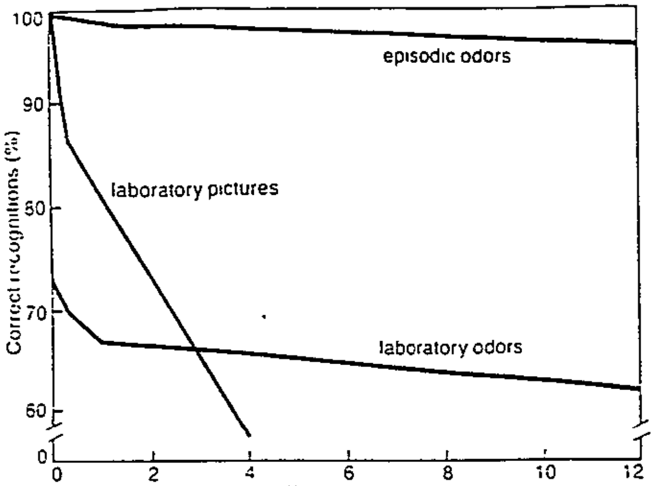


Figure 1. The special strength of odor memory—the relative permanence of the ability to recognize a given odor—is apparent when one compares it to recognition memory for pictures. The recognition of so-called episodic odors, odors associated with significant real-life experiences, remains close to the initial strength as time passes. By comparison, one's ability to recognize pictures shown in a laboratory experiment, while as strong initially as the ability to recognize episodic odors, decreases rapidly in a relatively short time. Laboratory odors are not recognized well after a minimal interval of time, but, like episodic odors, show very little long-term loss.

the stimuli were equally likely to be presented, whereas in the training experiments and in real life the chances vary over trials. Unequal probabilities mean that the number of alternative choices per odor in the set is reduced. More importantly, the experiments we conducted measured memory span and discriminability of stimuli, not familiarity, which of course will be different for college students and perfumers.

One avoids the naming problem by performing recognition experiments (Engen and Ross 1973), but naming remains an interesting phenomenon. Visual objects and their names are so closely linked that they are inseparable (Posner 1969) It seems unreasonable to assume, as apparently was done in the training experiments referred to above, that people learn names for visual objects but somehow fail to learn names for odors, leaving them incapable of showing the size of their odor memory Rather, it seems, as Nabokov suggests, that the linking of names and odors is inherently weak. This article reports recent research on the problem of naming odors and on the arrangement of odors in memory. It considers, not how many odors can be named, but how accurately an untrained person can do it and the nature of the words he or she uses for the task.

The associative power of odor

It has long been known by experimental psychologists that associating odors with items to be learned is no more beneficial than associating them with nonsense syllables (Heywood and Vortriede 1905). One recent experiment shows that people take longer to learn to associate odors with numerical digits than they do to associate abstract free-form line drawings with digits (Davis 1977). In another experiment, it was found that subjects were "sluggish" in learning new word associations to make odors more identifiable (Cain 1980). Therefore, it is not surprising that it should be difficult to recall the name of a familiar odor. Sumner (1962) experienced this problem earlier in attempting to develop a clinical test of olfaction. He presented 12 common odors—almond, lemon, camphor, coffee, clove, anise, chocolate, peppermint, nutmeg, tar, eucalyptus, and asafetida—to a control group of medical students to establish norms against which to evaluate patients. To his surprise, the typical medical student could identify only six such odors.

Considering this difficulty in naming, it is curious that odor classification has dominated research in olfaction. The working assumption is that stimuli cause sensations of certain qualities which subjects can describe and categorize. What is basically wrong with this approach is its emphasis on words, which have but a tenuous connection to odor perception. The reason for the continued popularity of odor classification is undoubtedly that classification has provided meaningful data on color perception correlated with both the stimulus and the relatively specific transduction characteristics of photoreceptors. There is a strong tendency to assume that all the senses fit such a stimulus-response model and to compare them with vision. Thus Henning's smell prism, reproduced in Figure 2, is still presented in textbooks as analogous to the color solid, although there is not a shred of evidence that the smell prism is valid as a model for understanding the stimuli and olfactory receptors or that the words "flowery," "putrid," "fruity," "resinous," "spicy," and "tarry" represent special odor qualities analogous to the colors blue, green, yellow, and red.

The fact that one may recognize an odor as familiar without being able to name it has been described as the

"The smell and taste of things . . . bear unfaltering, in the tiny and almost impalpable drop of their essence, the vast structure of recollection"

—Marcel Proust

"tip-of-the-nose" state (Lawless and Engen 1977) If the experimenter presents the correct odor name to a subject in that state, the subject will recognize it immediately and be puzzled about why he could not retrieve it in the first place. The reason is that the odor-name association is not symmetrical. We asked some 35 subjects (college students) two sets of questions. One pertained to the quality of an odor: for example, if the subject could name a similar odor, its general category, and objects which might produce it. The other set probed for information about the odor name: whether the subject could give a name with a similar meaning, sound, or number of syllables and letters. These questions were similar to those used in a study of the tip-of-the-tongue state (Brown and McNeill 1966). The tip-of-the-nose results were unusually clear-cut: the subjects' answers to the second set of questions indicated that they could not think of a similar word related to the target. They could, however, answer the first set of questions, showing that they knew something in general about the odor category or an object associated with it.

The tip-of-the-nose state differs in important ways from the tip-of-the-tongue state, in which a person can tell you something about the correct name, such as the initial letter, the number of syllables, or the general configuration of the word. In the tip-of-the-nose state, the person appears to have no such lexical information about odor names, a situation which undoubtedly influences the results in tests of odor perception.

Our recent experiments, to be described here, examined this problem closely. They included a task in which there were no prescribed names (an open-ended situation leaving it up to the subjects to recall a name for each odor) and a multiple-choice task in which alternative

'Memory can restore to life everything except smells, although nothing revives the past so completely as a smell that was once associated with it"

—Vladimir Nabokov

names were provided and which demanded simple recognition and no production of names at all.

The first task was simply to name odors spontaneously; it was left up to the subject to produce a name for each odor. The subject was presented common odors, one at a time, and asked to describe each with one word, if possible. It was emphasized that the odors would not be harmful or unpleasant. There was ample time between trials to prevent fatigue or adaptation

Two sets of odorants were used with two different groups The first set, presented to 48 subjects, 25 women and 23 men, included stimuli used by Cain and Krause (1979) called "brand names," which are considered relatively easy to identify. Bazooka bubble gum, Crayola crayons, Ivory bar soap, Johnson's baby powder, and Vick's Vaporub Five other odorants used were amyl acetate (banana oil), clove oil, citral (lemon), phenylethyl alcohol (rose), and methyl salicylate (mint or wintergreen) Each of the odorants was presented in cotton inside a widemouthed bottle 50 mm high and 25 mm in diameter, shavings were used for the solid substances (gum, crayon, and soap) There were no visual cues available to the subject.

The second set of odorants included some of the same odors (banana, lemon, and rose); they were microencapsulated stimuli produced by 3M, which are now widely used clinically (Doty et al. 1984), and which are used on the scratch-and-sniff card that accompanies this article The 13 odors used, in addition to the three [illegible]tioned, were grape, smoke, mint, gasoline, licorice, [illegible]sk, pine, onion, soap, and diesel. Each was presented on a strip of paper attached to a 100 × 150 mm card. The subject was asked to scratch the paper strips with a fingernail or a piece of sandpaper. This released the odorous molecules for the subject to sniff and name as succinctly as possible. Thirty new subjects, 17 men and 13 women, were tested in this task.

For the first set of 10 odorants including the brand names, the number of correct responses ranged from 1 to 9, with a mean of 4.4, or 44%, and a standard deviation of 1.9. The mean number of correct responses for the 5 brand names was 2.4, with a standard deviation of 1 3. The mean for the 5 other odorants was 2.1, with a standard deviation of 1.2. The difference between the two kinds of odorants is not significant, although it is in the expected direction. The difference between men and women is statistically significant, with means and standard deviations of 4.1 and 1.6 and 4 9 and 2.1 for men and women respectively. Regarding differences among the odors, the percentage of correct identification ranged from 27% for rose to 75% for Bazooka bubble gum.

For the scratch-and-sniff experiment with 13 odorants, the mean number of correct responses was 4.2, or 32%, ranging from 1 to 8. (The difference between the two sets in average percentages of correct identification—44% and 32%—is significant at the 1% level) Women again did significantly better than men, with a mean score of 4.9 compared to the men's 3 8. Regarding differences among odors, the results ranged from musk, which was never correctly identified, to licorice, which was identified correctly by 83% of the subjects. To summarize, although there are individual differences, people are not good at naming even very familiar odors The main point is that, on the average, fewer than half of the odors were correctly named, even with a liberal scoring system, accepting, for example, "gum" as a correct answer without either [illegible]ubble" or "Bazooka "

A number of factors affect this performance. First, some stimuli are better than others. The scratch-and-sniff odorants were weaker and apparently less saturated than those in the other set, which represented "the real thing" in the case of brand names. Second, in each set, as we have seen, some odors are more often correctly identified than others Part of the difference involves the scoring of the responses For example, what is a correct

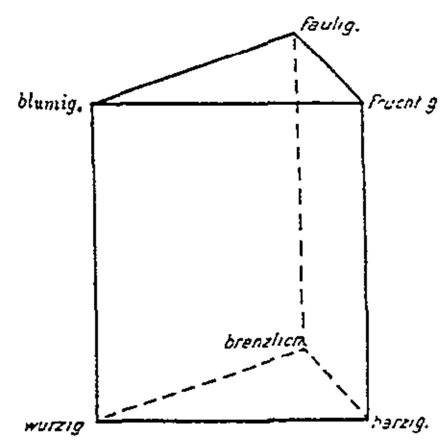


Figure 2. Despite the differences between visual and odor memory demonstrated in Figure 1, attempts to classify odors by analogy with classifications of visual stimuli have dominated research in olfaction. The smell prism shown here, developed by the German psychologist Hans Henning more than 70 years ago, can still be found in textbooks, even though its dimensions defined by the classifications flowery (*blumig*), putrid (*faulig*), fruity (*fruchtig*), spicy (*wurzig*), tarry (*brenzlich*), and resinous (*harzig*) do not provide a valid model for olfactory stimuli.

Table 1 Examples of odor descriptions

Johnson's baby powder (48 responses)

correct response (21)	Band-Aid	rose
no response (6)	bubble gum	suntan lotion
powder (3)	dentist's office	tissue
soap (3)	flower	toilet paper
air freshener	hand lotion	vanilla
baby oil	man's perfume	wax
baby wipes		

3M lemon Microfragrance (30 responses)

correct response (12)	cleaner, Lemon-fresh Pledge
don't know (3)	hard candy
air freshener	know, but can't recall
bathroom freshener	like pine
berry, as in magic marker	magic markers special
candy	nothing
citron citrus	orange
citrus fruit	some kind of fruit
citrus not a lemon or lime	

Table 2 Types of descriptions of odors (%)

	Correct response	Related odor	Associated object	Related sensation	No response
Natural odors	44	3	47	1	5
Micro-fragrances	33	5	45	7	10

Table 3 Semantic levels of reference to odors

Level	Example of verbal response
target	lemon
synonym or near synonym	citron
superordinate	fruit
general noun or word	smell

After Halliday and Hasan 1976

answer for musk? But beyond that, the difference must involve experience and psychophysical attributes.

Explaining the differing levels of performance of men and women is not easy. The present data support those reported by Cain (1980, 1982) and by Doty and his colleagues (1985). The reason for this reliable finding is not clear, but two hypotheses stand out. One is that it reflects hormonal differences, which could affect the sensory receptors and other levels of the perceptual system and make women more sensitive to odors than men are. There is good evidence for such a hormonal mechanism, with greater sensitivity during ovulation than during menstruation (Mair et al. 1978). However, it seems unlikely that the mechanism could have had an effect on the average results in the present tasks. The other hypothesis is that the difference reflects a greater verbal ability in women, which is well known from the literature on gender differences. An analysis of the effect of such differences by Hyde (1981) suggests that, although the difference in abilities is reliable and replicable, it may nonetheless be too small to influence performance in the present tasks. Still, the control of words over perception is very significant, as we will see, and the absence of given words must be considered the major factor in explaining the poor performance in this task by both men and women.

How odors are named

In addition to scoring the responses for accuracy of identification, we analyzed the kinds of words or descriptions used by subjects misidentifying odors. When a subject fails to describe the odor of lemon as lemon, what is the nature of the verbal response? It was expected that such an analysis would provide information about odor associations and thus the nature of odor memory. Table 1 presents representative data for two of the odors, Johnson's baby powder from one set and lemon from the other. Both are close to the average in terms of accuracy.

Table 2 summarizes the data obtained from all the subjects in both groups. Their responses fall into five broad categories: the correct response (e.g., lemon), a related odor word (fruity), an object associated with the odor (Pledge), a related sensation (bitter), and no response or "I don't know." The lexical nature of these responses was analyzed using the system described by Halliday and Hasan (1976) and illustrated in Table 3 with responses to the odor of lemon. It is a hierarchical taxonomy in which each level includes the meaning of the level above.

Applying this system to the present data shows that the target or the correct name is virtually the only obtained response like the categories shown in Table 3—that is, responses to odor are either the target or an object associated with the odor, or they are wrong. They do not conform to Miller's (1969) hypothesis, based on other data, that such lexical responses are generally more likely to be superordinates than subordinates of the name. Superordinate responses, which are described as "related odors" in Table 2, were obtained only about 3 to 5% of the time in the present study. They include identifying lemon as "citrus fruit" or smoke as "something burnt." Examples of the other semantic levels shown in Table 3 can also be found, but they too are uncommon. However, there are other responses which may be categorized semantically in terms of collocation, a different lexical system described by Halliday and Hasan. An example would be responding to the odor of onion as garlic, pizza, or spicy. Such typical responses suggest how odors are organized in memory. Moreover, collocation explains the kind of odor episodes described at the beginning of this article.

The results summarized in Table 2 indicate that people do categorize odors, but not with semantically cohesive general nouns. Rather, they do it in terms of the similarity of odors and the similarity of the context or kind of object in which odors may be perceived. Verbal responses to odors tend to be personal, referring to objects with which a person has had experience. Some objects are more highly specific (Wrigley's Juicy Fruit gum) than others (garlic powder, pizza). There are occasional references to taste qualities, such as sweetness.

Contrasts between odors and colors

The odor descriptions we obtained are not the kind one is led to expect by traditional odor classification systems, which would require responses analogous to so-called landmark colors such as red, green, blue, white, and black. Instead, odors are evidently named in a way that is similar to the naming of so-called accidental colors, or colors of familiar objects (Miller and Johnson-Laird 1976), showing the effect of environmental experience and learning. Odors may also be like the colors at the extremes of the visual spectrum, which are not as readily named as landmark colors (Chafe 1970).

On the stimulus side, there is no known physical or chemical correlate of odor similar to the well-established psychophysical correlates of the visual spectrum, and no evidence has been found yet to confirm the existence of olfactory receptor neurons analogous to the relatively specific cones (Gesteland 1986). Rather, odors may be coded individually according to patterns of activity in olfactory and other receptors. The olfactory system contains a rich supply of receptors capable of responding to highly diverse signals, which may function in assemblies together with different receptor neurons.

Perceptions of odors and colors operate according to different principles. To begin with, colors involve a lexical system which is organized in memory in a relatively abstract and rigidly controlled way. By contrast, odor perception is characterized by flexibility and adaptability, and a relatively concrete but open-ended nonverbal coding system. This manner of organization in turn affects odor memory. What is stored about odors is not likely to involve semantic categories which can be used to retrieve odors. Rather, odor memory involves perceptually unitary episodes—one is reminded again of those portrayed by Marcel Proust—described in an idiosyncratic lexicon (Engen 1982). Because the odor is an integral part of the episode, what is stored in memory is not interfered with or even forgotten through frequent experience with similar names or closely related sensory information, such as seeing petites madeleines in the baker's window every day. This is also the reason that it is difficult to use other people's descriptions in naming odors (Engen and Ross 1973), to remember such words when they are the correct answers on a test of odor perception (Sumner 1962), or to learn new names for odors (Davis 1977, Cain 1980).

Finally, there is a difference between odor and color in the direction of association between a name and a stimulus. Once one has presented a color chip and had a subject name it, the subject can conjure up a mental image of the color or the chip when the name is subsequently presented (Dorcus 1932). But in odor perception there is no such reciprocity: while perception of an odor may elicit an associated name, such a name cannot be used later to bring back the original odor perception. Odor memory is largely limited to recognition, as when a stimulus, a petite madeleine soaked in tea, produces the odor and the past episode in which it was first experienced. One does not retrieve an odor episode with words, but with odor. The strength of the association is weak from odor to name and nearly zero from name to odor. Cain (1980) points out that manufacturers of synthetic flavors benefit from this, because one is likely to perceive the odor of a product as matching its label, for example, strawberry, even though it is in fact a poor approximation of the real thing. Being provided with the name also resolves any tip-of-the-nose state. Having names available before an odor is smelled changes the perceptual problem and the naming process.

The effect of names

To elucidate further the effect of names, we designed a multiple-choice test using the same 13 scratch-and-sniff odorants given above. Four alternative names were provided for each odor that would be read before the subject scratched and sniffed. One was the correct name, and the other three were foils. It was the nature of these incorrect alternative responses that was the focus of our experiment. There were two different sets, each used for a different group of subjects. One set, selected from a standard odor test (Doty et al. 1984), represented quite different things from the target odor, while the other set referred to similar objects. For example, for the odor of grape the set with the diverse choices provided the incorrect alternative answers pizza, turpentine, and clove, while the alternatives for the categorized set were melon, plum, and strawberry. All the correct names and alternative responses are presented in Table 4. It is important to note that the odors represented by the categorized set are highly discriminable and not at all likely to be confused in direct comparisons.

A pool of 67 subjects, different from those who participated in the experiments already described, was

Table 4 Alternatives in a multiple-choice test (correct name in italics)

Diverse alternatives

a pizza	b turpentine	c. *grape*	c clove
a licorice	b clove	c chili	c *banana*
a *root beer*	b smoke	c banana	c watermelon
a tomato	b licorice	c strawberry	d *menthol*
a *gasoline*	b pizza	c peanut	c lilac
a cedar	b gasoline	c *lemon*	c root beer
a soap	b *licorice*	c black pepper	c peanut
a *musk*	b garlic	c turpentine	c lime
a lime	b *rose*	c mint	c bubble gum
a cola	b cinnamon	c *pine*	c coconut
a smoke	b whiskey	c pineapple	c *onion*
a *soap*	b fruit punch	c menthol	c pumpkin pie
a pumpkin pie	b *diesel*	c rose	c lemon

Categorized alternatives

a melon	b plum	c *grape*	d. strawberry
a pear	b apple	c pineapple	c *banana*
a charcoal	b *smoke*	c gas	c kerosene
a peppermint	b clove	c pine	d *menthol*
a *gasoline*	b. motor oil	c turpentine	d paint thinner
a grapefruit	b orange	c. *lemon*	c lime
a chocolate	b *licorice*	c peppermint	d caramel
a. *musk*	b bath oil	c hand lotion	d lipstick
a. lilac	b. *rose*	c. lily	d carnation
a cedar	b. furniture polish	c. *pine*	d turpentine
a garlic	b green pepper	c leek	d *onion*
a *soap*	b deodorant	c hand lotion	d toothpaste
a motor oil	b *diesel*	c gas	d kerosene

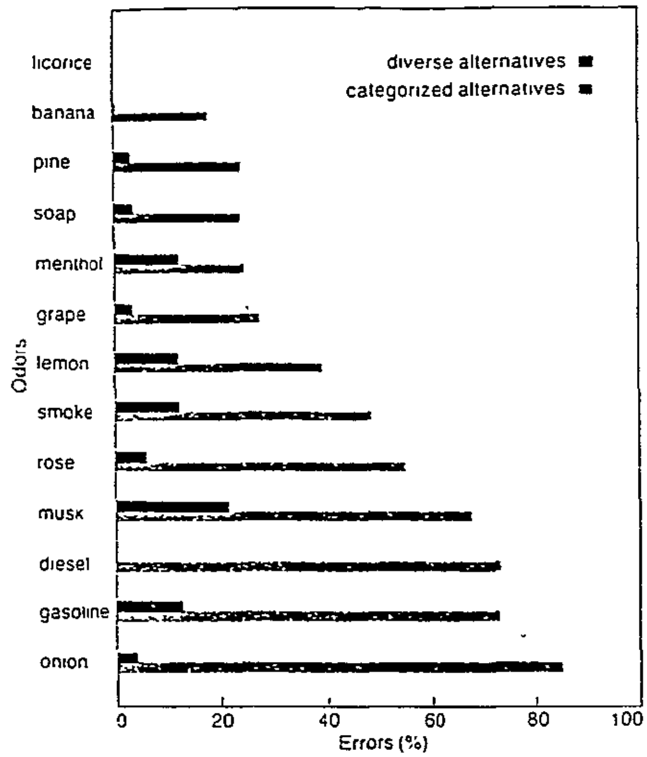


Figure 3. Because the association between odors and names is inherently weak and pliable, recognition of odors can easily be influenced when various verbal alternatives are presented to accompany the odors. This graph shows the mean percentages of two groups of subjects making errors in a multiple-choice test requiring them to select one of four names for each of 13 odors (see Table 4). The odorants were microencapsulated substances. The black bars show the percentages of subjects making an error when the choices represented diverse substances, for example, pizza, turpentine, grape, and clove, when grape is the correct answer. The gray bars show comparable results when the choices represented a category of similar substances, for example melon, plum, grape, and strawberry, when grape is the correct response. The poorer accuracy in the latter case appears to be proportional to the similarity of the odors represented by the incorrect alternatives and the target. The subjects managed as a group to identify without error only licorice (with both diverse and categorized alternatives) and diesel (with diverse alternatives).

divided at random into two groups. One group of 34 received an answer sheet with the correct names and the diverse set of alternatives, and another group of 33 received an answer sheet that contained the correct names and the categorized alternatives. The order of presentation of the 13 odors and the letter of the correct answer were the same for both groups. The group being tested with the diverse alternatives obtained a mean score of 93.3% correct answers, or 12.2 of the 13 odors, with a standard deviation of 1.1. By contrast, the group with the categorized alternatives scored only slightly over 50% correct with a mean of 6.6 for the 13 odors and a standard deviation of 2.2. This is a highly significant difference.

Clearly, the kind of names available presents an important problem (Rabin and Cain 1984). It is possible that a person may be tricked and describe onion as garlic because he has a verbal set of "garlic" provided by the foil without actually perceiving the odor as garlic. Resolution of this complex problem requires more than the present data, but one is inclined to the hypothesis that the results reflect perceptual effects rather than response bias.

The performance varied for the different odors, as can be seen in Figure 3, which shows the percentage of subjects who misidentified each odor in the two sets. For all except licorice, the percentages of error are higher for the group presented with the categorized alternatives. Some choices for the categorized set altered the performance more than did others. For example, in the case of onion the foil "garlic" was chosen by most of the subjects. Although familiarity and other factors are involved, the nature of the foils is the main explanation of the difference in results between the two sets. Onion and garlic are easily discriminated when compared side by side, but the problem here is that they also are similar. For the other odors the errors are distributed more evenly across the alternatives. In the case of diesel odor, for example, the percentages of errors for the categorized alternatives "motor oil," "gas," and "kerosene" are nearly equal. In general, the difference in the performance under the two conditions shown by the bars in Figure 3 reflects how well the misnomers were selected in terms of collocation. The categorized alternatives for banana, pine, and soap are not as adequate in this regard as those for diesel, gasoline, and onion.

The function of smell

The research results that I have surveyed lead us to conclude that the association between odors and their verbal descriptions is generally weak. An average person can identify only a small number of odors by verbal label, about half a dozen common ones, which is a level of ability far below that attained by the same person discriminating between odors presented side by side.

Psychophysical attributes of the stimuli and individual differences affect the results, but not substantially. However, one factor which stands out is the dominance of available names. Once a subject has an expectation of what the name of an odor might be, that name will largely control the sensation, which will be perceived as fitting the name. By contrast, the presentation of an odor before a name is available does not have a similar effect on the choice of names. In fact, the unavailability of a verbal response often leaves one in the tip-of-the-nose state, unable to recall a name even though the odor is familiar. That is, the associative strength of an odor-name pair is weak and asymmetric and is easily influenced by the verbal factor.

Analysis of the kinds of verbal responses obtained in free recall tests reveals that the responses are quite different from those used in describing and classifying colors. In general, odor perception is not organized lexically by nouns but around the similarity of objects causing odors and especially the contexts in which odors usually occur. A corollary is that they tend to be described in terms of personal references rather than by more general names like those proposed by traditional classification systems. Such references are similar to the words of a young child who might describe objects functionally—for example, an orange as "something to eat"—when an adult would use a more abstract term such as "citrus fruit."

The salient aspect of the sense of smell is the persistence of memories of episodes associated with odor. Such memories are retrieved with odor stimuli, as when one recognizes a familiar odor, but not with the odor name alone in the absence of stimulation. There is no pure recall of odor perceptions using a name as a paired associate. The reason is the inherently weak association between the odor and the name, which in turn reflects the fact that the odor name is not an important part of the episode. The main function of the sense of smell, then, is not to recall odors for cognitive reasons but to respond to odors actually encountered.

References

Brown, R W, and D McNeill 1966 The "tip of the tongue" phenomenon *J Verbal Learning Verbal Beh.* 5:325–37.

Cain, W. S 1979 To know with the nose. *Science* 203:467–70.

——— 1980 Chemosensation and cognition In *Olfaction and Taste*, ed H. van der Starre, vol 7, pp. 347–58 London IRL Press

——— 1982. Odor identification by males and females: Predictions vs. performance. *Chem Senses* 7:129–42.

Cain, W S, and R J Krause 1979 Olfactory testing: Rules for odor identification *Neurological Res* 1.1–9

Chafe, W. L 1970 *Meaning and the Structure of Language* Univ. Chicago Press

Davis, R G 1977 Acquisition and retention of verbal associations to olfactory and abstract visual stimuli of varying similarity. *J Exp Psychol: Human Learning and Memory* 3 37–51.

Desor, J A, and G K Beauchamp. 1974 The human capacity to transmit olfactory information. *Perception and Psychophysics* 16:551–56

Dorcus, R M. 1932 Habitual word association to colors as possible factor in advertising *J App Psychol* 16.277–87

Doty, R L., S. Applebaum, H Zusho, and R G Settle 1985. Sex differences in odor identification ability: A cross-cultural analysis *Neuropsychologia* 23 667–72.

Doty, R L, P. Shaman, and M Dann 1984 Development of the University of Pennsylvania Smell Identification Test A standardized microencapsulated test of olfactory function *Physiol Beh* 32 489–502

Engen, T. 1982. *Perception of Odors*. Academic Press

Engen, T., J. E. Kuisma, and P. D. Eimas. 1973 Short-term memory of odors *J. Exp Psychol.* 99:222–25.

Engen, T., and C Pfaffmann 1960. Absolute judgments of odor quality. *J Exp. Psychol* 59.23–26.

Engen, T., and B. M. Ross. 1973 Long-term memory of odors with and without verbal descriptions. *J. Exp Psychol* 100:221–27

Eriksen, C. W, and H. W Hake 1955 Multidimensional stimulus differences and accuracy of discrimination *J Exp. Psychol.* 50 153–60.

Gesteland, R C. 1986 Speculations on receptor cells as analyzers and filters. *Experientia* 42:287–91

Halliday, M. A. K., and R. Hasan 1976 *Cohesion in English*. Longman

Heywood, A, and H A Vortriede 1905 Some experiments on the associative power of smells *Am J Psychol* 16 537–41

Hyde, J S. 1981. How large are cognitive gender differences? A meta-analysis using w and d. *Am. Psychol* 36 892–901

Jones, F N 1968 Information content of olfactory quality In *Theories of Odors and Odor Measurement*, ed. N Tanyolac, pp 133–46 Bebek, Turkey. Robert Coll Res Center

Lawless, H., and T. Engen. 1977 Associations to odors Interference, mnemonics, and verbal labeling *J Exp Psychol: Human Learning and Memory* 3.52–59

Mair, R. G., J. A. Bouffard, T. Engen, and T H Morton. 1978. Olfactory sensitivity during the menstrual cycle. *Sensory Processes* 2:90–98.

Miller, G. A 1953 What is information measurement? *Am Psychol* 8:3–11.

——— 1969 The organization of lexical memory: Are word associations sufficient? In *The Pathology of Memory*, ed G A Talland and N C Waugh, pp 223–36 Academic Press

Miller, G. A, and P. N. Johnson-Laird 1976 *Language and Perception* Harvard Univ Press

Posner, M I. 1969. Representational systems for storing information in memory. In *The Pathology of Memory*, ed G. A Talland and N. C Waugh, pp. 173–95 Academic Press

Rabin, M D, and W. S. Cain 1984 Odor recognition Familiarity, identifiability, and encoding consistency. *J Exp Psychol Learning Memory, and Cognition* 10.316–25

Sumner, D. 1962 On testing the sense of smell *Lancet* 2 895–97

LABORATORIO No. 1

PARTE B

IDENTIFICACION DE SABORES BASICOS

Objetivo:

Identificar el sabor básico dominante presente en cada una de las soluciones que se proveerán.

Procedimiento:

Diversas concentraciones de sustancias representando los sabores básicos han sido preparadas para que usted las examine. Las concentraciones usadas son las siguientes:

Dulce -	sucrosa	1.0% w/v	(2.5 g/250 ml)
Salado -	cloruro de sodio	0.2% w/v	(0.5 g/250 ml)
Acido -	ácido cítrico	0.04% w/v	(0.1 g/250 ml)
Amargo -	cafeína	0.05% w/v	(0.125 g/250 ml)

LABORATORIO No. 1

PARTE B

PERCEPCION DE SABORES BASICOS

Nombre:_______________

Fecha:_______________

Reconocimiento de sabores básicos

Pruebe, en orden descendente, cada una de las soluciones en el orden indicado en la boleta. Las soluciones pueden tener un gusto dulce, ácido, salado o amargo. Entre las soluciones con sabores básicos puede haber una o más muestras que tienen solamente agua. Identifique el sabor de la solución de cada uno de los vasos codificados. Enjuáguese la boca con agua antes de degustar y también entre una muestra y otra. Para aclarar la boca se proporcionan también galletas.

Código	Sabor
____________	____________
____________	____________
____________	____________
____________	____________
____________	____________
____________	____________

RECONOCIMIENTO DE SABORES BASICOS

Panelista No							% CORRECTO
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							

Chapter 5

CHEMICAL SENSATION: TASTE

Linda M. Bartoshuk

TASTE ANATOMY

Four kinds of projections, or papillae, can be identified on the human tongue. The filiform papillae, which contain no taste buds, give the tongue its rough, filelike appearance. The fungiform papillae, shaped like button mushrooms, appear primarily on the tip of the tongue and along the edges of the front of the tongue. Taste buds are buried in the surface epithelium of these papillae. The foliate papillae consist of a set of 3 to 8 parallel folds on the rear edges of the tongue, with taste buds buried in the sides of the folds. The circumvallate papillae, nine to ten in number, are shaped like flat mounds surrounded by a trench; they form an inverted V on the rear of the tongue. Taste buds lie in the sides of the papillae and in the wall of the surrounding trench. Humans between 20 and 70 years of age have about 250 taste buds in each circumvallate papilla and a total of about 1300 taste buds in the foliate papillae.[1] A human fungiform papilla can contain up to 8 taste buds.[2]

Taste buds are also found in the soft palate (the soft part of the roof of the mouth), the pharynx (throat), larynx, esophagus, and epiglottis (which closes off the larynx to prevent the entrance of food and fluids during swallowing).[1]

Taste buds are globular clusters of cells arranged something like the segments of an orange. The taste bud is surrounded by skin cells that appear to be the source for new taste cells. When the newly dividing epithelial cells of the rat taste bud were radioactively labeled, daughter cells were found to move into the taste bud at the rate of 1 every 10 hours. Examination at various intervals of time showed that cells within the taste bud had a life span of about 10 days.[1] During its life cycle, a cell moves from the edge to the center of the taste bud. Since the nerve fibers do not move, the receptor cells are presumably innervated by different nerve fibers as they change location. This poses a problem for stable quality perception.

Mammalian and fish taste buds require intact innervation, for if the peripheral taste nerve is cut or crushed, the taste buds will degenerate. As the nerve fibers regenerate, taste buds will reappear.

The sheep has been used to study the early development of the taste bud.[2] In sheep, as in humans, taste buds develop in utero. The early taste buds in sheep, called presumptive taste buds, have a nerve supply when they appear at about 49 days, with one presumptive taste bud per papilla. Later, a papilla may contain two taste buds, and by birth (about 147 days), 3 to 4 buds. Single fiber responses from the chorda tympani of fetuses aged from 109 days to term were similar to those of lambs and adult sheep. That is, as early as 5 weeks before birth, the responses appeared similar to those of the adult.

The early development of taste buds suggests that the fetus may taste substances in amniotic fluid. This is especially important for the evaluation of the effects of early experience on taste preferences. At birth, an organism already may have had a considerable amount of taste experience.[2]

Taste nerve fibers branch before forming synapses with receptor cells. The receptor cells, innervated by a single fiber, can even be in separate taste buds.

Taste information is transmitted via three nerves: the chorda tympani (part of the facila or VIIth cranial nerve), the glossopharyngeal (IXth cranial nerve), and the vagus (Xth cranial nerve). The chorda tympani nerve innervates the fungiform papillae, and the glossopharyngeal nerve innervates the foliate and circumvallate papillae. The innervation of the taste buds found on other oral structures is not as well established. Taste buds on the soft palate are innervated primarily by the greater superficial petrosal nerve (which is a branch of the facial nerve), although the glossopharyngeal and vagus nerves may also play a role.[4] Taste buds on the laryngeal surface of the epiglottis are innervated by the superior laryngeal nerve (which is a branch of the vagus nerve). These receptors have been implicated in the chain of events leading to sudden infant death syndrome.[5]

The three nerves mediating taste project ipsilaterally to the solitary nucleus of the medulla. Athough the areas to which they project overlap to some extent, the chorda tympani (cranial nerve VII) terminates in the rostral end of the nucleus, the vagus (cranial nerve X) in the caudal end, and the glossopharyngeal (cranial nerve IX) in an area between these two.[8]

A taste relay in the pontine area of the rat receives ipsilateral projections from both rostral and caudal locations in the solitary nucleus. This suggests the interesting possibility that a particular pontine neuron may receive inputs from two different tongue areas.[7] Some pontine neurons project to the dorsal thalamus, but others project through the substantia innominata to the central nucleus of the amygdala, an area that may be important for the regulation of feeding and drinking.[8] Neurons of a third group project to both areas.[9,10] Very little is known about the possibility of similar pontine projections in other species.

Taste is located near the tongue portion of the somesthetic projections in the thalamus. The body surface is represented somatotopically in the somesthetic system, with the tongue at the extreme anterior end; the thalamic taste area is located near the tip of the thalamic tongue area.[7] The thalamic taste area receives bilateral projections from the pontine taste area in the rat.[11]

Two cortical areas responsive to electrical stimulation of the chorda tympani and glossopharyngeal nerves have been identified in the squirrel monkey.[6] One of these is in the tongue portion of the somatotopic projection of the body surface in somatic sensory area I. Spatial localization on the tongue is maintained in this area. The second area, located just rostral to the first, is in the anterior operculatrous NaCl substitute for people on low sodium diets.

If anions meet the structural requirements for bitter or sweet (see below), then this taste is added to the taste of the cations. For example, sodium benzoate is salty, sweet, and bitter because the sodium ion is salty and the benzoate ion is sweet and bitter

Salt Substitutes. Since the salty taste is produced by the positively charged cation in a salt, we look for substitutes among the nonsodium cations that are unlikely to be toxic and that are similar to sodium in size The ions most similar to sodium are lithium and potassium Lithium chloride cannot be used as a salt substitute because it is toxic. This leaves potassium chloride, which is the major ingredient in currently marketed salt substitutes. The bitter taste of KCl makes it unacceptable to many. However, the amount of bitterness perceived varies among individuals (see the discussion, below, on PTC sensitivity).

Sweet

Sweetness is produced by molecules with various structures. The familiar compounds include sugars, sugar alcohols, amino acids (D forms are usually sweet), and assorted organic compounds, including those either in use or of interest as artificial sweeteners (e.g., saccharin, cyclamate, neohesperidine dihydrochalcones, and aspartame). In addition, salts may taste sweet as well as salty. Lead and beryllium salts are dramatic examples; many salts, including NaCl, are sweet at low concentrations.

Despite the apparent diversity, there is a chemical structure common to these compounds Shallenberger and Acree[14] noted that sweet molecules contain what they called an AH, B system. A and B, separated by 3 Å, are portions of the molecule capable of forming hydrogen bonds. The A portion shares its hydrogen atom with the receptor membrane, while the membrane shares a hydrogen atom with B. A third portion, (γ), added by Kier and Shallenberger,[15] makes the sweet unit a triangle. Figure 5-2 shows the sweet triangle in D-glucose.

There are issues yet to be resolved. For example, some molecules contain more than one potential site for the triangle; therefore, the actual source of sweetness is not clear.

Bitter

Like sweetness, bitterness is produced by some inorganic and a variety of organic compounds. Although there is no single chemical structure associated with bitter, some generalizations concerning chemical structure and bitterness do exist. The alkaloids found naturally in plants (e g , caffeine, nicotine, solanine, cocaine, and strychnine) are generally bitter. The toxicity of these substances has been cited as a possible environmental pressure that may have promoted the evolution of the aversion to bitterness found in so many species.

Glycosides, derived from certain sugars, are also found in plants.[11] Some of the glycosides have been studied in detail to determine the effects of substitutions at various sites on the molecules.[16] These studies are very important to the development of any comprehensive theory of bitterness. However, they have also had a more practical impact. Certain substitutions produce compounds that can produce strong sweet tastes at low concentrations, e g., neohesperidine and naringin dihydrochalcone. These compounds may be useful as nonnutritive sweeteners

Another group of organic compounds found in plants, diterpenes, has also been analyzed for structure-bitterness relations.[17] An AH, B system similar to that of Shallenberger was suggested as the common link. However, in this case the AH and B are separated by 1.5 Å instead of 3 Å, which produces an intramolecular hydrogen bond between AH and B.

A variety of additional structure-bitterness generalizations have been suggested.[18] It is still too early to tell whether any or all of these can be related. However, no synthesis appears likely to occur soon.

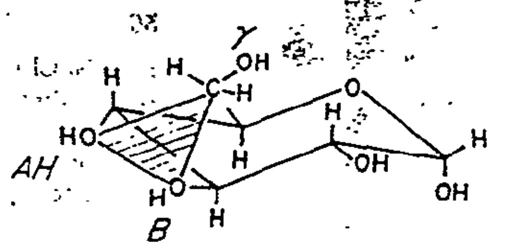
Taste Blindness to the Bitterness of PTC. PTC (phenylthiocarbamide or phenylthiourea) and other compounds containing the H-N-C=S grouping are of special interest because sensitivity to their bitter taste is genetically mediated by a single mendelian dominant gene. An individual with two recessive genes in his genotype has a high tolerance threshold (and low sensitivity) to PTC bitterness, leading to the classification *nontaster* or *taste blind*, while an individual with one or both dominant genes in his genotype has a low tolerance threshold (and high sensitivity) to PTC bitterness, leading to the classification *taster*. Saliva is often related to PTC tasting because a very early study[19] seemed to show that a taster had to have his saliva present in order to taste PTC. This is not supported by modern data.[20]

The exact requirements for the molecular grouping responsible for the taste of PTC have been questioned,[21,22] since molecules that do not contain exactly the H-N-C=S grouping, e.g., anetholtrithione and caffeine, have been found to produce a bimodal threshold distribution that is essentially the PTC distribution.[23] Of even more practical interest is the relation between PTC status and the bitterness of saccharin. Tasters of PTC tend to perceive more bitterness in saccharin than nontasters do at concentrations of saccharin used in low calorie foods and beverages.[24] Similarly, tasters of PTC tend to perceive more bitterness in potassium chloride and in sodium and potassium benzoate.[25]

Adaptation

Self-Adaptation

Adaptation refers to the diminished sensitivity that results from prolonged taste stimulation. Adaptation to saliva provides a good example of the completeness of adaptation in taste. Saliva contains enough NaCl to taste salty, yet most individuals do not notice a chronic salty taste in their mouths. This is because we are adapted to the NaCl in saliva. If saliva is removed with a thorough water rinse, the concentration of NaCl in saliva can be tasted.



D-GLUCOSE

Fig. 5–2. D-Glucose in the "chair" conformation. The AH, B, γ triangle that forms the sweet unit in Shallenberger's theory is indicated.

Taste adaptation tends to be complete (i.e., sensation disappears) to moderate intensities of sour, sweet, and bitter stimuli under laboratory conditions that keep stimulation truly constant. Under normal eating conditions, adaptation does not occur because the concentrations of the taste stimuli keep varying as foods and beverages are moved in the mouth and diluted with saliva.

Adaptation to the NaCl in saliva has important clinical consequences. An individual can only taste NaCl when its concentration is higher than the Na^+ concentration in saliva. Figure 5–3 shows an example of this. The data were obtained with the method of magnitude estimation (discussed below). The saltiness of NaCl is shown under three adaptation conditions: tongue rinsed with water to remove saliva, tongue adapted to resting saliva, and tongue adapted to the saliva produced by vigorous chewing. Chewing causes an increase in salivary flow rate. Since the sodium content of saliva increases with flow rate, chewing increases the salivary

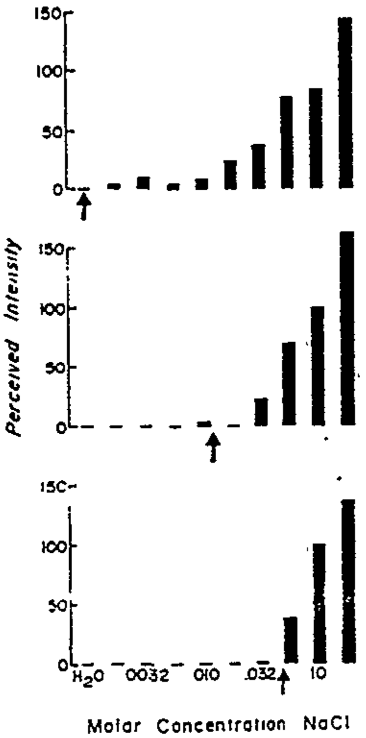


Fig. 5–3. Saltiness of NaCl (±1 S.E.) when the tongue is rinsed with water (*top panel*), when the tongue is adapted to resting saliva (*middle panel*, salivary Na indicated by *arrow*), and when the tongue is adapted to stimulated saliva (*bottom panel*). (From Bartoshuk, L.M.: The psychophysics of taste. Am. J. Clin. Nutr., *31*:1068, 1978.[59] By permission of American Society for Clinical Nutrition, © Am. J. Clin. Nutr.)

sodium concentration. Note that the effects of adaptation are greatest at concentrations just above the adapting concentration.

Cross-Adaptation

If adaptation to one substance decreases sensitivity to another substance, we say that cross-adaptation has occurred.

McBurney and his colleagues[26–28] have studied cross-adaptation within groups of substances that share a taste quality. Their results produced remarkably simple generalizations based on perceived taste quality. Salty substances cross-adapt. Similarly, sour substances cross-adapt. The sweet substances so far tested tend to cross-adapt, but a number of substances are still untested. Certain bitter stimuli show no cross-adaptation, but others do cross-adapt.

McBurney's results show impressive agreement with the chemistry of the taste qualities. That is, the tastes that cross-adapt are believed to have the same receptor mechanisms. A common chemical characteristic has been proposed for saltiness (i.e., the charge on the cation), and cross-adaptation does occur. Similarly, sour substances have hydrogen ions in common and cross-adapt. For bitter substances, a variety of chemical characteristics have been proposed, and cross-adaptation clearly fails for certain bitters. For sweet substances, a common chemical characteristic has been proposed, and cross-adaptation occurs among many sweet substances, but the possibility remains that sweet (like bitter) may prove to be heterogeneous.

Cross-adaptation does not generally occur across substances of different taste qualities.[29] This is additional support for the idea that the taste qualities are mediated by different receptor mechanisms.

The Taste of Water

We now know that a long-standing controversy over the "intrinsic" taste of water is related to adaptation. Water was initially considered to be a tasteless solvent, but by the 19th century several authors had expressed doubts about the tastelessness of distilled water. Öhrwall[70] and Henle[71] argued that distilled water is flat rather than tasteless. Several investigators[72–74] reported individual subjects who found that distilled water tasted bitter. In 1914, Brown[75] recorded evaluations of the taste of water from 100 subjects; he concluded that "The upshot of all these observations seems to be that water is not tasteless (in the broader sense of the word), that it tastes more like bitter than anything else, and that its taste can be neutralized by salt. . . . Furthermore, there is evidence that some of the weakest solutions of salt in water give a substance which is as nearly tasteless as any which can be produced" (pp. 253, 254).[35] Brown's comments proved prophetic. The amount of salt he recommended adding to "remove" the bitter taste of water bracketed the normal concentration of NaCl in saliva, the solution to which one is normally adapted.

As it turns out, the taste of water varies, depending on what precedes it on the tongue. Under normal circumstances, saliva bathes the tongue. Figure 5–4 shows that adaptation to saliva makes water taste bitter and sour. These results can be duplicated with pure NaCl as the adapting solution instead of saliva.[36] Thus the NaCl in saliva is viewed as the probable source of the apparent "intrinsic" taste of distilled water. This phenomenon can become particularly important in diseases that alter salivary electrolytes.

Figure 5–5 shows another water taste: the sweet taste of water following adaptation to various concentrations of the potassium salts of chlorogenic acid and cynarin, substances found in the globe artichoke. The sweetness of sucrose is shown on the same function for purposes of comparison. This effect is particularly interesting because not all individuals are equally sensitive to it.

In summary, all four basic taste qualities can be produced with a water stimulus by adapting to appropriate substances, but the most common water taste is sweet.[17]

Mixtures

When two lights of different colors are mixed, the resulting color is qualitatively distinct from the components and falls between them on the color circle. For example, red light mixed with yellow light fuses into orange.

When substances with different tastes are mixed, their taste qualities do not fuse but rather can usually be recognized. The intensities of the component tastes are often decreased (mixture suppression) and occasionally may even increase (enhancement or synergism). These interactions play an important role in the use of condiments or flavoring ingredients in foods. For example, the addition of salt to food not only adds a salty taste but also changes the perceived intensities of other tastes in the food.

There have been many attempts to formulate rigorous rules for mixture interactions, but none of these efforts have been completely successful. One reason for this is that mixture interactions might occur both in the stimulus itself (as chemical interactions among components) and at different levels of the nervous system.[31,38,39]

Clinically, mixture interactions are important for patients who have differential taste quality losses. The tastes that are no

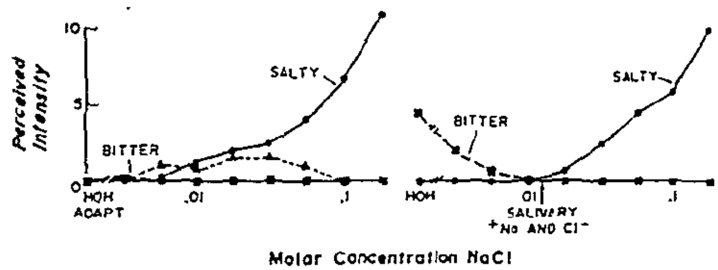


Fig. 5–4. Tastes of NaCl and water when the tongue is adapted to water (*left panel*) and when the tongue is adapted to saliva (*right panel*).

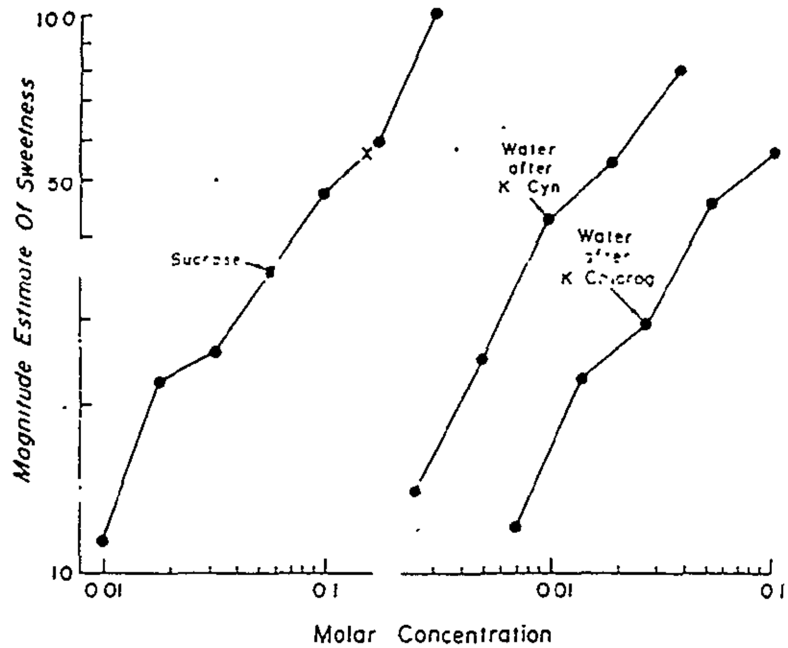


Fig. 5–5. A comparison of the sweetness of sucrose with the sweetness of water following adaptation to the potassium salts of cynarin and chlorogenic acid (constituents of globe artichoke). The X on the sucrose function marks the sweetness of 2 teaspoons of sucrose in 6 ounces of water (0.16 M). (From Bartoshuk, L.M., Lee, C.H., and Scarpellino, R.: Sweet taste of water induced by artichoke (*Cynara scolymus*). Science, *178*:988, 1972.[10] Copyright 1972 by the AAAS.)

longer perceived may fail to affect the remaining tastes in the normal manner. In this way a loss that is actually very specific to one quality can change the intensities of other qualities.

CLINICAL EVALUATION OF TASTE

Thresholds

Most clinical studies of taste function have been done with threshold measurements. In its simplest form, the threshold for a substance is that concentration that can be tasted 50% of the time; unfortunately, the actual procedures are anything but simple to carry out correctly. In addition to these problems, threshold measurements have an even more serious deficiency: they do not always mean what they seem to mean.

Figure 5–6 shows data collected in our laboratory on how the perceived intensity of NaCl grows with concentration. The three graphs reflect three different experiments. The functions in the upper left section show that sodium lauryl sulfate, the detergent commonly used in toothpaste, reduces the saltiness of NaCl by a constant percentage across all concentrations. The functions in the lower left section show that adaptation to the sodium in saliva has only a local effect on the saltiness of NaCl. The tongue becomes less sensitive to the adapting concentration, and higher concentrations show a diminishing effect. The functions in the lower right section show that radiation therapy (radiation applied to the head and neck region) reduced sensitivity to NaCl in an unusual way. Threshold remained normal, but higher concentrations (those actually encountered in the real world) were less salty than normal.

These three graphs illustrate the folly of attempting to extrapolate to real experi-

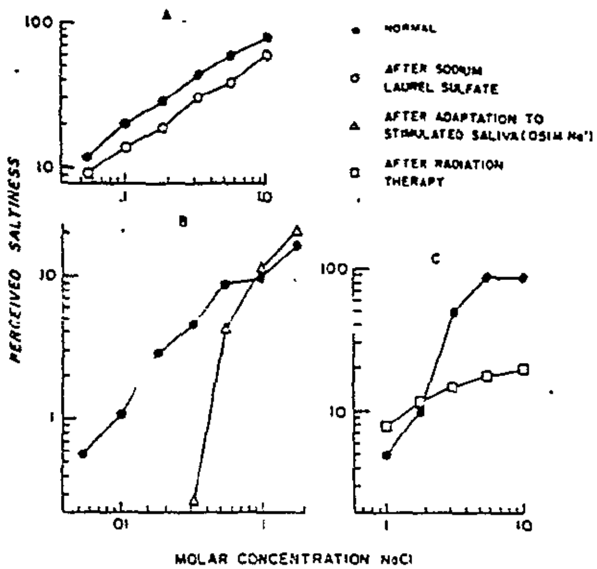


Fig. 5–6. The before and after comparisons of the taste of NaCl were all made within individuals. *A*, Saltiness of NaCl before and after administration of the detergent, sodium lauryl sulfate (data from DeSimone, J.A., Heck, G.L., and Bartoshuk, L.M.: Surface active taste modifiers: A comparison of the physical and psychophysical properties of gymnemic acid and sodium lauryl sulfate. Chem. Senses, *5*:317, 1980[55]). *B*, Saltiness of NaCl before and after adaptation to saliva. *C*, Saltiness of NaCl before and after radiation therapy (data are averages of first two and last two sessions, from Bartoshuk, L.M.: The psychophysics of taste. Am. J. Clin. Nutr., *31*:1068, 1978.[53] By permission of American Society for Clinical Nutrition. © Am. J. Clin. Nutr.).

ence from thresholds. Thresholds and the perceived intensities of higher concentrations need not be correlated because the slopes of psychophysical functions can vary.

Threshold measures can be of great value, but they are of most use in laboratory investigations of taste. Clinical assessments whose aim is to explain anomalous taste experiences need procedures that produce a scale of perceived intensity.

Suprathreshold Scales

Psychophysical functions, or plots of perceived intensity versus concentration, can be obtained by several methods. For example, a 9-point category scale is commonly used in food studies. Subjects rate intensity by assigning numbers from this scale, e.g., 1 is weakest and 9 is strongest. Category scales have one serious limitation: the numbers have only ordinal properties. That is, an 8 on such a scale cannot be assumed to represent an intensity that is necessarily twice as strong as a 4. Eight is simply more intense than 4.

Ratio methods avoid this difficulty. In one of the most popular of these methods, magnitude estimation, subjects are asked to assign numbers that are proportional to perceived intensity. Thus if a certain concentration of sucrose were called 4, then another concentration tasting twice as sweet would be called 8.[40]

Taste functions generated with this method tend to be power functions of the form, $\psi = \phi^{\beta}$, where ψ = perceived intensity, ϕ = concentration, and β varies, depending on the stimulus and the way in which it is tasted. If we take the logarithm of both sides, we have $\log \psi = \beta \log \phi$. This means that if we plot $\log \phi$ on the x axis and $\log \psi$ on the y axis, the function will be a straight line of slope β. (See Figure 5–6 for examples of such plots.) The slope of these functions provides a very simple index of how perceived intensity grows with concentration.

Psychophysical functions provide more information than thresholds, but there are important limitations to what can be inferred from the functions. The rates of growth of the functions (i.e., the exponents) can be compared across individuals. Absolute intensities cannot be compared.

Information generated by thresholds and psychophysical functions does not allow a direct comparison of perceived intensity across individuals. For example, two individuals might have identical thresholds and psychophysical functions for quinine with identical slopes, but we could still not conclude that a given concentration of quinine was equally bitter to both.

In response to the need for a method that can provide information about differences in absolute sensation magnitude across individuals, Stevens and Marks[41] have developed a new method—magnitude matching. In short, the method requires the subject to judge the sensation magnitudes of stimuli from two or more modalities that are presented within a single test session. The subject judges all of the sensations on a single, common scale of sensation magnitude, without regard to the modality. This method, used with taste, allows us to express taste intensity relative to another modality. For example, suppose individual *A* judges a particular sound to be twice as intense as a particular NaCl solution, while individual *B* judges the two to be equal. Their experiences of intensity cannot be identical for both stimuli. If both perceive the sound to have the same intensity, then *B* perceives the NaCl to be twice as strong as *A* does. Alternatively, if both perceive the taste to be the same, then *A* perceives the sound to be twice as strong as *B* does. Although we cannot know for sure which answer is correct, we can make one answer more and more reasonable by extending our testing to additional modalities.

TASTE ALTERATIONS IN DISEASE

Taste disorders are of interest for several reasons. They disturb many individuals and can lead to a loss of morale because of the loss of an important source of pleasure. Changes in food and fluid intake caused by taste changes might be expected to have serious nutritional consequences. Taste may also have diagnostic value or may at least provide some clues concerning the nature of some underlying disease.

The taste problems that have been described can be divided roughly into reports of sensory gain, sensory loss, distorted tastes, and the presence of unpleasant tastes. The discussion above on normal taste function provides some guidelines concerning the interpretation of these problems.

Sensory Gain

A few studies actually suggest increased taste sensitivity as a result of disease. However, the validity of such observations has been questioned.

Children with cystic fibrosis have been reported to show unusual taste sensitivity to NaCl.[42] However, other investigators have been unable to replicate these results.[43–45]

Addison's disease (adrenal insufficiency) has been reported to produce unusually low taste thresholds.[46,47] Taste sensitivity in individuals with Addison's disease has not, to the best of my knowledge, been re-examined since these early results. However, the validity of the result has been challenged on logical grounds[36] for the following reasons. Henkin and coworkers[47] have commented: "At the detection threshold, all of the test substances tasted bitter to the patients with adrenal insufficiency" (p. 732). Water (and dilute solutions) taste bitter if the tongue is adapted to NaCl (see the section above on the taste of water). Addison's disease raises the concentration of sodium in saliva,[48] and the testing method employed in the early study did not include rinses between stimuli. This suggests the possibility that the patients with Addison's disease, adapted to their abnormal saliva, were able to discriminate low concentrations of NaCl from water because they tasted the water as unusually bitter.

Salivary sodium is also elevated in cystic fibrosis. The low NaCl thresholds reported in that disease[12] may also have actually been water taste thresholds.

The appetite loss often resulting from cancer has been linked to a gain in taste sensitivity. DeWys and Walters[49] reported that some cancer patients have unusually high recognition thresholds for the bitter taste of urea. They suggested that increased bitter thresholds might produce an aversion to meat because meat contains small amounts of bitter substances. However, DeWys and Walters apparently made a statistical error.[50] They reported that 8 out of 50 cancer patients and 1 out of 23 controls had recognition thresholds below 0.09 M urea. These data yield a chi square value of 1.05 and not 10.5 as stated in their article. The value 10.5 would have been statistically significant, but 1.05 is not.

In summary, there are no generally accepted cases in which taste function is actually enhanced as the result of disease.

Sensory Loss

Sensory loss is clearly real, can have a variety of causes, and can take a variety of forms.

Several diseases have been suggested as causes for taste losses. For example, although the report of increased sensitivity to urea in cancer patients was not valid, reports of decreased sensitivity may be valid. Difficulties raised by the study of cancer patients illustrate another problem that results from attempts to link taste losses to disease. The therapies rather than the disease per se may be responsible for the sensory loss.

Radiation of the head and neck for cancer provides an example. Taste function is reduced or even abolished following radiation to certain areas. For many subjects there is at least partial recovery. However, the course of that recovery is unusual. Figures 5–7 and 5–8 show threshold and scaling data obtained from a woman undergoing radiation therapy. Note that her

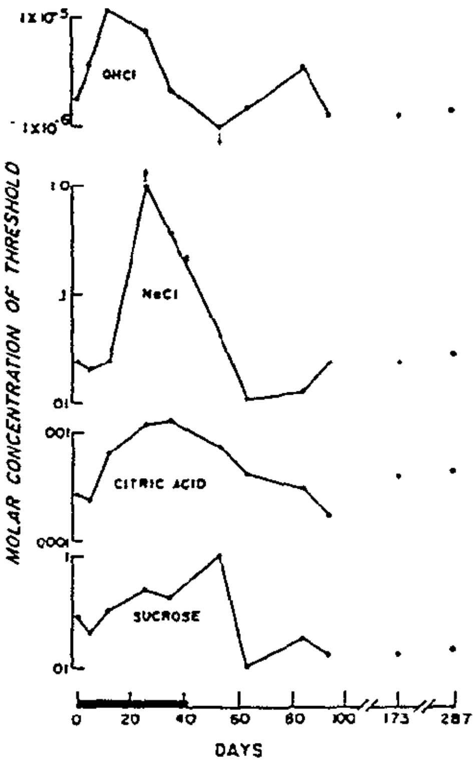


Fig. 5–7. Recognition thresholds (up-down, forced choice) obtained from a 52-year-old woman undergoing radiation therapy for cancer of the neck. The *black line* on the abscissa shows the time during which the patient was receiving radiation therapy. (From Bartoshuk, L.M.: The psychophysics of taste. Am. J. Clin. Nutr. *31*: 1068, 1978.[51] By permission of American Society for Clinical Nutrition, © Am. J. Clin. Nutr.)

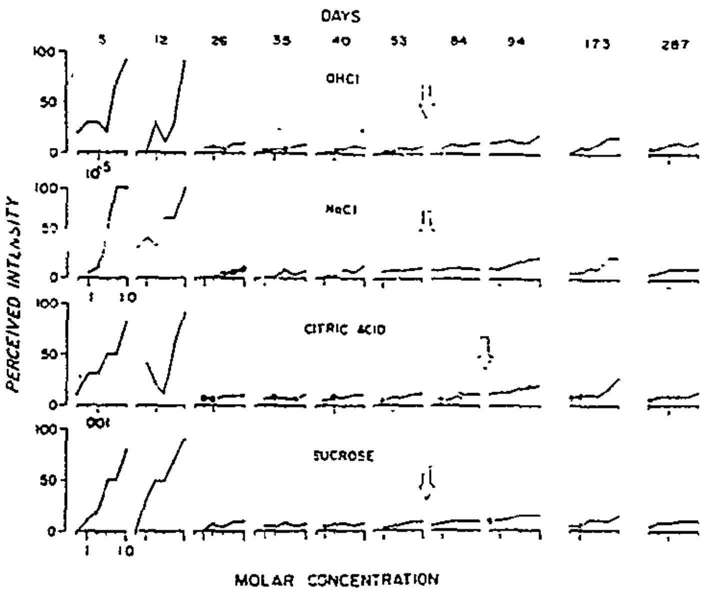


Fig. 5–8. Psychophysical functions for the patient whose thresholds are shown in Fig. 5–7. The *arrows* indicate the time at which the thresholds returned to normal. The *circled points* indicate an atypical quality report (e.g., "bitter" for sucrose). (From Bartoshuk, L.M.: The psychophysics of taste. Am. J. Clin. Nutr., *31*:1068, 1978. By permission of American Society for Clinical Nutrition, © Am. J. Clin. Nutr.)

thresholds had returned to normal after approximately 2 months, but her suprathreshold scales had not. They remained flattened and although this woman could taste, she lived in a pastel taste world. Mossman and Henkin[51] also observed this pattern of loss.

Radiation effects were used earlier (see Fig. 5–6) to illustrate how perceived intensities could be reduced even though thresholds were normal. The fact that thresholds are not infallible predictors of real world taste experiences limits the utility of studies of the effects of disease on taste because virtually all these studies used threshold measures. In addition, many of the older studies used threshold methodology that does not meet modern standards.

One of the simplest causes of taste loss is the severing of a nerve. The chorda tympani taste nerve crosses the eardrum on its way from the tongue to the brain. During certain types of ear surgery, the nerve may be cut. The front of the tongue on the same side as the surgery becomes insensitive to taste. Interestingly enough, the patient may hardly notice such a loss even if both chorda tympani nerves are cut. Input from the other taste nerves, the glossopharyngeal and the vagus, still leaves the patient with considerable taste function.

Bell's palsy can produce a taste loss because of damage to the facial nerve (the chorda tympani is a branch of the facial nerve).

Zinc deficiency has been reported to produce reductions in taste function. But the status of zinc remains controversial, in part because a single-blind study involving zinc administration produced taste improvement while a double-blind study failed to do so.[52] Henkin and his colleagues concluded, "The results indicate that treatment of unselected patients with taste and smell disorders with zinc sulfate or placebo is effectively equivalent." However, as they also noted, there may still be individuals in whom zinc deficiency is related to taste loss. This area remains of great interest to taste investigators and deserves further study.

Advancing age is often assumed to produce taste losses. However, this is based on a series of threshold studies. In fact, the threshold elevations are relatively small and in any case should not be used to predict losses in perceived intensity at higher concentrations. More recent experiments suggest that many healthy elderly people show no obvious taste losses,[33,34] but there are definitely taste losses in particularly elderly individuals. The essential question is whether age per se produces loss or whether the elderly have had more time to encounter a disorder leading to taste loss. The answer remains uncertain.

Smoking is often assumed to produce taste losses. However, threshold studies show essentially no taste differences before and after smoking. Scaling studies are needed to resolve this problem. A taste loss like that shown by the radiation data in Figure 5–8 could exist and be undetected by the threshold studies.

Distorted Tastes

Patients sometimes report that certain foods have abnormal tastes. This may reflect a differential loss of one or more qualities. For example, ketchup could taste unusually sour to an individual with a reduced ability to taste sweet, since it contains relatively large amounts of sugar that normally suppress the sour taste of the acids.

Detergents can produce temporary taste distortions. Sodium lauryl sulfate, a detergent commonly used in toothpastes and mouthwashes, can reduce the ability to taste sweet and add a bitter taste to acids.[35] This produces the well-known "orange juice effect": after brushing the teeth, orange juice tastes bitter as well as less sweet than usual.

Dysgeusia: The Presence of Unpleasant Tastes

This symptom is probably the most distressing to patients with taste symptoms, and it is also the least understood. Although there have been attempts to relate dysgeusia to abnormalities of the taste receptor membrane, such symptoms could have an entirely different origin. They may reflect the abnormal presence in the mouth of molecules with noxious tastes. There are several possible sources. Exhaled air may deposit molecules with unpleasant tastes. Saliva can contain substances with unpleasant tastes. In particular, some medications may be present in saliva. Another potential source of unpleasant tastes is crevicular fluid, the fluid in the gums near the teeth.[36] This fluid may be expressed into the mouth, especially by chewing. This suggests that dysgeusia might be treated by improving oral health.

One final source of noxious tastes is suggested by research on the venous taste phenomenon. Circulation time was once evaluated by injecting saccharin intravenously. It traveled to the mouth (thus giving a measure of circulation efficiency) and was then tasted. Electrophysiologic studies suggest that the saccharin is tasted in the blood and does not have to reach the taste receptors on the tongue surface. The possibility that disease states can introduce tasteable substances into the blood or that clinical treatment (e.g., chemotherapy) might do so should be evaluated.

REFERENCES

1. Bradley, R.M., Tongue topography. *In* Handbook of Sensory Physiology. Edited by L.M. Beidler. Vol. IV, Chemical Senses Section 2. Taste. New York, Springer-Verlag, 1971, pp. 1–30.
2. Bradley, R.M., and Mistretta, C.M. Fetal sensory receptors. Physiol. Rev., 55:352, 1975.
3. Beidler, L.M., and Smallman, R.L. Renewal of cells within taste buds. J. Cell Biol. 27:263, 1965.

BEST AVAILABLE

4. Cleaton-Jones, P : A denervation study of taste buds in the soft palate of the albino rat. Arch. Oral Biol., *21*:79, 1976.
5. Storey, H.T., and Johnson, P.: Laryngeal receptors initiating apnea in lambs. *In* Development of Upper Respiratory Anatomy and Function. Edited by J.F. Bosma and J. Showacre. Bethesda, Md.: U.S. DHEW, 1975, pp. 184–198.
6. Burton, H , and Benjamin, R.M.: Central projections of the gustatory system. *In* Handbook of Sensory Physiology. Vol IV, Chemical Senses Section 2· Taste. Edited by L.M. Beidler. New York, Springer-Verlag, 1971, pp. 148–164.
7. Norgren R., and Pfaffmann, C.: The pontine taste area in the rat. Brain Res , *91*.99, 1975.
8. Norgren R.: Taste pathways to hypothalamus and amygdala. J. Comp. Neurol , *166*:17, 1976.
9. Norgren, R., and Leonard, C M : Ascending central gustatory pathways. J. Comp Neurol., *150*: 217, 1973.
10. Norgren, R.: Gustatory afferents to ventral forebrain. Brain Res., *81*:285, 1974.
11. Ganchrow, D., and Erickson, R.P.. Thalamocortical relations in gustation. Brain Res., *36*: 289, 1972.
12. Norgren, R , and Wolf, G.· Projections of thalamic gustatory and lingual areas in the rat. Brain Res , *92*.123, 1975.
13. Pfaffmann, C Gustatory afferent impulses. J. Cell. Comp Physiol., *17*:243, 1941.
14. Shallenberger, R.S., and Acree, T.E.: Chemical structure of compounds and their sweet and bitter taste. *In* Handbook of Sensory Physiology IV. Chemical Senses Section 2: Taste. Edited by L.M. Beidler. New York, Springer-Verlag, 1971, pp. 221–277.
15. Shallenberger, R.S.: Chemical clues to the perception of sweetness. *In* Sensory Properties. Edited by G.G. Birch, J.G. Brennan, and K.J. Parker. London, Applied Science Publishers, Ltd., 1977, pp 91–100.
16. Horowitz, R.M., and Gentili, B Taste and structure in phenolic glycosides J Agric. Food Chem , *17*:696, 1969
17. Kubota, T., and Kubo, I.. Bitterness and chemical structure. Nature, *223*:97, 1969.
18. Moncrieff, R W.: The Chemical Senses. London, Leonard Hill, 1967.
19. Cohen, J , and Ogden, D.P.. Taste blindness to PTC as a function of saliva. Science, *110* 532, 1949
20. Hall, M.J , Bartoshuk, L.M , Cain, W S , and Stevens, J.C : PTC taste blindness and the taste of caffeine. Nature, 253:442, 1975
21. Fischer, R. Gustatory, behavioral and pharmacological manifestations of chemoreception in man. *In* Gustation and Olfaction Edited by G. Ohloff and A.F. Thomas New York, Academic Press, 1971, pp. 187–237.
22. Beets, M.G.J.. Structure-Activity Relationships in Human Chemoreception. London, Applied Science Publishers, Ltd., 1978.
23. Kalmus, H.: Genetics of taste. *In* Handbook of Sensory Physiology. Vol. IV. Chemical Senses Section 2· Taste. Edited by L.M Beidler New York, Springer-Verlag, 1971, pp. 165–178.
24. Bartoshuk, L.M : Bitter taste of saccharin: Related to the genetic ability to taste the bitter substance 6-n-propylthiouracil (PROP). Science, *205*·934, 1979.
25. Rifkin, B , and Bartoshuk, L.M. The bitter taste of potassium chloride and sodium benzoate is related to the genetic ability to taste 6-*n*-propylthiouracil. Paper presented at the Meeting of the Association for Chemoreception Sciences Sarasota, Florida, 1980
26. Smith, D.V , and McBurney, D.H. Gustatory cross adaptation. Does a single mechanism code the salty taste? J Exp Psychol , *80*·101, 1969
27. McBurney, D H , Smith, D V , and Shick, T R : Gustatory cross adaptation Sourness and bitterness Percept Psychophys *11* 228 1972
28. McBurney, D H Gustatory cross adaptation between sweet tasting compounds Percept. Psychophys , *11*.225, 1972
29. McBurney, D H , and Bartoshuk, L M Interactions between stimuli with different taste qualities. Physiol. Behav., *10* 1101, 1973.
30. Öhrwall, H Untersuchungen über den Geschmackssinn Skandinav Archiv Physiol., 2.1. 1891.
31. Henle, J.: Ueber den Geschmackssinn Anthropologische Vortrage, 2 3, 1880
32. Camerer W : Ueber die Abhängigkeit des Geschmackssinns von der gereizten Stelle der Mundhöhle. Z. Biol , *6*·440, 1870.
33. Kiesow, F · Beitrage zur physiologischen Psychologie des Geschmackssinnes. Phil. Stud , *10*.523, 1894
34. Skramlik, E von: Mischungsgleichungen im Gebiete des Geschmackssinnes. Z. Psychol. Physiol. Sinnesorgane, *53*:36, 1922
35. Brown, W.: The judgment of very weak sensory stimuli University of California Publications in Psychology, *1* 199, 1914.
36. Bartoshuk, L M · NaCl thresholds in man: Thresholds for water taste or NaCl taste? J Comp Physiol Psychol , *87*.310 1974
37. McBurney, D.H , and Shick, T.R. Taste and water taste of twenty-six compounds for man Percept Psychophys , *10*·249, 1971
38. Lawless, H T : Evidence for neural inhibition in bittersweet mixtures. Ph.D. thesis, Brown University, 1978
39. Bartoshuk, L.M · Sensory analysis of the taste of NaCl. *In* Biological and Behavioral Aspects of Salt Intake. Edited by M Kare, M.J Fregley, and R.A Bernard New York, Academic Press, 1980. pp 83–98
40. Marks, L.E.. Sensory Processes. The New Psychophysics. New York, Academic Press, 1974
41. Stevens J C., and Marks L.E Cross-modality matching functions generated by magnitude estimation Percept. Psychophys., 27 379, 1980
42. Henkin, R I , and Powell, G F · Increased sensitivity of taste and smell in cystic fibrosis Science, *138* 1107, 1962
43. Wotman, S., Mandel, I D., Khotim, S , Thompson, R.H., Kutscher, A H., Zegarelli, E V., and Denning C R : Salt taste thresholds and cystic fibrosis. Am. J Dis Child., *108* 372, 1964
44. Hertz, J., Cain, W S , Bartoshuk, L M , and

PRUEBA DE
EVALUACION DE DUREZA

A continuación se le presentan 9 productos con diferente grado de dureza. El #1 debe ser el más sueve y el #9 el más duro.

Sabiendo que dureza es definida como la fuerza necesaria para comprimir un alimento entre las muelas o entre la lengua y el paladar, evalúe cada una de las muestras y comente si está de acuerdo o no con la escala que se le dió.

Ponga la muestra entre los dientes molares y evalúe la fuerza necesaria para comprimirlo.

Valor en Escala	Producto	Comentarios
1	Queso Crema	____________
2	Clara de huevo	____________
3	Queso Americano	____________
4	Salchicha	____________
5	Aceituna	____________
6	Manías	____________
7	Zanahoria	____________
8	Almendras Peladas	____________
9	Confite/duro	____________

CLASSIFICATION OF TEXTURAL CHARACTERISTICS

MECHANICAL

Primary	Secondav
Hardness	
Cohesiveness	Fracturability
	Chewiness
	Gumminess
Viscosity	
Springiness	
Adhesiveness	

GEOMETRICAL

Particle size and shape
Particle shape and orientation

OTHER

Moisture content	
Fat content	Oiliness
	Greasiness

Ref. Szczesniak 1963, J. Food Sci. 28,385-389.

Relations between textural properties and popular nomenclature

Mechanical characteristics

Primary parameters	Secondary parameters	Popular terms
Hardness		Soft → Firm → Hard
Cohesiveness	Brittleness	Crumbly → Crunchy → Brittle
	Chewiness	Tender → Chewy → Tough
	Gumminess	Short → Mealy → Pasty → Gummy
Viscosity		Thin → Viscous
Springiness		Plastic → Elastic
Adhesiveness		Sticky → Tacky → Gooey

Geometrical characteristics

Class	Examples
Particle size and shape	Gritty, Grainy, Coarse, etc.
Particle shape and orientation	Fibrous, Cellular, Crystalline, etc.

Other characteristics

Primary parameters	Secondary parameters	Popular terms
Moisture content		Dry → Moist → Wet → Watery
Fat content	Oiliness	Oily
	Greasiness	Greasy

Ref. Szczesniak, 1963 J. Food Sci. 28, 385-389.

DEFINITIONS OF MECHANICAL TEXTURE TERMS

	ORIGINAL (physical/sensory)	MODIFIED SENSORY
HARDNESS	force to attain a given deformation	force to compress between molar teeth, or between tongue and palate
COHESIVENESS	strength of internal bonds	amount of deformation with molars before rupture
VISCOSITY	rate of flow per unit force	force to draw liquid from spoon over tongue
SPRINGINESS	rate at which a deformed material goes back to its undeformed position upon force removal	perceived degree and speed of return to original height after partial compression with molar teeth
ADHESIVENESS	work to overcome attractive forces between food and other surfaces	tongue force to remove material adhering to mouth
FRACTURABILITY	force with which the material fractures	horizontal force with which the food moves away from the point of compressive, shattering forces
CHEWINESS	energy to disintegrate solid food to a state ready for swallowing	no. chews to masticate the food at 1 chew/sec
GUMMINESS	energy to disintegrate semi-solid food to a state ready for swallowing	denseness that persists through mastication

Ref. Szczesniak, 1963 J. Food Sci. 28, 38 5-389.

Civille and Liska, 1975 J. Texture Studies 6, 19-31.

EVALUACION DE ADHESIVIDAD

A continuación se le presentan 5 productos con diferente grado de adhesividad.

Lleve la muestra a la boca, presionela contra el paladar (con la lengua) y evalúe la fuerza necesaria para removerla del paladar. El #1 en la escala es el menos adhesivo, y el # 5 el más adhesivo.

Valor en Escala	Producto	Comentarios
1	Manteca Vegetal	____________
2	Crema de Trigo	____________
3	Turrón de Angeles	____________
4	Queso Crema	____________
5	Mantequilla de Maní	____________

*La distancia entre puntos de la escala puede no ser indéntica.

PRUEBA DE UMBRALES DE PERCEPCION

Nombre ________________ No. Panelista ______ Fecha ________

Usted recibirá una serie de concentraciones de muestras de un tipo de estímulo (dulce, salado, ácido o amargo). Las muestras están arregladas en orden ascendente en concentración.

Primero, enjuague su boca con el agua control, para familiarizarse con los sabores. No trague las muestras. Empiece la evaluación con la primera muestra a su izquierda y continue con la siguiente hacia la derecha, solamente pruebe cada muestra una vez. No pruebe las muestras anteriores nuevamente.

Describa la sensación percibida (dulce, salada, ácida o amarga). Además evalúe la intensidad de la sensación utilizando la siguiente escala:

0 = igual que el agua control
? = diferente del agua pero no puedo describirlo
1 = muy débil
2 = débil
3 = pronunciado (definitivo, claro)
4 = fuerte
5 = muy fuerte

No. de muestra	Sensación	Intensidad

Describa la sensación percibida (dulce, salada, ácida o amarga). Además evalúe la intensidad de la sensación utilizando la siguiente escala:

0 =	igual que el agua control
? =	diferente del agua pero no puedo describirlo
1 =	muy débil
2 =	débil
3 =	pronunciado (definitivo, claro)
4 =	fuerte
5 =	muy fuerte

No. de muestra	**Sensación**	**Intensidad**
____________	_________________	___________
____________	_________________	___________
____________	_________________	___________
____________	_________________	___________
____________	_________________	___________
____________	_________________	___________
____________	_________________	___________
____________	_________________	___________
____________	_________________	___________
____________	_________________	___________

Describa la sensación percibida (dulce, salada, ácida o amarga). Además evalúe la intensidad de la sensación utilizando la siguiente escala:

0 =	igual que el agua control
? =	diferente del agua pero no puedo describirlo
1 =	muy débil
2 =	débil
3 =	pronunciado (definitivo, claro)
4 =	fuerte
5 =	muy fuerte

No. de muestra	Sensación	Intensidad
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________

Describa la sensación percibida (dulce, salada, ácida o amarga). Además evalúe la intensidad de la sensación utilizando la siguiente escala:

0 = igual que el agua control
? = diferente del agua pero no puedo describirlo
1 = muy débil
2 = débil
3 = pronunciado (definitivo, claro)
4 = fuerte
5 = muy fuerte

No. de muestra	Sensación	Intensidad
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________

Describa la sensación percibida (dulce, salada, ácida o amarga). Además evalúe la intensidad de la sensación utilizando la siguiente escala:

0 = igual que el agua control
? = diferente del agua pero no puedo describirlo
1 = muy débil
2 = débil
3 = pronunciado (definitivo, claro)
4 = fuerte
5 = muy fuerte

No. de muestra	Sensación	Intensidad
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________
________	____________	________

TABULACION DE RESULTADOS

Resultados de Prueba de Umbrales de ________________

PANELISTA No.	Concentración en g/100 ml. solución
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	

18	
19	
20	

Comentarios:

LABORATORIO No. 2

PRUEBAS DE DIFERENCIA

INTRODUCCION

Las pruebas de diferencia se utilizan para evaluar si existe o no diferencia entre 2 o más productos, si esta puede ser detectada y si los evaluados son capáces de discriminar.

Los métodos que se utilizan en esta práctica son los más utilizados en la industria de alimentos.

DEFINICIONES

A. Duo-Trio: en la prueba de duo-trio se presentan tres muestras, dos de ellas son idénticas. Dos muestras se codifican con números aleatorios de 3 dígitos, una muestra se codifica como referencia. Una de las muestras codificadas es idéntica a la referencia. El objeto de la prueba es indicar cuál es la muestra diferente de la referencia. Si las muestras son A y B el siguiente orden de muestras es posible:

R(A)AB	R(B)AB
R(A)BA	R(B)BA

B. Triángulo: en la prueba de triángulo, tres muestras son presentadas, dos de ellas son idénticas. Las tres muestras están codificadas con números aleatorios de 3 dígitos. El objeto es seleccionar la muestra que en el grupo de tres es diferente de las dos restantes. Si las muestras son A y B, las siguientes variaciones en ordenamiento son posibles:

AAB	BBA
ABA	BAB
BAA	ABB

C. Prueba de comparación múltiple: en las pruebas de comparación múltiple, una referencia o estandard identificada con "R", es presentada a los penalistas con varias muestras codificadas. La tarea del panelista es comparar cada una de las muestras codificadas contra la referencia para determinar si existen diferencias. Esta prueba puede ser utilizada muy eficientemente para evaluar 4 ó 5 muestras al mismo tiempo. Pequeñas diferencias entre las muestras pueden ser detectadas, además nos da información sobre la dirección y magnitud de la diferencia.

PROCEDIMIENTO

Los participantes serán divididos en 4 grupos de 5 personas cada uno. Cada uno de los 4 grupos será asignado a un área de la cocina y se le indicará que prepare las muestras y prepare la prueba (codificación de vasos, boletas, servir muestras) utilizando uno de los 3 métodos anteriores. Cada grupo deberá preparar muestras para todos los participantes (20).

Area 1 - Duo-trio comparando________________________

____

Area 2 - Prueba de triángulo comparando__________________

____

Area 3 - Comparación múltiple comparando_________________

____

Cada estudiante evaluará 2 sets de muestras para cada prueba. Las boletas a utilizarse se han adjuntado. Ponga un orden aleatorio al servir las muestras.

Anotar los resultados en el pizarrón y luego analizarlos usando:

- Tablas proveídas en Roessler et al., 1979.

- Tablas proveídas por Watts et al., 1992.

- Análisis de varianza.

LABORATORIO No. 2

PRUEBA DE DUO-TRIO

Panelista #:__________________ Fecha:___________

Alimento:____

Instrucciones: evalúe las muestras de izquierda a derecha. La muestra de la izquierda es la referencia. Determine cual de las otras dos muestras (codificadas) es igual a la referencia e indíquelo marcando con una X.

Set 1	Ref.	Código ______	Código______
	[redacted]		

Set 1	Ref.	Código______	Código______
	[redacted]		

Comentarios

____

____

____

____

LABORATORIO No. 2 - PARTE B
ANALISIS DE DATOS

PRUEBA DE DUO-TRIO

Panelista #	Muestra seleccionada set 1		Muestra seleccionada set 2	
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
TOTAL				
PROBABILIDAD				

LABORATORIO No. 2

PRUEBA DE TRIANGULO

Panelista #______________ Fecha_________________

Tipo de Alimento________________________________

Instrucciones: A continuación se presentan tres muestras de ______________. Dos de estas muestras son iguales y una es diferente. Evalúe las muestras en el orden que aparece en la boleta y ponga una marca (X) al lado del código de la muestra que es diferente.

Set 1	Código	La muestra diferente es
	______	________
	______	________
	______	________

Set 2	Código	La muestra diferente es
	______	________
	______	________
	______	________

LABORATORIO No. 2 - PARTE C
ANALISIS DE DATOS

PRUEBA DE TRIANGULO

Panelista #	Muestra seleccionada - set 1		Muestra seleccionada - set 2	
	Diferente	Duplicada	Diferente	Duplicada
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
TOTAL				
PROBABILIDAD				

COMPARACION CONTRA REFERENCIA

Producto______________________ Panelista No.______________________

______________________ Fecha__________

Instrucciones:

A continuación se le presentan 5 muestras de galletas (codificadas) y una muestra control (R). Pruebe la muestra "R", luego pruebe cada una de las muestras codificadas y evalúe el grado de diferencia existente con respecto al control. Por favor tome en cuenta todos los atributos de la muestra (aroma, color, sabor, textura, etc.).

Código	Grado de Diferencia contra Referencia
____________	____________
____________	____________
____________	____________
____________	____________
____________	____________

Escala de Grado de Diferencia

0 = no hay diferencia
1 = diferencia muy pequeña
2 = diferencia pequeña
3 = diferencia moderada
4 = diferencia grande
5 = diferencia extrema

LABORATORIO No. 2 - PARTE D
ANALISIS DE DATOS

PRUEBA DE COMPARACION MULTIPLE

Panelista #					TOTAL
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
TOTAL					
PROBABILIDAD					

[Reprinted from Food Technology, 1963, Vol. 17, No. 9, Pages 45–48, 51].

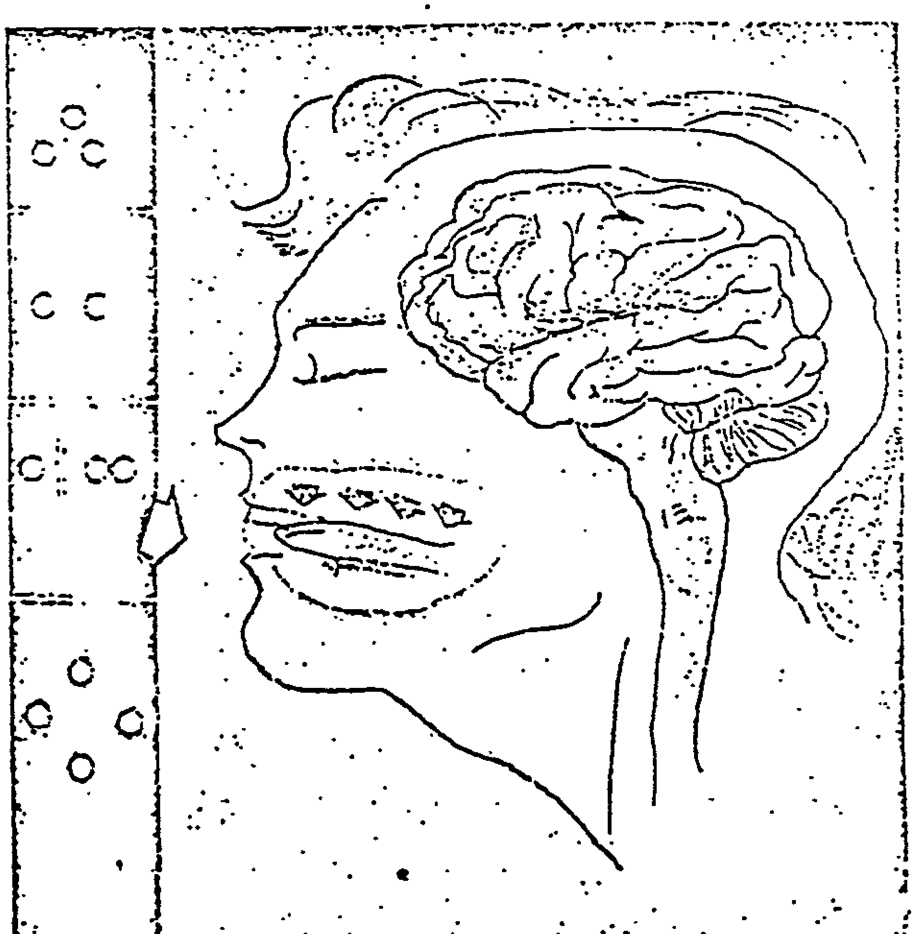
SENSORY TESTING OF DIFFERENCES IN TASTE

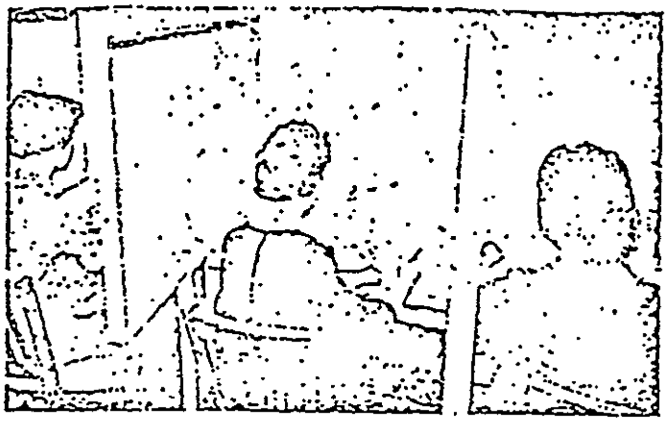
I. METHODS

Elsie H. Dawson, Jennie L. Brogdon, and Sandra McManus

■ Flavor-difference testing has been defined as "a comparison or test of quality variation without indication of preference" (Kramer, 1959). In contrast to consumer-preference and flavor-acceptance tests, difference tests are concerned only with whether a detectable difference exists between two or more treatments. Flavor-difference tests include triangle, paired-comparison, duo-trio, multiple-comparison, descriptive-term, profile, and dilution tests. These methods of evaluation are used in quality control and in research and development work on foods and beverages such as, for example, the effect of formulation or processing change on a product, the effect of packaging materials on

• This is one of two articles on sensory testing for flavor differences. Flavor-difference testing is widely used for quality control and for research and development work on foods and beverages. It is necessary, therefore, that the most appropriate method and procedures be followed in order to obtain reliable results. Knowledge accumulated and progress in the last decade have placed sensory-difference testing on a more scientific basis. Information gained from collected data and experimentation will further develop and advance techniques of difference testing that will be beneficial to the food industry in the future.





flavor, and the effect of pesticides on the flavor of fruits and vegetables.

The detection of differences among treatments is usually made by a small number of well chosen panel members, who serve essentially as a laboratory instrument. As with any laboratory instrument, the precision of the results depends on the precision of the tool and the conditions under which that tool is used. The factors for determining precision to be considered here are: a) control of environmental conditions at testing, b) test method, and c) selection of panel members.

Table 1. Morning and afternoon paired difference tests on bouillon prepared with beef bouillon cubes and four kinds of water.

Pair[a]	Taste difference			
	A.M.		P.M.	
	Yes	No	Yes	No
AB	13**	0	15**	3
BA	14**	0	14**	3
AC	11*	2	16**	2
CA	13**	1	15**	2
AD	10*	3	12	5
DA	11*	2	16**	2
BC	11*	2	13*	4
CB	12**	1	12	6
BD	13**	1	14*	4
DB	12**	1	16**	1
CD	10	6	14*	5
DC	9	3	14*	2
Sum	139**	22	171**	39
Percent	86	14	81	19

* Significantly greater number of yes answers than expected by chance at the 5% level.

** Significantly greater number of yes answers than expected by chance at the 1% level.

[a] The water solutions used in preparing the beef bouillon were: A—water plus chlorine; B—water plus iron; C—water; and D—water plus calcium.

Control of Environmental Conditions

Although the panel is operating as a laboratory instrument, the ratings depend on human judgments, which may be influenced by physical conditions in the person or in the environment. Many of these conditions might also affect flavor acceptance and preference tests.

The judging room should be free from distractions (Mitchell, 1957a), odor-free, air-conditioned, or with other means of proper ventilation and temperature control. If the room and furnishings are off-white or light or neutral gray, the panelists will not be distracted by color. Lighting should be uniform, and adjustable if color needs to be masked. Separate booths prevent the exchange of expressions or impressions between panel members, ensuring individual responses. If care is not exercised, round-table discussion can create qualitative flavor differences where none exist (Foster *et al.*, 1955).

Table 2. *Incorrect* identifications in the first and second half of triangle and paired comparison test series.

Method	Caffeine	NaCl	Sucrose	Tartaric acid	Total
Triangles[a]					
First 3	51	51	50	46	198*
Second 3	32	40	39	28	139
Pairs[b]					
First 5	43	24	23	32*	122**
Second 5	26	11	23	15	75

* Significantly greater number of incorrect identifications than expected by chance at the 5% level.

** Significantly greater number of incorrect identifications than expected by chance at the 1% level.

[a] Each of 6 panel members tasted 3 levels of concentration of each basic taste in distilled water. Six triangles were tasted for either 2 or 3 replications, depending on number of correct answers.

[b] Each of 6 panel members tasted 3 levels of concentration for each basic taste in distilled water. Ten pairs were tasted for either 1 or 2 replications, depending on number of correct answers.

Several investigators have reported contradictory results for other factors influencing panel precision, such as time of day, number of samples per session, and time intervals between samples (Janowitz and Grossman, 1949; Goetzl *et al.*, 1950; Mitchell, 1957b). Our laboratory found no significant differences in results between morning (11 A.M.) and afternoon (3 P.M.) sessions in a paired comparison study of bouillon reconstituted with different kinds of water (Table 1).

The optimum number of samples that can be tested at one session without taste fatigue depends on the product (Brandt and Hutchison, 1956; Mitchell, 1956b; Pfaffmann *et al.*, 1954; Ward and Boggs, 1951). More samples can be tested at one session if the product is bland than if less bland (Kramer *et al.*, 1961; Laue *et al.*, 1954).

Our laboratory studied the effect of taste fatigue on basic taste-dilution tests using triangle and paired comparison tests. The panel made significantly more *incorrect* identifications in the first half, indicating better taste acuity in the second half (Table 2). The fact that the panel did not make more incorrect identifications in the second half indicates no taste fatigue.

Similar results were found in our laboratory in a study using instant bouillon reconstituted with different kinds of water. The effect on panel evaluation of a given pair of samples served as the first pair in the session was studied. In seven pairs of samples tested 20 times each, the results for the first and last pairs were as follows:

Can you detect a difference in taste?	Yes	No
First pair	13	7
Seventh pair	19**	1

Note above that differences detected in flavor were significant at the 1% (**) level when the samples were served last, but not when served first.

Panel members tend to use all available information in making their judgments. Therefore, samples should be prepared and served as uniformly as possible in all aspects not related to flavor. Size of sample, temperature, texture, appearance, and color must be controlled. The actual identity of each sample must be concealed by coding. If possible, the product should be tasted by the panelist in the condition in which the food is normally consumed.

Information about the variable to be studied can be of great help in increasing the sensitivity of discrimination tests (Dyer and Abrams, 1953). However, panel members will be influenced in their decisions by their knowledge of the stimulus variable and by the information given them, regardless of whether it contradicts perceptual experience in the test situation (Raffensperger and Pilgrim, 1956). Therefore, instructions should be clear, concise, and appropriate to the experiment.

This brief summary of factors is evidence that no one statement or recommendation can cover these areas. Experience with the specific product must guide decisions on the control of many variables. Further information for clarifying standardized conditions can be found in a number of sources, e.g., Dawson and Harris (1951), Kramer and Twigg (1962), and Arthur D. Little, Inc. (1958).

Test Methods

The methods used for evaluating flavor differences include: triangle, paired-comparison, duo-trio, multiple-comparison, scoring, ranking, single-stimulus, descriptive-term, dilution, and flavor-profile tests. These methods have been reviewed and discussed by Boggs and Hanson (1949), Caul (1957), Dawson and Harris (1951), Kramer and Twigg (1962), Lockhart (1951), Peryam (1958), Schwartz and Foster (1957), Terry *et al.* (1952), and others.

The objectives of the difference test must be clearly defined before a test method can be selected. The investigator must determine: a) whether it is sufficient to determine only if a difference exists; b) whether the direction, extent, and importance of the difference must be known; or c) whether a complete analysis and description of the flavor is needed. The efficiency and simplicity of the method and the appropriateness to the problem must also be considered (Peryam, 1958).

The efficiency of a method considers both the statistical aspects and the amount of work required to achieve a given degree of discrimination. The method selected should give maximum information with the least cost and effort. A means for determining efficiency-of-variables procedures, such as multiple-comparison method, has been developed in terms of the number of tastings required for demonstrating statistical significance at a chosen probability level (Wiley *et al.*, 1957).

Other factors to be considered include the simplicity of the test (Byer and Abrams, 1953), the consistency of the panel, the amount of testing to be done, whether testing is to be continuous or intermittent, and whether materials are constant or varied

Two or Three Sample Tests

Triangle and duo-trio tests are used when two treatments are to be evaluated or when it is advantageous to refer continuously to a reference sample used as an unknown. The results of these tests indicate whether a difference exists between two treatments with no indication of the direction and extent of the differences, although triangle and duo-trio methods have been modified to obtain a quality judgment. Because the objectives of the test are changed, if intensity of difference or a quality judgment is included, a different method should be used (Kramer *et al.*, 1961; Peryam, 1958).

The results of triangle and duo-trio method may be influenced by the order in which samples are served. In duo-trio tests, discrimination is better if the stronger flavor or unusual flavor is presented as the odd sample (Brandt and Hutchison, 1956; Mitchell, 1956a). In triangle tests, detecting differences can be improved by holding the odd sample constant (Pfaffmann *et al.*, 1954) and by keeping samples 1 and 3 as duplicates (Harries, 1956). Panel members have shown a bias toward selecting the middle sample of the triangle test as the odd sample (Berg *et al.*, 1955; Harrison and Elder, 1950).

Paired-comparison tests are used to evaluate two samples at a time. If more than two treatments are being tested, each treatment can be compared with every other in the series (Terry *et al.*, 1952).

When the panel member is required to indicate only whether the samples in a pair are the same or different, the information obtained is similar to that obtained in triangle and duo-trio tests. In most paired-comparison tests the dimension of the flavor difference is compared on the basis of a specified criterion such as "which sample is sweeter?" Thus, the extent of the difference is not determined. Paired-comparison tests of this type can be used only when the criterion can be defined by the investigator and understood by each panel member. Paired tests are subject to bias as to position, temperature, and other factors (Gridgeman, 1958).

Triangle, duo-trio, and paired tests are applied for quality control and for product development and improvement when products approach complete homogeneity within lots. These methods are also used to determine the accuracy and reliability of panel members. Advantages are that small differences between samples can be determined, and direct comparison of the samples re-

Test Sensitivity. A number of comparisons have been made of the sensitivity of the triangle, paired, and duo-trio tests.

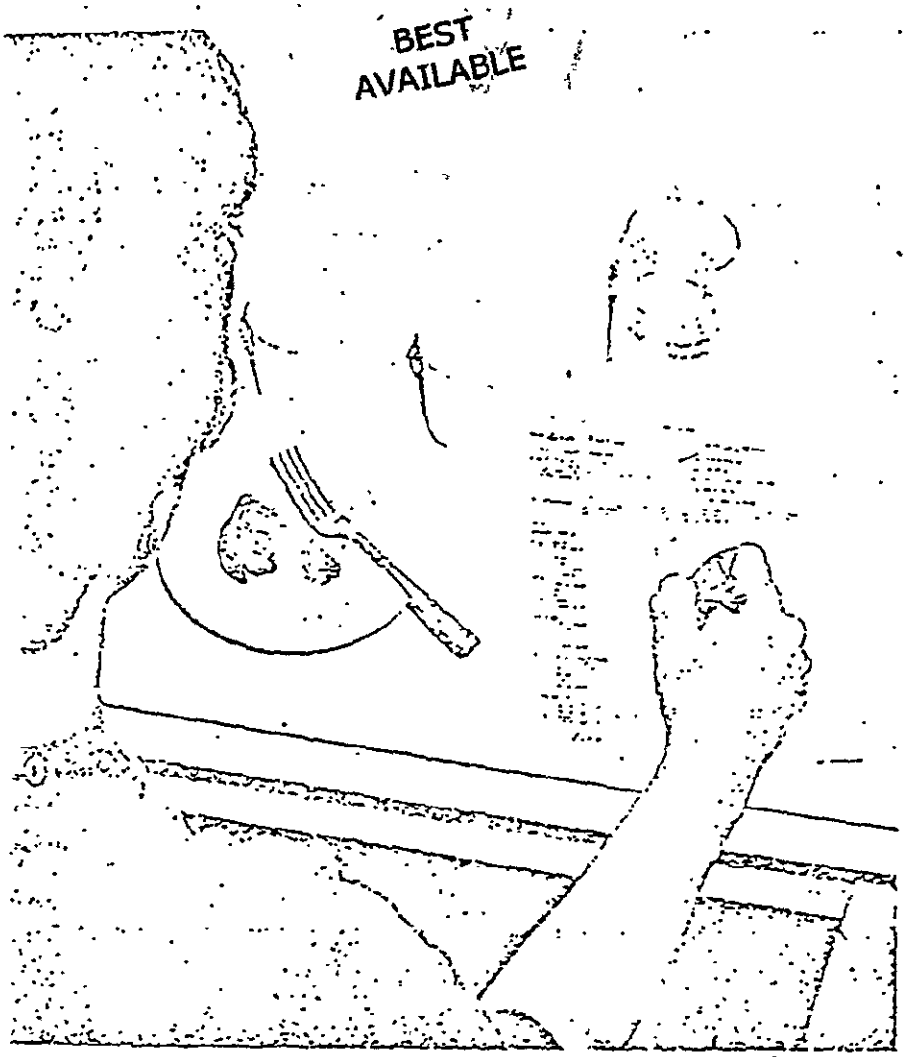
The paired test was reported superior to triangle tests when quinine and dextrose aqueous solutions were tasted (Byer and Abrams, 1953). However, Harrison and Elder (1950) concluded that the triangle test is statistically more efficient than the paired comparison. Dawson and Dochterman (1951) found no difference in precision between the triangle test and the paired-comparison test on chocolate fudge as a basis for selecting reliable panel members. It was felt that more confidence would be placed in using the triangle test, however, since judges who could not identify duplicate samples could be eliminated. Pfaffmann *et al.* (1954) found that, when the flavor dimension of orange drink was specified, paired tests were better than the triangle test, although paired tests were not superior when flavor dimension was not specified.

Paired comparison tests and triangle tests were reported to be equally sensitive, and appreciably superior to duo-trio tests, with aqueous solutions of primary flavors, tomato juice, and ground beef (Gridgeman, 1955). The paired test was usually more economical than the triangle and duo-trio tests on an aliquot-for-aliquot basis; the duo-trio test was the least economical.

Triangle and paired tests were compared in our laboratory in experiments with ground rib patties. The meat was from cattle treated with two levels of an insecticide. Neither method was superior to the other in terms of panel sensitivity. The paired tests were superior, however, from the standpoint of panel time and the amount of laboratory preparation required.

Multiple-Sample Tests. A multiple-comparison test could be selected when a number of treatments are to be compared to a reference sample. Multiple-comparison tests have been developed for the evaluation of canned foods (Mahoney *et al.*, 1957) and of foods exposed to pesticides (Wiley *et al.*, 1957). The direction, extent, and importance of the difference can be determined.

The advantages of the multiple-comparison method over the triangle method are: a) detects smaller differences between treated and untreated samples; b) gives additional information about the direction and importance of the differences; c) requires less time and fewer samples; d) is more efficient when panels have not been specially



selected or trained; and c) is not influenced by small differences in color and texture (Hogue and Briant, 1957; Kramer and Ditman, 1956; Mahoney *et al.*, 1957).

Scoring Methods. Numerical scoring of samples is applicable when the dimension of differences can be predetermined and arranged on a numerically graduated scale. The degree, direction, and extent of differences can be determined. A number of samples can be tested at one session, depending on the product.

The scale may be descriptive in nature, or be anchored on specimens of the product for reference and direct comparison with the experimental sample. The number of gradations on the scale will depend on the number of intervals a panelist can distinguish. All panel members require training and must understand the criterion in the same way.

Scores for products may be affected by order of presentation, for example, when samples follow a non-standard or a standard sample (Carlin *et al.*, 1956). Harrison and Elder (1950) state that the standard must be presented twice—once as an "open" standard and again as a blind sample. The blind sample will receive a score slightly different from that of the open sample. Harries (1956) found that scores were affected by physical presentation on the table when samples are presented in a straight line.

Ranking may be used to specify the dimension of the difference as to the intensity of a characteristic, and when actual values are not needed or are difficult to provide. The panelist must be thoroughly familiar with the particular characteristic under study.

Ranking is a fast method of discriminating multiple samples. Fixed scales can be used with a control or reference sample.

Scoring and ranking methods appeared equally efficient in detecting differences in a number of foods; however, the scoring method indicated degrees of differences that were not shown by ranking (Dawson and Dochterman, 1951; Packs, 1954). Kramer *et al.* (1961) could not indicate a definite advantage favoring the use of either scoring or ranking procedures.

The flavor-profile test is the best known method for obtaining a qualitative as well as quantitative description of the individual characteristics of the total flavor complex of the product. Descriptive words for each individual flavor note and number are obtained that indicate the relative strength of each note on some suitable scale. With this method it is possible to indicate degrees of differences between two samples, the degree of blending, degrees of similarities, and the over-all impression of the product.

Considerable knowledge of flavor is required for interpretation of flavor-profile results that cannot be analyzed statistically but can be plotted or graphed. A written record of a flavor can be obtained and used as a reference.

The flavor-profile method requires education in flavor and odor sensations, keen interest, intelligence, and honesty on the part of the panelist. Flavor-profile techniques have been reviewed in detail by Sjöström *et al.* (1957) and Caul (1957).

Dilution tests are used to summarize the flavor of the product in terms of percent dilution or as a ratio that reflects the actual amount of odor or flavor detected (Hanson *et al.*, 1954; Kramlich and Pearson, 1958). Dilution indexes have been reported for several Polish foods (Tilgner, 1962a). For example, the flavor of raspberry syrup was of standard quality at a dilution of 1:150, and of prime quality at 1:200. The index permits establishment of anchor points for descriptive terms used in rating odor and flavor of foods and beverages. Dilution tests have been studied also as a means for selecting and maintaining a reference standard for scoring or difference testing methods (Tilgner, 1962b).

The method requires suitable standards for comparison and for dilution of the test material, and is limited to foods that can be made homogeneous without affecting flavor.

Dilution tests have been extended to provide a flavor profile. The panel members record the profile for samples at dilutions between the identification threshold and the natural undiluted product. The results indicate what, how much, and at what level each sensory component exists in the product. A dilution flavor profile for an instant coffee showed at a dilution level of 0.05 to 0.10% a sweet note that disappeared and was superseded by a bitter note at 0.1 to 1%. A sour note emerged at 0.4% and remained up to 0.6%; a burnt note was sensed at dilutions of 0.2 to 1%. At normal strength (1%), only the bitter and burnt notes were in evidence (Tilgner, 1962a).

Statistical Techniques

Statistical techniques for difference testing have not been considered here. References most frequently used in our food-quality laboratory are: Bradley, 1953; Cochran and Cox, 1957; Duncan, 1955; Ezekiel and Fox, 1959; Federer, 1955; Gridgeman, 1955; Hopkins and Gridgeman, 1955; Kramer, 1955, 1956, 1957; Radkins, 1958; and Terry *et al.*, 1952.

REFERENCES

Arthur D. Little, Inc. 1958. "Flavor Research and Food Acceptance." Reinhold Publishing Corp., New York.

Berg, H. W., F. Filipello, E. Hinreiner, and A. D. Webb. 1955. Evaluation of thresholds and minimum difference concentrations for various constituents of wines. I. Water solutions of pure substances. *Food Technol.* 9, 23.

Boggs, M. M., and H. L. Hanson. 1949. Analysis of foods by sensory difference tests. *Advances in Food Research* 2, 219.

Bradley, R. A. 1953. Some statistical methods in taste testing and quality evaluation. *Biometrics* 9, 22.

Brandt, D. A., and E. P. Hutchison. 1956. Retention of taste sensitivity. *Food Technol.* 10, 419.

Byer, A. J., and D. Abrams. 1953. A comparison of the triangular and two-sample taste-test methods. *Food Technol.* 7, 185.

Carlin, A. F., O. Kempthorne, and J. Gordon. 1956. Some aspects of numerical scoring in subjective evaluation of foods. *Food Research* 21, 273.

Caul, J. F. 1957. The profile method of flavor analysis. *Advances in Food Research* 7, 1.

Cochran, W. G., and G. M. Cox. 1957. "Experimental Designs." 2d ed. John Wiley and Sons, New York.

Dawson, E. H., and E. F. Dochterman. 1951. A comparison of sensory methods of measuring differences in food qualities. *Food Technol.* 5, 79.

Dawson, E. H., and B. L. Harris. 1951. Sensory methods for measuring differences in food quality. U. S. Dept. Agr. Infor. Bull. 34.

Duncan, D. B. 1955. Multiple range tests for correlated and heteroscedastic means. *Biometrics* 13, 164.

Ezekiel, M., and K. A. Fox. 1959. "Methods of Correlation and Regression Analysis." John Wiley and Sons, New York.

Federer, W. T. 1955. "Experimental Design, Theory and Application." Macmillan, New York.

Foster, D., C. Pratt, and N. Schwartz. 1955. Variation in flavor judgments in a group situation. *Food Research* 20, 539.

Goetzl, F. R., A. J. Ahokas, and J. G. Payne. 1950. Occurrence in normal individuals of diurnal variation in acuity of the sense of taste for sucrose. *J. Appl. Physiol.* 2, 619.

Gridgeman, N. T. 1955. Taste comparisons: Two samples or three? *Food Technol.* 9, 148.

Gridgeman, N. T. 1958. Psychophysical bias in taste testing by pair comparison, with special reference to position and temperature. *Food Research* 23, 217.

Hanson, H. L., L. Kline, and H. Lineweaver. 1951. A dilution method for the determination of relative flavor stability of egg solids. Food acceptance testing methodology. Quartermaster Food and Container Inst., Chicago, p. 20.

Harries, J. M. 1956. Positional bias in sensory assessments. *Food Technol.* 10, 86.

Harrison, S., and L. W. Elder. 1950. Some applications of statistics to laboratory taste testing. *Food Technol.* 4, 434.

Hogue, D. V., and A. M. Briant. 1957. Determining flavor differences in crops treated with pesticides. I. A comparison of a triangle and a multiple comparison method. *Food Research* 22, 351.

Hopkins, J. W., and N. T. Gridgeman. 1955. Comparative sensitivity of pair and triad flavor intensity difference tests. *Biometrics* 11, 64.

Janowitz, H. D., and M. I. Grossman. 1949. Gusto-olfactory thresholds in relation to appetite and hunger sensations. *J. Appl. Physiol.* 2, 217.

Kramer, A. 1959. Glossary of some terms used in the sensory (panel) evaluations of foods and beverages. *Food Technol.* 13, 733.

Kramer, A., and L. P. Ditman. 1956. A simplified variables taste panel method for detecting flavor changes in vegetables treated with pesticides. *Food Technol.* 10, 155.

Kramer, A., and B. A. Twigg. 1962. "Fundamentals of Quality Control for the Food Industry." p. 105. Avi Publishing Co., Westport, Conn.

Kramer, A., E. F. Murphy, A. M. Briant, M. Wang, and M. E. Kirkpatrick. 1961. Studies in taste panel methodology. *J. Agr. Food Chem.* 9, 224.

Kramer, C. Y. 1955. A method of choosing judges for a sensory experiment. *Food Research* 20, 492.

Kramer, C. Y. 1956. Additional tables for a method of choosing judges for a sensory experiment. *Food Research* 21, 598.

Kramer, C. Y. 1957. Extension of multiple range tests to group correlated adjusted means. *Biometrics* 13, 13.

Kramlich, W. E., and A. M. Pearson. 1958. Some preliminary studies on meat flavor. *Food Research* 23, 567.

Laue, E. A., N. H. Ishler, and G. A. Bullman. 1954. Reliability of taste testing and consumer testing methods. I. Fatigue in taste testing. *Food Technol.* 8, 387.

Lockhart, E. E. 1951. Binomial systems and organoleptic analysis. *Food Technol.* 5, 428.

Mahoney, C. H., H. L. Stier, and E. A. Crosby. 1957. Evaluating flavor differences in canned foods. I. Genesis of the simplified procedure for making flavor difference tests. *Food Technol.* 11, 29.

Mitchell, J. W. 1956a. The effect of assignment of testing materials to the paired and odd position in the duo-trio taste difference test. *Food Technol.* 10, 169.

Mitchell, J. W. 1956b. Duration of sensitivity in trio taste testing. *Food Technol.* 10, 201.

Mitchell, J. W. 1957a. Problems in taste difference testing. I. Test environment. *Food Technol.* 11, 476.

Mitchell, J. W. 1957b. Problems in taste difference testing. II. Subject variability due to time of the day and day of the week. *Food Technol.* 11, 477.

Parks, A. B. 1954. Ranking versus scoring in palatability tests using small trained panels. Food acceptance testing methodology. Quartermaster Food and Container Inst., Chicago, p. 40.

Peryam, D. R. 1958. Sensory difference tests. *Food Technol.* 12, 231.

Pfaffmann, C., H. Schlosberg, and J. Cornsweet. 1954. Variables affecting difference tests. Food acceptance testing methodology. Quartermaster Food and Container Inst., Chicago, p. 4.

Radkins, A. P. 1958. Sequential analysis in organoleptic research: triangle, paired, duo-trio tests. *Food Research* 23, 225.

Raffensperger, E. L., and F. J. Pilgrim. 1956. Knowledge of the stimulus variable as an aid in discrimination tests. *Food Technol.* 10, 254.

Schwartz N., and D. Foster. 1957. Methods for rating quality and intensity of the psychological properties of food. *Food Technol.* 11, 15.

Sjöström, L. B., S. E. Cairncross, and J. F. Caul. 1957. Methodology of the flavor profile. *Food Technol.* 11, 20.

Terry, M. E., R. A. Bradley, and L. L. Davis. 1952. New designs and techniques for organoleptic testing. *Food Technol.* 6, 250.

Tilgner, D. J. 1962a. Dilution tests for odor and flavor analysis. *Food Technol.* 16 (2), 26.

Tilgner, D. J. 1962b. Anchored sensory evaluation tests—a status report. *Food Technol.* 16 (3), 47.

Ward, A. C., and M. M. Boggs. 1951. Comparison of scoring results for two and four samples of corn per taste session. *Food Technol.* 5, 219.

Wiley, R. C., A. M. Briant, I. S. Fagerson, E. F. Murphy, and J. H. Sabry. 1957. The Northeast regional approach to collaborative panel testing. *Food Technol.* 11, 43.

THE AUTHORS are with the Human Nutrition Research Division, Agricultural Research Service, U. S. Department of Agriculture, Beltsville, Md.

Presented at the 22nd Annual Meeting of the Institute of Food Technologists, Miami Beach, Florida, June 10-14, 1962.

SENSORY TESTING OF DIFFERENCES IN TASTE

II. SELECTION OF PANEL MEMBERS

Elsie H. Dawson, Jennie L. Brogdon, and Sandra McManus

• THIS IS ONE OF TWO *articles on sensory testing for flavor differences. Flavor-difference testing is widely used for quality control and for research and development work on foods and beverages. It is necessary, therefore, that the most appropriate method and procedures be followed in order to obtain reliable results. Knowledge accumulated and progress in the last decade have placed sensory-difference testing on a more scientific basis. Information gained from collected data and experimentation will further develop and advance techniques of difference testing that will be beneficial to the food industry in the future. Part I, Methods, appeared in September* Food Technology.

■ IT IS IMPORTANT to select panelists who have a superior ability to detect differences. The candidates considered for a panel should exhibit intelligence, comprehension, concentration, sustained interest, and motivation toward sensory testing. They must, of course, have interest in the problem. Panelists should be able to detect fine differences in specific attributes of foods, and to give reproducible judgments in testing the same samples at different times. A panel of 3 to 10 is usually adequate, the size depending on the variability associated with the experiment and on the magnitude of the difference between samples to be detected.

The sensitivity and reliability of an individual varies from time to time (Boggs *et al.*, 1960). Variability is even greater between individuals—in taste thresholds, in degrees of difference they can distinguish, and in ability to give reproducible judgments. Health, age, sex, smoking, and emotional factors have been considered possible causes for variability in taste acuity among individuals. Panel candidates for a specific product may have to be excluded on the basis of some of these criteria.

Important Factors

Individuals should be healthy. Minor infections of the nose and throat may affect flavor perception, although previous illness from hay fever or sinus may not affect taste acuity (Kotschevar, 1956). Nutritional health of an individual may also affect sensitivity. Abnormal taste responses have been found in vitamin-A-deficient rats (Bernard *et al.*, 1961; Seifter *et al.*, 1960). Human subjects deficient in salt showed increased taste sensitivity for salt although sensitivity to sweet, sour, and bitter showed no such change (Yensen, 1958). Reduction in blood sugar level by intraperitoneal injection of insulin was not associated with a change in taste sensitivity, although preference for sugar was enhanced (Pfaffmann and Hagstrom, 1955).

Age of the individuals should be considered, although reports vary greatly as to the effect of age on panel members' acuity (Bates, 1960; Kotschevar, 1956). However, all sensory threshold sensitivities seem to decrease exponentially with age, when the intensities of the threshold stimuli are expressed in terms of psychological magnitude scales (Hinchcliffe, 1962).

There may be sex-associated differences in ability to perceive flavor. Pangborn (1959) found much lower thresholds in females than in males of college age. Kotschevar (1956) found little or no effect due to sex when panel members evaluated meat. For difference tests, which do not require a representative sample of the public, panel members should be selected for sensitivity, and sex differences should be ignored.

Physiological factors (not reviewed here) are discussed in a recent publication edited by Kare and Halpern (1961).

Pretastes

Ability to discriminate may also be affected by substances tasted prior to flavor evaluation (Fabian and Blum, 1943; Pilgrim *et al.*, 1955). Dallenbach and Dallenbach (1943) noted that the effect on taste acuity of tasting one food before another was reported in the early 1800's. Berg *et al.* (1955) reported that alcohol enhances the sweetness of a sucrose solution. Sucrose significantly decreased the perceived intensity of caffeine, but there was no evidence that sucrose affected the perception of even strong solutions of salt (Kamenetzky and Pilgrim, 1958). Kamen and Pilgrim (1959) reported the effect of suprathreshold primary-taste stimuli on subsequently tasted primary-taste solutions. Mixtures of stimuli have also been studied as to their effect on taste perception (Hinreiner *et al.*, 1955a; Pangborn, 1960, 1961; Valdés *et al.*, 1956).

In a recent experiment, we investigated the effect of several pretastes on the ability to distinguish differences in the taste of potatoes with and without added sodium chloride. Triangle tests were employed in a randomized block design (Cochran and Cox, 1957). The pretasted substances were tomato juice, milk, apple juice, raw apple, and coffee served at 10°C; bouillon and coffee served at 50°C; and distilled water at room temperature (21°C). Two minutes elapsed between the tasting of the pretasted substances and the evaluation of the potatoes.

The panel differentiated between the samples with and without added sodium chloride more often (significant at the 1% level) when the pretaste was distilled water, hot bouillon, and hot coffee (Table 1) and (at the 5% level) tomato juice. The panel did not differentiate between the samples a significant number of times after the other pretasted substances.

Table 1. Effect of pretaste on the taste acuity for potatoes with and without sodium chloride.

Pretasted substance	Temperature (°C)	Correct selections of different sample[a] (%)
Distilled water	21	[illegible]**
Bouillon	50	[illegible]**
Coffee	50	59**
Tomato juice	10	53*
Apple juice	10	47
Raw apple	10	47
Milk	10	44
Coffee	10	38

* Significantly greater percentage than expected by chance at the 5% level.

** Significantly greater percentage than expected by chance at the 1% level.

[a] The number of correct selections of the different sample in 32 triangle tests. The different sample could be the one with or without added sodium chloride.

The temperature of the pretaste appeared to have an effect on detecting flavor differences. All of the pretastes served at room temperature and at 50°C yielded significant differences at the 1% level, but there were no significant differences at this level when the pretastes were served at 10°C (refrigerator-chilled).

Variability in panel performance can be decreased by selecting sensitive individuals and training and checking their performances.

Taste Thresholds

Threshold tests with solutions of pure chemicals have been used for many years to select panel members and to detect differences in the sensitivity of individuals to sweet, salt, sour, and bitter solutions. In our laboratory, we determined thresholds of prospective panel members for the four basic or primary tastes as a basis for a study on methodology of difference tests.

Triangle dilution tests were made by two procedures to compare thresholds when the taste is known to the panelist (*Test A*) and when it is not (Test B). Chemically pure solutions of sucrose, caffeine, and sodium chloride were used in various levels of concentration. In test A each triangle consisted of two samples of distilled water and one of the basic taste solution. In test B the like samples in each triangle were either the basic taste being tested or distilled water. The person's threshold in each test was determined by the lowest molar dilution identified a significant number of times at the 1% level.

The *difference* threshold was established in test A, when the panel was informed of the taste being tested and asked only to select the sample in the triangle containing the basic taste. The *taste* threshold was established in test B, when each panelist was required to name the taste being tested and to identify the sample or samples containing the basic taste.

The *difference* threshold is the lowest molar concentration distinguishable from distilled water without identification of the taste. Some other terms used for the difference threshold are sensitivity, subthreshold taste, subliminal, and threshold of sensation (Fabian and Blum, 1943; Richter and Campbell, 1940; Richter and MacLean, 1939). The *taste* threshold is the minimum concentration in which the judge can recognize and identify the taste. Another term used for taste threshold is gustatory threshold (Janowitz and Grossman, 1949; Goetzl *et al.*, 1950b).

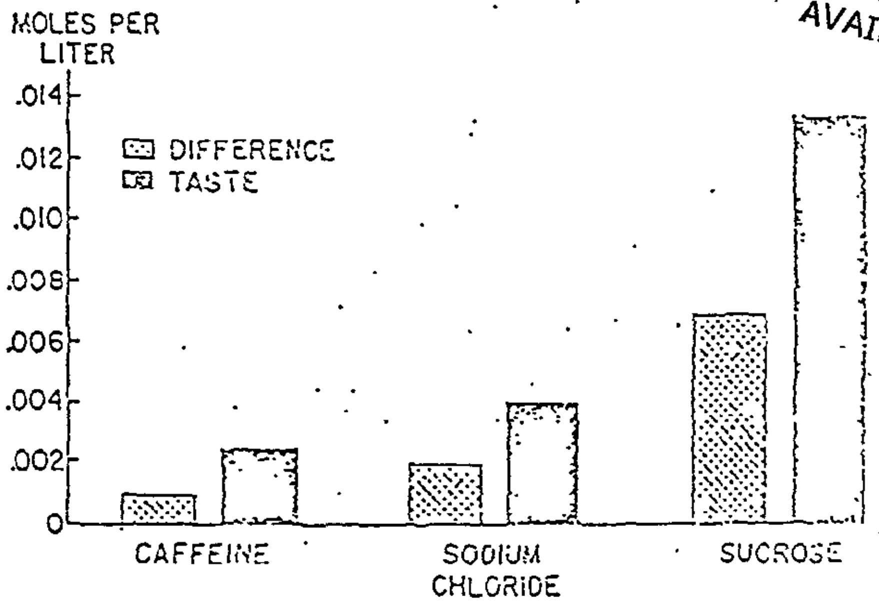


Fig. 1. Comparison of panel means for "difference" and "taste" thresholds.

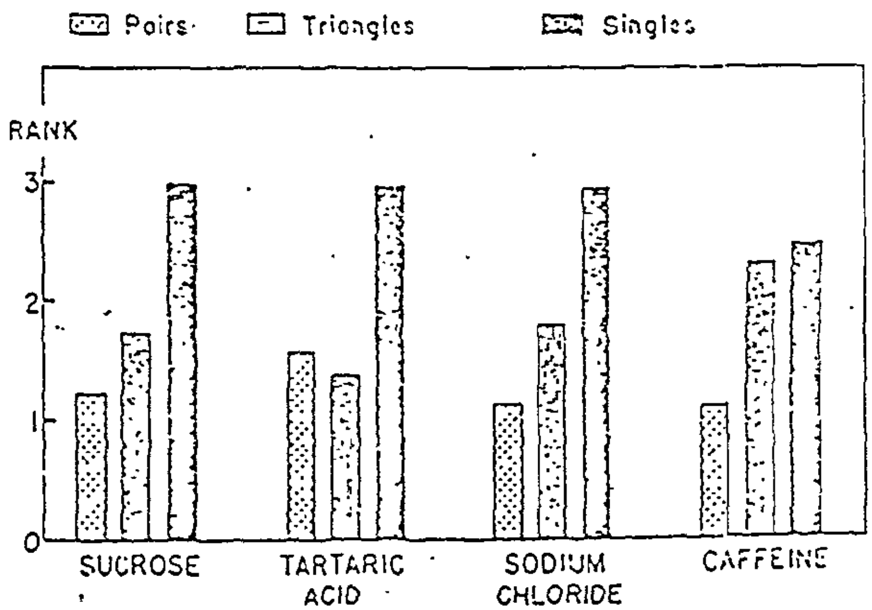


Fig. 2. Comparison of methods for determining taste thresholds by ranking of panel means. Rank of 1 is the lowest mean threshold for six panel members; rank of 3, the highest.

Fig. 1 shows the panel thresholds for sucrose, caffeine, and sodium chloride. The *taste* thresholds for the panel (Test B) were considerably higher than the *difference* thresholds for the panel (Test A). Similar results have been reported in the literature (Pangborn, 1959; Richter and Campbell, 1940; Richter and MacLean, 1939; Fabian and Blum, 1943).

Since taste thresholds are considerably higher than difference thresholds, it is important in establishing taste thresholds that the panelist identify the taste being tested.

Triangle, paired, and single-stimulus dilution tests were compared to determine the most efficient method of establishing taste thresholds for the four basic tastes.

Panel members were told that they were being tested for taste sensitivity to the four basic tastes, but were not told the order in which the tastes were offered. The six panelists were required to name the taste being tested and identify samples containing the taste. Randomized block designs were used for all methods.

In the triangle test, 6 triangles were served at each of 3 sessions and for 3 levels of concentration. The "like" samples in each triangle were either the basic taste being tested or distilled water. In the paired comparison test, 10 pairs of samples for each of 2 sessions were served for 3 levels of concentration. In both tests, panelists were required to identify the basic taste being tested a significant number of times at the 1% level.

In the single-stimulus tests, the panelists tasted 8 solutions of one primary taste in order of increasing molar concentration and indicated in which beaker they were positive of the identity of the taste. This test was repeated 3 times for each basic taste. The threshold concentration was the average of the molar concentrations for the 3 replications.

When results of the methods were compared, lowest thresholds were usually demonstrated by the pairs, next lowest by triangles, and the highest thresholds by the single-stimulus procedure (Fig. 2). In reporting taste thresholds, the methods used must be described accurately since they have a bearing on the results. When only two samples were compared, the taste appeared to be identified by contrast more easily than in the triangle method, where one out of three samples was different. In the single-stimulus test, a higher concentration of the taste was needed for identification.

The paired-comparison and triangle tests required 2 to 3 times as many sessions as the single-stimulus method, and 2½ to 3 times as long per session.

The concentrations, in moles per liter, for the eight solutions of each substance were:

Concentration number	Sucrose	Tartaric acid	Sodium chloride	Caffeine
1	0.0064	0.00004	0.0021	0.0004
2	.0128	.00008	.0016	.0006
3	.0186	.00016	.0064	.0008
4	.0256	.00032	.0128	.0016
5	.0410	.00064	.0166	.0032
6	.0555	.00128	.0256	.0051
7	.0819	.00256	.0341	.0085
8	.1024	.00512	.0512	.0125

Also, the paired and triangle tests required 5 to 6½ times as many samples for statistical significance as the single-stimulus method. The time required for laboratory preparation was also much greater for the paired and triangle methods.

Individuals exhibit very different reactions to taste stimuli as shown by threshold values for the basic tastes. Our findings agree with those of Blakeslee (1932) in that no one of the panel was the poorest (or best) taster in respect to all substances tested. Some were relatively acute (or poor) tasters for all substances tested.

Training

In a third study the taste thresholds of trained and untrained panelists were compared by single-stimulus dilution tests. The tests were given to six experienced panelists whose taste thresholds had previously been established by triangle and paired dilution tests. The single-stimulus tests were repeated for a panel of 12 staff members whose basic taste thresholds were unknown and who had not taken the other tests.

The trained panel had significantly lower thresholds than the untrained panel for sodium chloride, sucrose, and tartaric acid (Fig. 3). However, the panels did not differ significantly in their thresholds for caffeine. Some of the panel members reported a carry-over in taste in the series of caffeine. Neilson (1957) reported that, in tasting caffeine, there is a 30-second delay before the peak of bitterness is reached and that the bitter taste lasts over 2 minutes. Pangborn (1959) reported a decided decrease in all thresholds with practice, but there is a wide variation between individuals in degree of such improvement.

Training of panel members increases sensory acuity; makes certain that the panel has uniform understanding of the properties to be evaluated, the system of evaluation, and the relationship between quality or intensity of sensory stimuli; and minimizes the effects of irrelevant factors.

The panel as a whole may improve with training (Bennett et al., 1956; Ehrenberg and Shewan, 1953), although short selection and training programs may have no clear-cut effect

BEST AVAILABLE

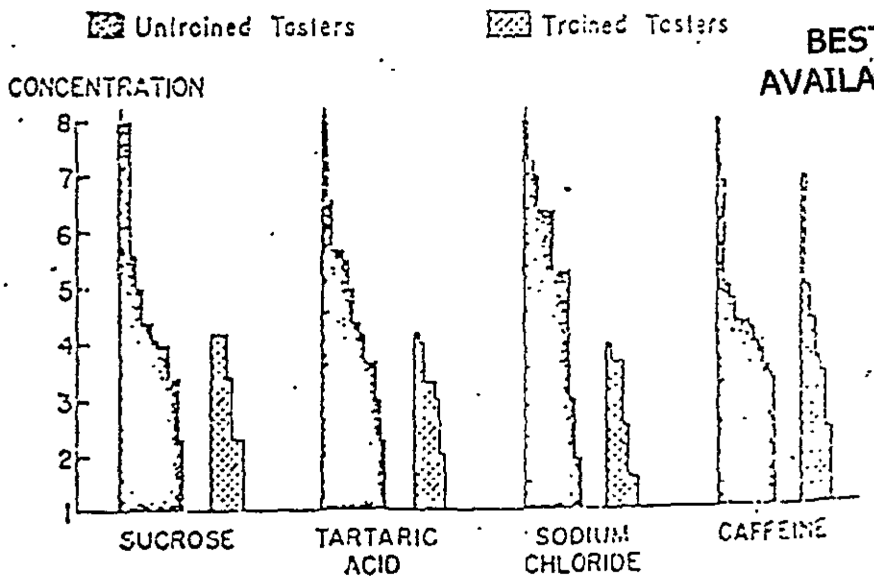


Fig. 3. Taste thresholds for twelve untrained and six trained tasters.

on the performance of the panel (Schlosberg *et al.*, 1954).

Special Tastes

Panel members may be chosen on the basis of their threshold sensitivities to substances other than the basic tastes, and trained with these substances. Solutions of insecticides (Dawson *et al.*, 1953; Gould *et al.*, 1951), sodium bisulfite (Boggs and Ward, 1950), imitation peppers (Peryam and Swartz, 1951), and meat extracts (McLean *et al.*, 1959) have been used.

However, evidence is only limited that the ability to detect differences in foods can be predicted by a panel member's threshold value (Hall *et al.*, 1959). Mackey and Jones (1954) reported that a judge who was good for tasting water solutions would not necessarily be a good judge for tasting foods. A judge who could readily distinguish a taste in one food was not always able to distinguish the same taste in other foods. Kirkpatrick *et al.* (1957) reported that a person acutely sensitive to one off-flavor in reconstituted dry milk may not be sensitive to all off-flavors.

Thresholds for different-textured foods with added primary tastes were affected by the type of food, texture, and levels of constituents added, and varied with each judge (Mackey and Valassi, 1956). Thresholds were lower for water solutions than for the same substances in foods. Campbell *et al.* (1958) and Hinreiner *et al.* (1955b) reported higher thresholds for substances added to beverages than for respective thresholds in water solutions. Mackey (1958) reported that the solvent medium affects discernment of taste substances. Primary tastes were easiest to detect in water, next in water and cellulose, and most difficult in oil. Apparently, substances must dissolve in the saliva before they can be tasted.

Girardot *et al.* (1952) stated that selecting panels on the basis of sensitivity to the four basic tastes is of limited value because only one factor of a complex situation is considered. Those authors reported a method of selecting panel members with the material to be tested. Procedures using the food to be evaluated for panel selection have also been reported by Kramer and Twigg (1962), Moser *et al.* (1950) and Wiley *et al.* (1957). The procedures involve testing the ability of panel members to discriminate between samples and/or to give reproducible qualitative judgments. The most reliable and sensitive persons are retained for the panel.

Odor Sensitivity

Panel members are also selected on their ability to distinguish odors as well as flavors. In screening over 600 people by odor-perception and odor-recognition series, no person was found completely anosmic (Caul, 1957). Olfactory thresholds are apparently related to food intake since variations in thresholds do not occur when noon meals are omitted (Goetzl *et al.*, 1950a). Methods of determining odor identifications have been reported by many persons, including Pilgrim and Schutz (1957), Johnston and Sandoval (1960) and Johnston and Parks (1960).

REFERENCES

Baten, W. D. 1950. Reaction of age groups to organoleptic tests. *Food Technol.* 4, 277.

Bennett, G., B. M. Spahr, and M. L. Dodds. 1956. The value of training a sensory test panel. *Food Technol.* 10, 205.

Berg, H. W., F. Filipello, E. Hinreiner, and A. D. Webb. 1955. Evaluation of thresholds and minimum difference concentrations for various constituents of wines. II. Sweetness: the effect of ethyl alcohol, organic acids and tannin. *Food Technol.* 9, 138.

Bernard, R. A., B. P. Halpern, and M. R. Kare. 1961. Effect of vitamin A deficiency on taste. *Proc. Soc. Exptl. Biol. Med.* 108, 784.

Blakeslee, A. F. 1932. Genetics of sensory thresholds: Taste for phenyl thio carbamide. *Proc. Natl. Acad. Sci.* 18, 120.

Boggs, M. M., and A. C. Ward. 1950. Scoring techniques for sulfited foods. *Food Technol.* 4, 282.

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

Campbell, C. L., R. K. Dawes, S. Deolalkar, and M. C. Merritt. 1958. Effect of certain chemicals in water on the flavor of brewed coffee. *Food Research* 23, 575.

Caul, J. F. 1957. The profile method of flavor analysis. *Advances in Food Research* 7, 1.

Cochran, W. G., and G. M. Cox. 1957. "Experimental Designs" 2d ed. John Wiley and Sons, New York.

Dallenbach, J. W., and K. M. Dallenbach. 1943. The effects of bitter adaptation on sensitivity. *Am. J. Psychol.* 56, 21.

Dawson, E. H., G. L. Gilpin, M. E. Kirkpatrick, and C. A. Weigel. 1953. Flavor of selected vegetables grown in soil treated with isomers of benzene hexachloride. *J. Agr. Food Chem.* 1, 399.

Ehrenberg, A. S. C., and J. M. Shewan. 1953. The objective approach to sensory tests of food. *J. Sci. Food Agr.* 10, 482.

Fabian, F. W., and H. B. Blum. 1943. Relative taste potency of some basic food constituents and their competitive and compensatory action. *Food Research* 8, 179.

Girardot, N. F., D. R. Peryam, and R. Shapiro. 1952. Selection of sensory testing panels. *Food Technol.* 6, 140.

Goetzl, F. R., M. S. Abel, and A. J. Ahokas. 1950a. Occurrence in normal individuals of diurnal variation of olfactory acuity. *J. Appl. Physiol.* 2, 553.

Goetzl, F. R., A. J. Ahokas, and J. G. Payne. 1950b. Occurrence in normal individuals of diurnal variation in acuity of the sense of taste for sucrose. *J. Appl. Physiol.* 2, 619.

Gould, W. A., J. P. Sleesman, R. W. Rings, M. Lynn, F. Krantz, Jr., and H. D. Brown. 1951. Flavor evaluations of canned fruits and vegetables treated with newer organic insecticides. *Food Technol.* 5, 129.

Hall, B. A., M. Tarver, and J. G. McDonald. 1959. A method for screening flavor panel members and its application to a two sample difference test. *Food Technol.* 3, 699.

Hinchcliffe, R. 1958. Aging and sensory thresholds. *J. Gerontol.* 17, 45

Hinreiner, E., F. Filipello, A. D. Webb, and H. W. Berg. 1955a. Evaluation of thresholds and minimum difference concentrations for various constituents of wines. III. Ethyl alcohol, glycerol and acidity in aqueous solutions. *Food Technol.* 9, 351.

Hinreiner, E., F. Filipello, H. W. Berg, and A. D. Webb. 1955b. Evaluation of thresholds and minimum difference concentrations for various constituents of wines. IV. Detectable differences in wine. *Food Technol.* 9, 489.

Janowitz, H. D., and M. I. Grossman. 1949. Gusto-olfactory thresholds in relation to appetite and hunger sensations. *J. Appl. Physiol.* 2, 217.

Johnston, J. W., Jr., and A. R. Parks. 1960. Odor-intensity and the stereochemical theory of olfaction. *Proc. Sci. Sec. Toilet Goods Assoc.* 34, 4

Johnston, J. W., Jr., and A. Sandoval. 1960. Organoleptic quality and the stereochemical theory of olfaction. *Proc. Sci. Sec. Toilet Goods Assoc.* 33, 3.

Kamen, J. M., and F. J. Pilgrim. 1959. Interaction of taste qualities. *Am. Psychologist* 11, 429.

Kamenetzky, J., and F. J. Pilgrim. 1958. The effect of sucrose upon the perceived intensity of salt and bitter. *Am. Psychologist* 13, 412.

Kare, M. R., and B. P. Halpern. 1961. "The Physiological and Behavioral Aspects of Taste." The University of Chicago Press, Chicago.

Kirkpatrick, M. E., J. C. Lamb, E. H. Dawson, and J. N. Eisen. 1957. Selection of a taste panel for evaluating the quality of processed milk. *Food Technol.* 11, 3.

Kotschevar, L. H. 1956. Taste testing frozen meat cooked before and after thawing. *J. Am. Dietet. Assoc.* 32, 444.

Kramer, A., and B. A. Twigg. 1962. "Fundamentals of Quality Control for the Food Industry." p. 105. Avi Publishing Co., Westport, Connecticut.

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

Mackey, A. O., and P. Jones. 1954. Selection of members of a food tasting panel: Discernment of primary tastes in water solution compared

BEST AVAILABLE

with judging ability for foods. *Food Technol.* 8, 527.

Mackey, A. O., and K. Valassi. 1956. The discernment of primary tastes in the presence of different food textures. *Food Technol.* 10, 238.

McLean, R. A., A. M. Gaddis, F. G. Suggs, and P. F. Crowe. 1959. Evaluation of flavors in meat by the use of aqueous extracts. *Food Technol.* 13, 256.

Moser, H. A., H. J. Dutton, C. D. Evans, and J. C. Cowan. 1950. Conducting a taste panel for the evaluation of edible oils. *Food Technol.* 4, 105.

Neilson, A. J. 1957. Time-intensity studies. *Drug & Cosmetic Ind.* 80, 452.

Pangborn, R. M. 1959. Influence of hunger on sweetness preferences and taste thresholds. *Am. J. Clin. Nutrition* 7, 280.

Pangborn, R. M. 1960. Taste interrelationships. *Food Research* 25, 245.

Pangborn, R. M. 1961. Taste interrelationships. II. Suprathreshold solutions of sucrose and citric acid. *J. Food Sci.* 26, 648.

Peryam, D. R., and V. W. Swartz. 1951. Methodology for sensory evaluation of imitation peppers. *Food Technol.* 5, 207.

Pfaffmann, C., and E. C. Hagstrom. 1955. Factors influencing taste sensitivity to sugar. *Am. J. Physiol.* 183, 651.

Pilgrim, F. J., and H. G. Schutz. 1957. Measurement of the qualitative and quantitative attributes of flavor. Chemistry of natural food flavors—a symposium. Quartermaster Food and Container Inst. for the Armed Forces. Chicago, p. 47.

Pilgrim, F. J., H. G. Schutz, and D. R. Peryam. 1955. Influence of monosodium glutamate on taste perception. *Food Research* 20, 310.

Richter, C. P., and K. H. Campbell. 1940. Sucrose taste thresholds of rats and humans. *Am. J. Physiol.* 128, 291.

Richter, C. P., and A. MacLean. 1939. Salt taste thresholds of humans. *Am. J. Physiol.* 126, 1.

Schlosberg, H., C. Pfaffmann, J. Cornsweet, and R. Pierrel. 1954. Selection and training of panels. Food acceptance testing methodology. Quartermaster Food and Container Inst., Chicago, p. 45.

Seifter, E., H. Rothberg, and J. I. Berkman. 1960. Effect of vitamin A deficiency on taste perception of rats. Abstracts of papers. Am. Chem. Soc. 137th Meeting, Cleveland, p. 17N.

Valdés, R. M., E. H. Hinreiner, and M. J. Simone. 1956. Effect of sucrose and organic acids on apparent flavor intensity. I. Aqueous solutions. *Food Technol.* 10, 282.

Wiley, R. C., A. M. Briant, I. S. Fagerson, E. F. Murphy, and J. H. Sabry. 1957. The Northeast regional approach to collaborative panel testing. *Food Technol.* 11, 43.

Yensen, R. 1958. Influence of salt deficiency on taste sensitivity in human subjects. *Nature* 181, 1472.

THE AUTHORS are with the Human Nutrition Research Division, Agricultural Research Service, U. S. Department of Agriculture, Beltsville, Md.

Presented at the 22nd Annual Meeting of the Institute of Food Technologists, Miami Beach, Florida, June, 1962.

Sensory Testing of Differences in Taste

I. Methods · II. Selection of Panel Members

Prepared for: U. S. Department of Agriculture, Agricultural Research Service, Federal Center Building, Hyattsville, Maryland

EXPANDED STATISTICAL TABLES FOR ESTIMATING SIGNIFICANCE IN PAIRED-PREFERENCE, PAIRED-DIFFERENCE, DUO-TRIO AND TRIANGLE TESTS

E. B. ROESSLER, R. M. PANGBORN, J. L. SIDEL and H. STONE

ABSTRACT

Two sets of expanded tables have been compiled for use in determining significance in paired-difference and triangle tests (one-tailed) and in paired-preference tests (two-tailed). One set of tables lists the number of correct responses (or agreeing judgments) for trials ranging from 7–100, at $p < 0.05$, 0.04, 0.03, 0.02, 0.01, 0.005 and 0.001. These tables are convenient for a quick estimate of significance of laboratory sensory data as well as consumer responses. The second set of tables gives the probabilities of obtaining a given number of correct (or agreeing) judgments in trials ranging from 5–50. These probability tables provide a more precise estimate of significance, which may be needed in more critical research or in making decisions of considerable importance. Some examples are given, with guidelines for the proper use of these tables and the interpretation of significance based upon them.

TABLES used to determine significance in discrimination and preference tests usually indicate the number of judgments required at only three levels of significance, i.e., the 5%, 1% and 0.1% levels (Roessler et al., 1948, 1956). Stone and Sidel (1978) have pointed out the inconsistencies in the entries in subsequent tables of this type (Amerine et al., 1965; Larmond, 1970; Stald and Einstein, 1973). Since these tables appear to have been constructed using different criteria, it is recommended that exact probabilities be obtained from tables of the cumulative binomial probability distribution, or in the event that such tables are not available, that approximate probabilities be computed using the normal curve. This procedure leaves no doubt as to the true probability. To be *almost* significant at a certain probability level is not the same as being significant at that level. Assurance of significance of the occurrence of an event at a particular level of α requires that the probability of the event occurring is equal to or less than α. We cannot concur with Basker (1976) that his contrived tables of probability for triangle testing by individual judges are a satisfactory approximation which can be used "instead of unwieldy tables of exact significance levels."

Since many investigators prefer, and will continue to use, tables giving the number of judgments required for significance at various levels, it is desirable to have tables with more than three probability levels. Otherwise, conclusions may disregard valuable information. For example, in 46 trials of a triangle test, 22 correct judgments are required for significance at $p < 0.05$. However, 22 correct judgments are also significant at $p < 0.03$, which represents considerably better performance and will be overlooked with tables giving only the usual three levels of $p < 0.05$, 0.01, and 0.001.

Tables for correct judgments

Table 1 gives the numbers of correct judgments required for significance in the paired-difference and duo-trio tests. These are one-tailed tests as only one response is correct, and [illegible] = ½ (Amerine et al., 1965). Table 2 gives the same informat[illegible] for the triangle test which is one-tailed with $p = 1/3$. Table [illegible] for use in paired presentation for preference which is [illegible]

Table 1 – Minimum numbers of correct judgments to establish significance at various probability levels for paired-difference and duo-trio tests (one-tailed, p = ½)[a]

No. of trials (n)	Probability levels: 0.05	0.04	0.03	0.02	0.01	0.005	0.001
7	7	7	7	7	7		
8	7	7	8	8	8	8	
9	8	8	8	8	9	9	
10	9	9	9	9	10	10	10
11	9	9	10	10	10	11	11
12	10	10	10	10	11	11	12
13	10	11	11	11	12	12	13
14	11	11	11	12	12	13	13
15	12	12	12	12	13	13	14
16	12	12	13	13	14	14	15
17	13	13	13	14	14	15	16
18	13	14	14	14	15	15	16
19	14	14	15	15	15	16	17
20	15	15	15	16	16	17	18
21	15	15	16	16	17	17	18
22	16	16	16	17	17	18	19
23	16	17	17	17	18	19	20
24	17	17	18	18	19	19	20
25	18	18	18	19	19	20	21
26	18	18	19	19	20	20	22
27	19	19	19	20	20	21	22
28	19	20	20	20	21	22	23
29	20	20	21	21	22	22	24
30	20	21	21	22	22	23	24
31	21	21	22	22	23	24	25
32	22	22	22	23	24	24	26
33	22	23	23	23	24	25	26
34	23	23	23	24	25	25	27
35	23	24	24	25	25	26	27
36	24	24	25	25	26	27	28
37	24	25	25	26	26	27	29
38	25	25	26	26	27	28	29
39	26	26	26	27	28	28	30
40	26	27	27	27	28	29	30
41	27	27	27	28	29	30	31
42	27	28	28	29	29	30	32
43	28	28	29	29	30	31	32
44	28	29	29	30	31	31	33
45	29	29	30	30	31	32	34
46	30	30	30	31	32	33	34
47	30	30	31	31	32	33	35
48	31	31	31	32	33	34	36
49	31	32	32	33	34	34	36
50	32	32	33	33	34	35	37
60	37	38	38	39	40	41	43
70	43	43	44	45	46	47	49
80	48	49	49	50	51	52	55
90	54	54	55	56	57	58	61
100	59	60	60	61	63	64	66

[a] Values (X) not appearing in table may be derived from $X = (z\sqrt{n} + n + 1)/2$. See text.

Author Roessler is with the Dept. of Mathematics, Univ. of California, Davis, CA 95616. Author Pangborn is with the Dept. of Food Science & Technology, Univ. of California, Davis, CA 95616. Authors Sidel and Stone are with Tragon Corp., Palo Alto, CA 94302.

0022-1147/78/0003-0940$02.25/0

Table 2—Minimum numbers of correct judgments to establish significance at various probability levels for the triangle test (one-tailed, p = 1/3)[a]

No. of trials (n)	Probability levels						
	0.05	0.04	0.03	0.02	0.01	0.005	0.001
5	4	5	5	5	5	5	
6	5	5	5	5	6	6	
7	5	6	6	6	6	7	7
8	6	6	6	6	7	7	8
9	6	7	7	7	7	8	8
10	7	7	7	7	8	8	9
11	7	7	8	8	8	9	10
12	8	8	8	8	9	9	10
13	8	8	9	9	9	10	11
14	9	9	9	9	10	10	11
15	9	9	10	10	10	11	12
16	9	10	10	10	11	11	12
17	10	10	10	11	11	12	13
18	10	11	11	11	12	12	13
19	11	11	11	12	12	13	14
20	11	11	12	12	13	13	14
21	12	12	12	13	13	14	15
22	12	12	13	13	14	14	15
23	12	13	13	13	14	15	16
24	13	13	13	14	15	15	16
25	13	14	14	14	15	16	17
26	14	14	14	15	15	16	17
27	14	14	15	15	16	17	18
28	15	15	15	16	16	17	18
29	15	15	16	16	17	17	19
30	15	16	16	16	17	18	19
31	16	16	16	17	18	18	20
32	16	16	17	17	18	19	20
33	17	17	17	18	18	19	21
34	17	17	18	18	19	20	21
35	17	18	18	19	19	20	22
36	18	18	18	19	20	20	22
37	18	18	19	19	20	21	22
38	19	19	19	20	21	21	23
39	19	19	20	20	21	22	23
40	19	20	20	21	21	22	24
41	20	20	20	21	22	23	24
42	20	20	21	21	22	23	25
43	20	21	21	22	23	24	25
44	21	21	22	22	23	24	26
45	21	22	22	23	24	24	26
46	22	22	22	23	24	25	27
47	22	22	23	23	24	25	27
48	22	23	23	24	25	26	27
49	23	23	24	24	25	26	28
50	23	24	24	25	26	26	28
60	27	27	28	29	30	31	33
70	31	31	32	33	34	35	37
80	35	35	36	36	38	39	41
90	38	39	40	40	42	43	45
100	42	43	43	44	45	47	49

[a] Values (X) not appearing in table may be derived from: $X = 0.4714 z \sqrt{n} + [(2n + 3)/6]$. See text.

Table 3—Minimum numbers of agreeing judgments necessary to establish significance at various probability levels for the paired-preference test (two-tailed, p = ½)[a]

No. of trials (n)	Probability levels						
	0.05	0.04	0.03	0.02	0.01	0.005	0.001
7	7	7	7	7			
8	8	8	8	8	8		
9	8	8	9	9	9	9	
10	9	9	9	10	10	10	
11	10	10	10	10	11	11	11
12	10	10	11	11	11	12	12
13	11	11	11	12	12	12	13
14	12	12	12	12	13	13	14
15	12	12	13	13	13	14	14
16	13	13	13	14	14	14	15
17	13	14	14	14	15	15	16
18	14	14	15	15	15	16	17
19	15	15	15	15	16	16	17
20	15	16	16	16	17	17	18
21	16	16	16	17	17	18	19
22	17	17	17	17	18	18	19
23	17	17	18	18	19	19	20
24	18	18	18	19	19	20	21
25	18	19	19	19	20	20	21
26	19	19	19	20	20	21	22
27	20	20	20	20	21	22	23
28	20	20	21	21	22	22	23
29	21	21	21	22	22	23	24
30	21	22	22	22	23	24	25
31	22	22	22	23	24	24	25
32	23	23	23	23	24	25	26
33	23	23	24	24	25	25	27
34	24	24	24	25	25	26	27
35	24	25	25	25	26	27	28
36	25	25	25	26	27	27	29
37	25	26	26	26	27	28	29
38	26	26	27	27	28	29	30
39	27	27	27	28	28	29	31
40	27	27	28	28	29	30	31
41	28	28	28	29	30	30	32
42	28	29	29	29	30	31	32
43	29	29	30	30	31	32	33
44	29	30	30	30	31	32	34
45	30	30	31	31	32	33	34
46	31	31	31	32	33	33	35
47	31	31	32	32	33	34	36
48	32	32	32	33	34	35	36
49	32	33	33	34	34	35	37
50	33	33	34	34	35	36	37
60	39	39	39	40	41	42	44
70	44	45	45	46	47	48	50
80	50	50	51	51	52	53	55
90	55	56	56	57	58	59	61
100	61	61	62	63	64	65	67

[a] Values (X) not appearing in table may be derived from: $X = (z\sqrt{n} + n + 1)/2$. See text.

tailed (as either sample may be preferred) and p = ½ (Amerine et al., 1965). For numbers of trials not shown in the tables, numbers of required judgments may be obtained from tables of the cumulative binomial probabilities or, excellent approximations may be computed from the following formulas:

Tables 1 and 3: $X = (z\sqrt{n} + n + 1)/2$

Table 2: $X = 0.4717 z \sqrt{n} + [(2n + 3)/6]$

where n is the number of trials, and the minimum number of correct (agreeing) judgments is X, if X is a whole number, or the next higher integer if X is not a whole number, and where z is taken from the following table.

Values of z							
Probability (α)	0.05	0.04	0.03	0.02	0.01	0.005	0.001
Tables 1 and 2	1.64	1.75	1.88	2.05	2.33	2.58	3.10
Table 3	1.96	2.05	2.17	2.33	2.58	2.81	3.30

Table 4—Probability of X or more correct judgments in n trials (one-tailed, p = ½)[a]

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
[illegible]		969	812	500	188	031																														
[illegible]		984	891	656	344	109	016																													
[illegible]		992	938	773	500	227	062	008																												
[illegible]		996	965	855	637	363	145	035	004																											
[illegible]		998	980	910	746	500	254	090	020	002																										
[illegible]		999	989	945	828	623	377	172	055	011	001																									
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[a] NOTE: Initial decimal point has been omitted.

[A]s an example of the use of the formulas for values exceeding tabular entries, consider the case of preferences between two canned meats by 150 consumers. Since Table 3 [exte]nds only to n = 100, one would apply the formula for the [paire]d-preference test (two-tailed, p = ½):

$$X = \frac{z\sqrt{150} + 150 + 1}{2}$$

[I]nserting the z values corresponding to probabilities of 0.05, 0.04, 0.03, 0.02, 0.01, 0.005 and 0.001, the minimum num[ber] of agreeing judgments (X) are 88, 88, 89, 90, 92, 93 and [95 r]espectively.

[T]ables of probabilities

[Ta]bles 1–3 are convenient to consult for quick estimates of [sign]ificance. For much greater accuracy, exact probabilities [s]hould be calculated as recommended by Stone and Sidel [(]1978). Tables 4, 5 and 6 list the probabilities of obtaining X [or m]ore correct responses (one-tailed) or agreeing judgments [(two-]tailed) in n trials.[1] Note that the initial decimal point has [b]een omitted in the body of the tables to save space, i.e., the [v]alue of 969 should be 0.969.

[If] an experimenter collected 50 paired comparisons to [dete]rmine if an expert panel could detect the added vanilla in [a] pudding mix, and obtained 31 correct responses, consulta[tion] of Table 1 would lead him to conclude there was no [sign]ificant difference, i.e., not significant at $p < 0.05$. How[eve]r, consultation of Table 4 shows that the probability of [o]btaining 31 out of 50 judgments in a paired-difference test by [c]hance alone is 0.059. The experimenter would not only be [more] accurate in reporting that there were significant dif[fe]rences between the pudding formulas at p = 0.059, but [m]ight change his mind regarding a "go, no-go" decision on the [pro]duct. If he is willing to take a chance at the 5% level, i.e., [of] being wrong five times out of a hundred trials, might he [n]ot also be willing to take a chance at the equally-arbitrary [cu]t-off of 5.9%?

[A]n understanding of probability is essential for the correct [use o]f these tables in order to avoid misinterpretation of the [sig]nificance of results from sensory testing. Furthermore, dis[c]retion is needed in the interpretation of n in these tables. For

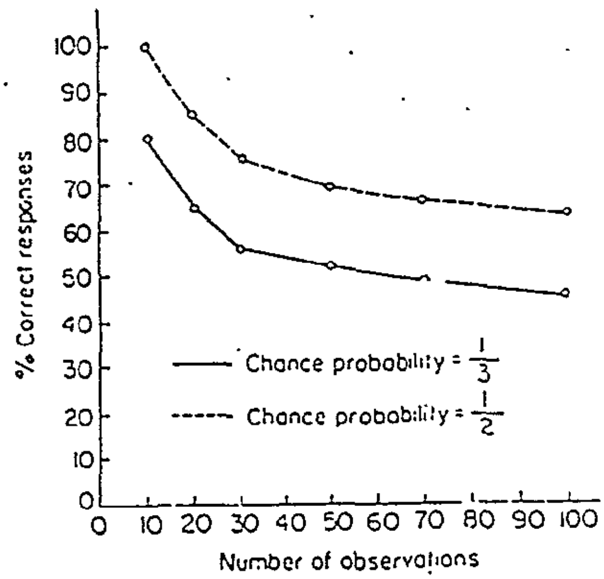


Fig. 1—Plot of the percent correct responses necessary for significance at p < 0.01 for one-tailed tests when chance probabilities are 1/3 (triangle tests) or 1/2 (paired-difference tests). Reprinted from Amerine et al. (1965) with permission from Academic Press, New York.

example, for n = 20, the same values apply for 1 judge × 20 sessions as for 20 judges × 1 session, but the interpretation of the result, obviously, is different.

Tables 1–3 list the number of correct decisions for a maximum n of 100. However, these values should not be misconstrued as recommendations for the maximum or minimum number of trials. In most situations, in fact, discrimination testing requires less than 50 trials in order to reach a reliable and valid decision. It is important to recognize that as n increases, the percent correct observations needed for statistical significance decreases, but in a nonlinear fashion. Figure 1 is a graphic representation of correct responses needed for significance at $p < 0.01$ for chance probabilities of ½ (paired tests)

n \ X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
5	864	539	210	045	004																							
6	912	649	320	100	018	001																						
7	941	737	429	173	045	007																						
8	961	805	632	259	088	020	003																					
9	974	857	623	350	145	042	008	001																				
10	983	896	701	441	213	077	020	003																				
11	988	925	766	527	289	127	039	009	001																			
12	992	946	819	607	368	178	066	019	[illegible]	001																		
13	995	961	861	673	448	241	104	035	009	002																		
14	997	973	895	720	524	310	149	058	017	004	001																	
15	998	981	921	791	596	382	203	088	031	008	002																	
16	998	986	941	834	661	453	763	176	050	016	004	001																
17	993	930	956	870	719	572	326	172	075	027	008	002																
18	929	933	067	895	769	568	391	223	108	043	014	004	001															
19		995	976	921	812	648	457	279	146	065	024	007	002															
20		997	982	940	848	703	521	339	191	092	038	013	004	001														
21		998	987	954	879	751	581	399	240	125	056	021	007	002														
22		998	991	965	904	794	638	400	293	163	079	033	012	003	001													
23		999	993	974	924	831	690	519	349	206	107	048	019	006	002													
24		999	995	980	941	862	737	576	406	254	140	068	028	010	003	001												
25		999	996	985	954	888	778	630	462	304	178	092	042	016	006	002												
26			997	989	964	910	815	679	518	357	220	121	058	025	009	003	001											
27			998	992	972	928	847	725	572	411	268	154	079	036	014	005	002											
28			999	994	979	943	874	765	623	464	314	191	104	050	022	008	003	001										
29			999	996	984	955	897	801	670	517	364	232	133	068	031	013	005	001										
30			999	997	988	965	916	823	714	568	415	276	166	090	043	019	007	002	001									
31				998	991	972	932	861	754	617	466	322	203	115	059	027	011	004	001									
32				998	993	978	946	885	783	662	516	370	243	144	078	038	016	006	002	001								
33				999	955	983	957	905	821	705	565	419	285	177	100	051	023	010	004	001								
34				999	996	987	965	922	849	744	612	468	330	213	126	067	033	014	006	002	001							
35				999	997	990	973	937	873	779	650	516	376	252	155	087	044	020	009	003	001							
36					998	992	978	949	895	810	697	562	422	293	187	109	054	025	012	005	002	001						
37					998	994	983	959	913	838	735	607	469	336	223	135	075	038	018	007	003	001						
38					999	996	987	967	928	863	769	650	515	381	261	164	095	051	025	011	004	002	001					
39					999	997	990	973	941	885	800	689	560	425	301	196	118	066	033	016	007	003	001					
40					999	997	992	979	952	903	829	726	603	470	342	231	144	083	044	021	010	004	001					
41						998	994	983	961	920	854	761	644	515	385	268	173	104	057	029	014	006	002	001				
42						999	995	987	968	933	876	791	683	558	428	307	205	127	073	038	019	008	003	001				
43						999	996	990	974	945	895	820	719	600	471	347	239	153	091	050	025	012	005	002	001			
44						999	997	992	980	955	912	845	753	639	514	389	275	182	111	063	033	016	007	003	001			
45						999	998	994	984	963	926	867	783	677	556	430	313	213	135	079	043	022	010	004	002	001		
46							998	995	987	970	938	887	811	713	598	472	352	246	161	098	055	029	014	006	003	001		
47							999	996	990	976	949	904	836	745	625	514	392	282	189	113	070	038	019	009	004	002	001	
48							999	997	992	980	958	919	859	776	672	554	433	318	220	142	086	048	025	012	006	002	001	
49							999	998	994	984	965	932	879	803	706	593	473	356	253	168	105	061	033	017	009	003	001	
50							999	998	995	987	972	943	896	829	739	631	513	395	287	195	126	076	042	022	011	005	002	001

a NOTE: Initial decimal point has been omitted.

and 1/3 (triangle tests), showing that beyond an n of 30, the percent correct responses changes very slightly.

In Tables 1 and 2, for n = 100, note that only 59% correct responses are required in a paired-difference test, while only 42% correct responses are required in a triangle test to attain significance at $p < 0.05$. Statistical significance is achieved, but from a *practical* point of view, this may not be an important difference, as in both cases, only nine correct decisions above chance are necessary to be significant at $p < 0.05$. The behavioral implications and interpretation of statistical tests

—*Continued on page 947*

Table 6—Probability of X or more agreeing judgments in n trials (two-tailed, p = ½)[a]

n \ X	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
5	375	312	062																																
6		644	219	031																															
7			453	125	016																														
8			727	289	070	008																													
9				508	180	039	004																												
10				754	344	109	021	002																											
11					549	227	065	011	001																										
12					774	388	146	039	006																										
13						581	267	092	022	003																									
14						791	424	180	057	013	002																								
15							607	302	118	035	007	001																							
16							804	454	210	077	021	004	001																						
17								629	332	143	049	013	002																						
18								815	481	238	096	031	008	001																					
19									648	359	167	064	019	004	001																				
20									824	503	263	115	041	012	003																				
21										664	383	189	078	027	007	001																			
22										832	523	286	134	052	017	004	001																		
23											678	405	210	093	035	011	003																		
24											839	541	307	152	064	023	007	002																	
25												690	424	230	108	043	015	004	001																
26												845	557	327	169	076	029	009	002	001															
27													701	442	248	122	052	019	006	002															
28													851	572	345	185	087	036	013	004	001														
29														711	458	265	136	061	024	008	002	001													
30														856	585	362	200	099	043	016	005	001													
31															720	473	281	150	071	029	011	003	001												
32															860	597	377	215	100	050	020	007	002	001											
33																728	487	296	163	080	035	014	005	001											
34																864	608	392	229	121	058	024	009	003	001										
35																	736	500	310	175	090	041	017	006	002										
36																	868	617	405	243	132	065	029	011	004	001									
37																		743	511	324	188	099	047	020	008	003	001								
38																		871	627	418	256	143	073	034	014	005	002								
39																			749	522	337	200	108	053	024	009	003	001							
40																			875	636	430	268	154	081	038	017	006	002	001						
41																				755	533	349	211	117	060	028	012	004	001						
42																				878	644	441	280	164	088	044	020	008	003	001					
43																					761	542	360	227	126	066	032	014	005	002	001				
44																					880	652	451	291	174	096	049	023	010	004	001				
45																						764	551	371	233	135	072	036	016	007	003	001			
46																						883	659	461	302	184	104	054	026	011	005	002	001		
47																							771	560	387	243	144	077	040	019	008	003	001		
48																							885	665	471	312	193	111	059	029	013	006	002	001	
49																								775	568	392	253	151	085	044	021	009	004	001	
50																								888	672	480	322	203	119	065	033	015	007	003	001

a NOTE: Initial decimal point has been omitted.

EXPANDED STATISTICAL TABLES FOR ESTIMATING TESTS . . . From page 943

xtend well beyond the relatively straightforward computation for estimation of statistical significance. Texts by Edwards (1965), McCall (1970), Huff (1974) and Reichmann (1961), among others, provide extensive statistical information and recommendations for applications to the behavioral sciences.

REFERENCES

merine, M.A., Pangborn, R.M. and Roessler, E.B. 1965. "Principles of Sensory Evaluation of Foods," Academic Press, New York.

ker, D. 1976. Comparison of discrimination ability between taste panel assessors. Chemical Senses & Flavor 2: 207.

wards, A.L. 1965. "Statistical Methods for the Behavioral Sciences," Rinehart, New York.

ff, D. 1974. "How to Lie with Statistics," Penguin Books, Philadelphia.

mond, E. 1970. "Methods for Sensory Evaluation of Food," Publ. #1284, Canada Dept. of Agriculture, Ottawa.

McCall, R.B. 1970. "Fundamental Statistics for Psychology," Harcourt, Brace, and World, New York.

Reichmann, W.J. 1961. "Use and Abuse of Statistics," Methuen & Co., Ltd., London.

Roessler, E.B., Warren, J. and Guymon, J.F. 1948. Significance in triangular taste tests. Food Res. 13: 503.

Roessler, E.B., Baker, G.A. and Amerine, M.A. 1953. Corrected normal and chi-square approximations to the binomial distribution in organoleptic tests. Food Res. 18: 625.

Roessler, E.B., Baker, G.A. and Amerine, M.A. 1956. One-tailed and two-tailed test in organoleptic comparisons. Food Res. 21: 117.

Stahl, W.H. and Einstein, M.A. 1973. Sensory testing methods. In "Encyclopedia of Industrial Chemical Analysis," 17: 608

Stone, H. and Sidel, J.L. 1978. Computing exact probabilities in sensory discrimination tests. J. Food Sci. 43:

Ms received 7/25/77; revised 11/4/77; accepted 11/10/77.

[1] Values not appearing in the table may be obtained from "Tables of the Cumulative Binomial Probability Distribution." Annals of the Computation Laboratory of Harvard University, Harvard University Press, Cambridge, Mass., 35:1955.

TABLE T7
Triangle Test for Difference: Critical Number (Minimum) of Correct Answers

Entries are the minimum number of correct responses required for significance at the stated significance level (i.e., column) for the corresponding number of respondents "n" (i.e., row). Reject the assumption of "no d[illegible] responses is greater than or equal to the tabled value.

n	Significance level (%) 10	5	1	0.1
3	3	3	—	—
4	4	4	—	—
5	4	4	5	—
6	5	5	6	—
7	5	5	6	7
8	5	6	7	8
9	6	6	7	8
10	6	7	8	9
11	7	7	8	10
12	7	8	9	10
13	8	8	9	11
14	8	9	10	11
15	8	9	10	12
16	9	9	11	12
17	9	10	11	13
18	10	10	12	13
19	10	11	12	14
20	10	11	13	14
21	11	12	13	15
22	11	12	14	15
23	12	12	14	16
24	12	13	15	16
25	12	13	15	17

n	Significance level (%) 10	5	1	0.1
26	13	14	15	17
27	13	14	16	18
28	14	15	16	18
29	[illegible]	[illegible]	[illegible]	19
30	14	15	17	19
31	15	16	18	20
32	15	16	18	20
33	15	17	18	21
34	16	17	19	21
35	16	17	19	22
36	17	18	20	22
42	19	20	22	25
48	21	22	25	27
54	23	25	27	30
60	26	27	30	33
66	28	29	32	35
72	30	32	34	38
78	32	34	37	40
84	35	36	39	43
90	37	38	42	45
96	39	41	44	48

Note. For values of n not in the table compute $z = (k - (1/3)n)/\sqrt{(2/9)n}$, where k is the number of correct answers. Compare the computed value of z to the critical value of a standard normal random variable, i.e., the values in the last row of Table T4 ($z_\alpha = t_{\alpha,\infty}$).

TABLE T8
Duo-Trio Test for Difference or One-Sided Paired Comparison Test for Difference: Critical Number (Minimum) of Correct Answers

Entries are the minimum number of correct responses required for significance at the stated significance level (i.e., column) for the corresponding number of respondents "n" (i.e., row). Reject the assumption of "no difference" if the number of correct responses is greater than or equal to the tabled value

n	Significance level (%) 10	5	1	0.1
4	4	—	—	—
5	5	5	—	—
6	6	6	—	—
7	6	7	7	—
8	7	7	8	—
9	7	8	9	—
10	8	9	10	10
11	9	9	10	11
12	9	10	11	12
13	10	10	12	13
14	10	11	12	13
15	11	12	13	14
16	12	12	14	15
17	12	13	14	16
18	13	13	15	16
19	13	14	15	17
20	14	15	16	18
21	14	15	17	18
22	15	16	17	19
23	16	16	18	20
24	16	17	19	20
25	17	18	19	21
26	17	18	20	22
27	18	19	20	22
28	18	19	21	23
29	19	20	22	24
30	20	20	22	24
31	20	21	23	25
32	21	22	24	26
33	21	22	24	26
34	22	23	25	27
35	22	23	25	27
36	23	24	26	28
40	25	26	28	31
44	27	28	31	33
48	29	31	33	36
52	32	33	35	38
56	34	35	38	40
60	36	37	40	43
64	38	40	42	45
68	40	42	45	48
72	42	44	47	50
76	45	46	49	52
80	47	48	51	55
84	49	51	54	57
88	51	53	56	59
92	53	55	58	62
96	55	57	60	64
100	57	59	63	66

Note: For values of n not in the table compute $z = (k - 0.5\,n)/\sqrt{0.25\,n}$, where k is the number of correct answers. Compare the computed value of z to the critical value of a standard normal random variable, i.e., the values in the last row of Table T4 ($z_\alpha = t_{\alpha,\infty}$).

TABLE T9
Two-Sided Paired Comparison Test for Difference: Critical Number (Minimum) of Correct Answers

Entries are the minimum number of correct responses required for significance at the stated significance level (i.e., column) for the corresponding number of respondents "n" (i.e., row). Reject the assumption of "no difference" if the number of correct responses is greater than or equal to the tabled value.

n	Significance level (%) 10	5	1	0.1	n	Significance level (%) 10	5	1	0.1
5	5	—	—	—	31	21	22	24	25
					32	22	23	24	26
6	6	6	—	—	33	22	23	25	27
7	7	7	—	—	34	23	24	25	27
8	7	8	8	—	35	23	24	26	28
9	8	8	9	—					
10	9	9	10	—	36	24	25	27	29
					40	26	27	29	31
11	9	10	11	11	44	28	29	31	34
12	10	10	11	12	48	31	32	34	36
13	10	11	12	13	52	33	34	36	39
14	11	12	13	14					
15	12	12	13	14					
16	12	13	14	15	56	35	36	39	41
17	13	13	15	16	60	37	39	41	44
18	13	14	15	17	64	40	41	43	46
19	14	15	16	17	68	42	43	46	48
20	15	15	17	18	72	44	45	48	51
21	15	16	17	19	76	46	48	50	53
22	16	17	18	19	80	48	50	52	56
23	16	17	19	20	84	51	52	55	58
24	17	18	19	21	88	53	54	57	60
25	18	18	20	21	92	55	56	59	63
					96	57	59	62	65
26	18	19	20	22	100	59	61	64	67
27	19	20	21	23					
28	19	20	22	23					
29	20	21	22	24					
30	20	21	23	25					

Note: For values of n not in the table compute $z = (k - 0.5\,n)/\sqrt{0.25\,n}$, where k is the number of correct answers. Compare the computed value of z to the critical value of a standard normal random variable, i.e., the values in the last row in Table T4($z_{\alpha} = t_{\alpha,\infty}$).

TABLE T10
Two-out-of-Five Test for Difference: Critical Number (Minimum) of Correct Answers

Entries are the minimum number of correct responses required for significance at the stated significance level (i.e., column) for the corresponding number of respondents (n, i.e., row). Reject the assumption of "no difference" if the number of correct responses is greater than or equal to the tabled value.

	Significance level (%)			
n	10	5	1	0.1
2	2	2	2	—
3	2	2	3	3
4	2	3	3	4
5	2	3	3	4
6	3	3	4	5
7	3	3	4	5
8	3	3	4	5
9	3	4	4	5
10	3	4	5	6
11	3	4	5	6
12	4	4	5	6
13	4	4	5	6
14	4	4	5	7
15	4	5	6	7
16	4	5	6	7
17	4	5	6	7
18	4	5	6	8
19	5	5	6	8
20	5	5	7	8
21	5	6	7	8
22	5	6	7	8
23	5	6	7	9
24	5	6	7	9
25	5	6	7	9
26	6	6	8	9
27	6	6	8	9
28	6	7	8	10
29	6	7	8	10
30	6	7	8	10
31	6	7	8	10
32	6	7	9	10
33	7	7	9	11
34	7	7	9	11
35	7	8	9	11

	Significance level (%)			
n	10	5	1	0.1
36	7	8	9	11
37	7	8	9	11
38	7	8	10	11
39	7	8	10	12
40	7	8	10	12
41	8	8	10	12
42	8	9	10	12
43	8	9	10	12
44	8	9	11	12
45	8	9	11	13
46	8	9	11	13
47	8	9	11	13
48	9	9	11	13
49	9	10	11	13
50	9	10	11	14
51	9	10	12	14
52	9	10	12	14
53	9	10	12	14
54	9	10	12	14
55	9	10	12	14
56	10	10	12	14
57	10	11	12	15
58	10	11	13	15
59	10	11	13	15
60	10	11	13	15
70	11	12	14	17
80	13	14	16	18
90	14	15	17	20
100	15	16	19	21

Note: For values of n not in the table compute $z = (k - 0.1\,n)/\sqrt{0.09\,n}$, where k is the number of correct answers. Compare the computed value of z to the critical value of a standard normal random variable, i.e., the values in the last row in Table T4($z_\alpha = t_{\alpha,\infty}$).

TABLE T11
Triangle Test for Similarity: Critical Number (Maximum) of Correct Answers

Accept the null hypothesis of no difference with 100(1 − β)% confidence if the number of correct responses is less than or equal to the number in the table that corresponds to the specified values of n, β, and p_d, where p_d is the proportion of the population that can distinguish the samples

		p_d						p_d			
n	β	0.15	0.20	0.25	0.30	n	β	0.15	0.20	0.25	0.30
18	0.001	—	—	—	—	45	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	15
	0.05	—	—	—	—		0.05	—	15	16	17
	0.10	—	—	—	6		0.10	—	16	17	19
21	0.001	—	—	—	—	48	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	17
	0.05	—	—	—	—		0.05	—	16	17	19
	0.10	—	—	7	7		0.10	—	17	19	20
24	0.001	—	—	—	—	51	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	18
	0.05	—	—	—	8		0.05	—	17	19	20
	0.10	—	—	8	9		0.10	17	18	20	22
27	0.001	—	—	—	—	54	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	18	19
	0.05	—	—	—	9		0.05	—	18	20	22
	0.10	—	—	9	10		0.10	18	20	21	23
30	0.001	—	—	—	—	57	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	19	21
	0.05	—	—	10	11		0.05	—	19	21	23
	0.10	—	10	10	11		0.10	19	21	23	25
33	0.001	—	—	—	—	60	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	20	22
	0.05	—	—	11	12		0.05	—	21	23	25
	0.10	—	11	12	13		0.10	20	22	24	26
36	0.001	—	—	—	—	66	0.001	—	—	—	22
	0.01	—	—	—	—		0.01	—	—	23	25
	0.05	—	—	12	13		0.05	—	23	25	28
	0.10	—	12	13	14		0.10	22	25	27	29
39	0.001	—	—	—	—	72	0.001	—	—	—	24
	0.01	—	—	—	13		0.01	—	—	25	28
	0.05	—	—	13	15		0.05	—	26	28	30
	0.10	—	13	15	16		0.10	25	27	30	32
42	0.001	—	—	—	—	78	0.001	—	—	—	27
	0.01	—	—	—	14		0.01	—	—	28	30
	0.05	—	—	15	16		0.05	26	28	31	33
	0.10	—	14	16	17		0.10	27	30	32	35

Note: For values of n not in the table calculate the 100(1 − β)% upper one-tailed confidence interval — $(1.5(x/n) - 0.5) + (1.5)z_\beta\sqrt{(nx - x^2)/n^3}$, where x is the number of correct answers from the study, n is the number of respondents, and z_β is the upper-β critical value of a standard normal deviate. It may be concluded with 100(1 − β)% confidence that the true proportion of distinguishers in the population is not greater than the calculated value. To find z_β use the last row of Table T4, substituting α for β.

TABLE T12
Duo-Trio Test for Similarity or Two-Sided Paired Comparison Test for Similarity: Critical Number (Maximum) of Correct Answers

Accept the null hypothesis of no difference with 100(1 − β)% confidence if the number of correct responses is less than [illegible] in the table that corresponds to the specified values of n, β, and p_d, where p_d is the proportion of the population that can distinguish the samples

		p_d						p_d			
n	β	0.15	0.20	0.25	0.30	n	β	0.15	0.20	0.25	0.30
24	0.001	—	—	—	—	56	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	—
	0.05	—	—	—	—		0.05	—	—	28	29
	0.10	—	—	—	12		0.10	—	28	29	31
28	0.001	—	—	—	—	60	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	—
	0.05	—	—	—	—		0.05	—	—	30	32
	0.10	—	—	—	14		0.10	—	30	32	33
32	0.001	—	—	—	—	64	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	32
	0.05	—	—	—	—		0.05	—	—	33	34
	0.10	—	—	—	16		0.10	—	32	34	36
36	0.001	—	—	—	—	68	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	34
	0.05	—	—	—	18		0.05	—	—	35	37
	0.10	—	—	18	19		0.10	—	35	36	38
40	0.001	—	—	—	—	72	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	36
	0.05	—	—	—	20		0.05	—	—	37	39
	0.10	—	—	20	21		0.10	—	37	39	41
44	0.001	—	—	—	—	76	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	39
	0.05	—	—	—	22		0.05	—	38	40	41
	0.10	—	—	22	24		0.10	—	39	41	43
48	0.001	—	—	—	—	80	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	41
	0.05	—	—	—	25		0.05	—	40	42	44
	0.10	—	—	25	26		0.10	—	41	43	46
52	0.001	—	—	—	—	84	0.001	—	—	—	—
	0.01	—	—	—	—		0.01	—	—	—	43
	0.05	—	—	26	27		0.05	—	42	44	46
	0.10	—	26	27	28		0.10	—	44	46	48

Note. For values of n not in the table calculate the 100(1 − β)% upper one-tailed confidence interval — $(2(x/n) - 1) + (2)z_\beta\sqrt{(nx - x^2)/n^3}$, where x is the number of correct answers from the study, n is the number of respondents, and z_β is the upper-β critical value of a standard normal deviate. It may be concluded with 100(1 − β)% confidence that the true proportion of distinguishers in the population is not greater than the calculated value. To find z_β use the last row of Table T4, substituting α for β.

LABORATORIO No. 3

ANALISIS DESCRIPTIVO

OBJETIVO

Demostrar las etapas incluidas en el análisis descriptivo de un producto alimentario: selección y entrenamiento de panelistas, desarrollo de un listado de atributos a ser evaluados, desarrollo de escalas apropiadas para medir esos atributos y evaluación de las muestras utilizando esas escalas.

MATERIALES Y EQUIPO

- Muestras producto
- Estándares para definir términos
- Boletas

PROCEDIMIENTO

1. El instructor dará información general sobre el producto a evaluarse

2. El estudiante evaluará las muestras proveídas e identificará términos que caractericen el producto en cuanto a apariencia, sabor y textura. Un término de apariencia, 2 de sabor y 3 de texturaa serán suficientes en esta práctica.

3. Cuatro subgrupos serán formados. Un miembro de cada subgrupo recogerá la información generada por los 5 miembros y hará un listado, tratando de eliminar términos que sean redundantes.

4. Una discusión general será llevada a cabo, donde el encargado de la práctica:

   a. Pedirá a los líderes de los subgrupos el listado de los atributos sugeridos para ordenarlos en secuencia lógica en la pizarra.
   b. Discutirá con los participantes, qué términos son independientes y cuáles son redundantes (que describen la misma característica).
   c. Obtendrá un consenso general de los términos descriptivos desarrollados para este grupo de productos.

5. El grupo y el instructor decidirán qué parámetros deben medirse en las escalas lineales no estructuradas y seleccionarán los términos que indiquen el inicio y final de cada escala. Adicionalmente se podrán utilizar escalas con el método "Spectrum", donde el evaluador escribe el número que indica la intensidad

percibida.

6. Ejemplo de productos o materiales a usarse como estádares para ejemplificar los puntos finales de algunas escalas, serán presentadas y definiciones de cada término serán dadas. (En situación real, los panelistas deben volver a evaluar las muestras presentadas auxiliándose de los estánderes para comprender el significado de los términos.

7. Cada estudiante evaluará las muestras y describirá la intensidad en las escalas desarrolladas en el punto 5, utilizando la boleta diseñada para ello. Discusiones de grupo serán llevadas a cabo para definir términos, eliminar términos redundantes, y lograr más homogeneidad entre los panelistas.

8. Medición de las escalas lineales para conocer la intensidad de cada uno de los atributos. Tabulación de datos.

9. Análisis de datos utilizando estadística descriptiva, diagramas de estrella y análisis de varianza.

CONCLUSION

El método de análisis descriptivo utiliza panelistas entrenados que tengan cierta experiencia con el producto que se evalúa. Se requiere que el grupo de panelistas sea capaz de seleccionar y definir aquellos términos que son importantes en la caracterización completa del producto. El grupo debe construir las escalas que presentarán un rango de intensidad para cada uno de los atributos a evaluarse y debe ser capaz de usar las escalas para evaluar consistentemente las características en el rango de productos descrito.

En esta práctica se hará descripción breve del uso de este método. Descripción en detalle así como uso de diferentes productos y análisis estadístico más sofisticado se darán en un curso de Análisis Sensorial Avanzado.

DESCRIPCION DE TERMINOS

Por favor escriba 10 palabras que usted crea puedan describir las características de apariencia, aromáticas, sabor y textura de las muestras. Evite el uso de términos afectivos como "bueno", "malo", "agradable","desagradable", etc. o términos que describan el producto tal como "sabor a...".

Utilice la boleta que se le presenta a continuación. Escriba los términos en la columna de la izquierda y evalúe la presencia de ellos en c/u de las muestras, haciendo una marca (X) en la línea que corresponda. Todavía no evalúe intensidad.

ANALISIS DESCRIPTIVO

Panelista No.________________

Producto__________________

CARACTERISTICAS	CODIGOS DE MUESTRAS	
APARIENCIA		
AROMATICOS		
GUSTOS BASICOS		
TEXTURA		
SENSACIONES BUCALES		
SABOR RESIDUAL		

PRUEBA DE PREFERENCIA

Por favor ordene las muestras de acuerdo a preferencia. 1 La más aceptada y la 2 la menos aceptada.

1. ______________

2. ______________

EVALUACION DE INTENSIDAD

INSTRUCCIONES

1. Despúes de conocer los puntos finales de las escalas, evalúe la intensidad que las muestras tienen para cada una de las diferentes características.

2. Evalúe las muestras en el orden asignado, escalas lineales no estructuradas para cada característica escribiendo sobre la escala pequeñas líneas verticales en el punto que mejor describa la intensidad percibida para cada muestra. Escriba el código de la muestra sobre cada línea vertical.

3. Método Spectrum

   Utilizando la boleta con los abributos definidos, escriba los números (del 0 al 15) que mejor describan la intensidad percibida en cada una de las muestras.

ORDEN DE EVALUACION ______ ______ ______ ______

TABLA PARA TABULACION DE DATOS
ANALISIS DESCRIPTIVO

Panelista No.	MUESTRAS			
	______	______	______	______
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
PROMEDIO DE PUNTAJES				

Journal of Food Protection, Vol. 47, No. 5, Pages 403-409 (May 1984)

Applications of Descriptive Analysis

M. GILLETTE

McCormick & Co., Inc., Corporate Research & Development Laboratories, 202 Wight Avenue, Hunt Valley, Maryland 21031

(Received for publication October 11, 1983)

ABSTRACT

Descriptive Analysis is a method of sensory evaluation that identifies, describes and quantifies the sensory attributes of a product. Descriptive Analysis is a valuable tool for providing information on appearance, aroma, flavor and/or texture of food products, and is used effectively for product and process development, shelf life studies, product improvement, quality assurance and control, and sensory-objective correlations in the food and flavor industry. Each application is discussed with examples.

The "Descriptive Analysis" technique of sensory evaluation identifies, describes, and quantitates sensory (visual, textural, auditory, olfactory, and gustatory) qualities of a given product. This technique requires a panel of 5-10 trained persons who are thoroughly familiar with the product's sensory characteristics and who can accurately and precisely communicate their perceptions. There are a variety of commonly used forms of descriptive analysis, including the "Flavor Profile" method of Arthur D. Little Inc. (*1*) and the "Quantitative Descriptive Analysis" method of Tragon Corporation (*11,12*). Many sensory groups devise their own customized methods of descriptive analysis, but all methods yield basically the same type of information.

If collected appropriately, data from descriptive analysis panels can be analyzed statistically by use of common statistical methods such as the t-test and analysis of variance. Data are often presented graphically in a variety of forms ranging from histograms to circular graphs. The circular graph technique (Fig. 1) provides an immediate "picture" of how a product tastes and, therefore, is very useful in communicating the results of a descriptive panel. To interpret the graph, realize that the center point represents zero on the descriptive scale, and points radiating outward from the center are increasing in magnitude. On this particular graph (Fig. 1), the test product A-5 is significantly stronger than the target in oily and toasted aromas, and significantly weaker in total flavor strength. The flavor graph clearly illustrates that the two products do not smell or taste the same.

It is important to note that not all differences on the flavor profile are statistically significant (i.e., flour/

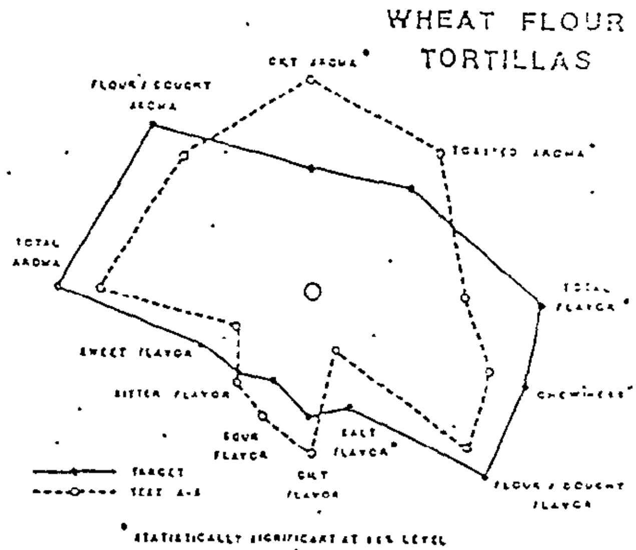


Figure 1. *Aroma and flavor profiles for target and test wheat flour tortillas.*

doughy aroma and flavor), and that the relative importance of each attribute may be different. For example, a statistically significant difference between the target and test A-5 (Fig. 1) in sweetness would be considered less severe of a problem than a statistically significant difference in bitterness.

Descriptive Analysis is most often used as a technical tool to aid in development or improvement of a product, as well as to delineate problem areas in shelf-life. It is very useful in helping to understand the sensory qualities of a product, but it is not the appropriate test to be used when preference or acceptability judgements are required. The descriptive analysis technique can be used most satisfactorily, however, in conjunction with hedonic tests to explain affective results. In the food and flavor industry, Descriptive Analysis can be applied to the following seven activities.

PRODUCT DEVELOPMENT

Descriptive Analysis provides a wealth of vital information to the food technologist or flavorist. Initially, the descriptive panel provides a description of the selected target (Fig. 2), and/or competitive products (Fig. 3). This

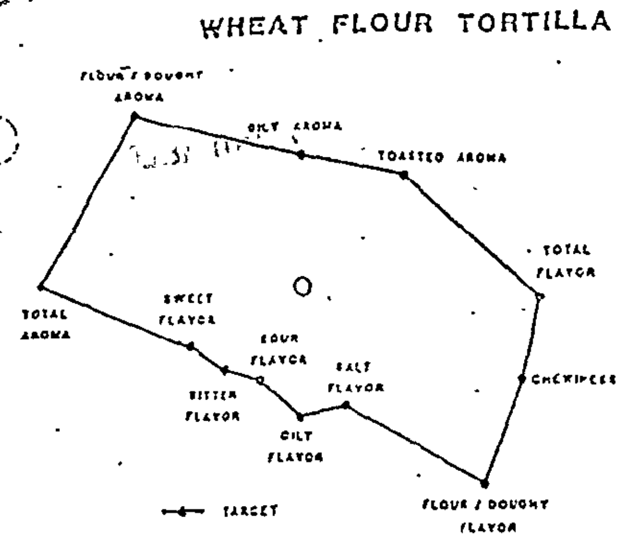


Figure 2. *Aroma and flavor profile for target wheat flour tortilla.*

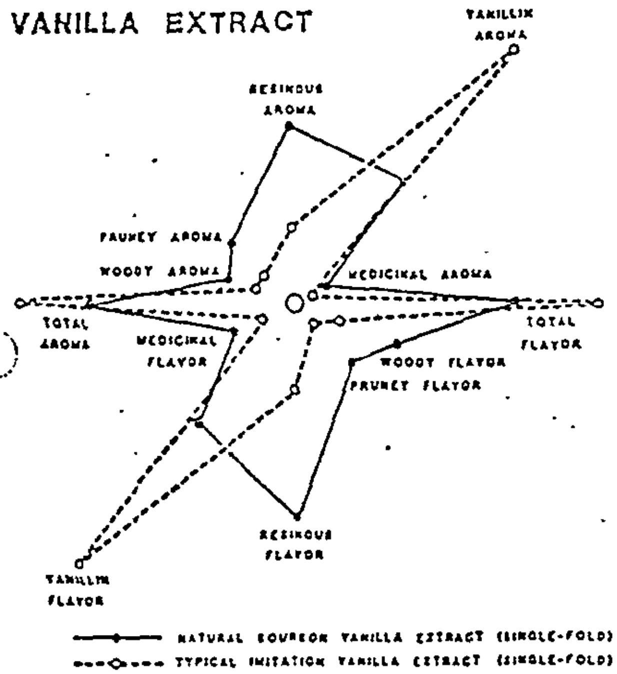


Figure 3. *Aroma and flavor profiles for natural bourbon vanilla extract and for a typical imitation vanilla extract.*

information provides the technologist with the necessary direction he/she needs to take to match the sensory qualities of the target. As product development progresses, the descriptive panel provides continual guidance to the technologist, helping to aim formulations in the desired direction. The objective assessment provided by the descriptive panel is particularly important during this phase of product development, because the food technologist or flavorist becomes easily fatigued or biased when evaluating his/her own "creation." The target profile provided by the descriptive panel assures a more rigorous product-development effort, as a predefined target has to be successfully matched.

Finally, when the product development objective has been achieved and pilot-plant scale-ups are underway, the descriptive panel is used to confirm that no changes in

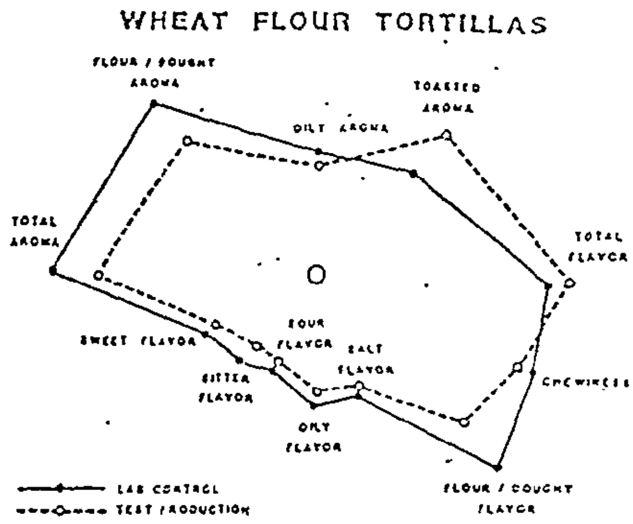


Figure 4. *Aroma and flavor profiles for laboratory and production wheat flour tortillas.*

sensory quality have been caused by the scale-up (Fig. 4).

SHELF-LIFE

New and reformulated products often need to be tested to determine their estimated shelf-life. This testing generally involves microbiological and sensory analyses. The descriptive panel provides the food technologist with detailed descriptions on how the color, aroma, flavor, and/or texture of the product change over time. If only a simple difference test (i.e., triangle or duo-trio test) were unwisely used, a product might fail shelf-life testing when it was actually *improving* with age (i.e., wine or vanilla extract) Similary, a descriptive panel is uniquely suited to point out the deteriorative changes or undesirable sensory qualities that might (Fig. 5) or might not (Fig. 6) develop with product aging. While chemical tests are

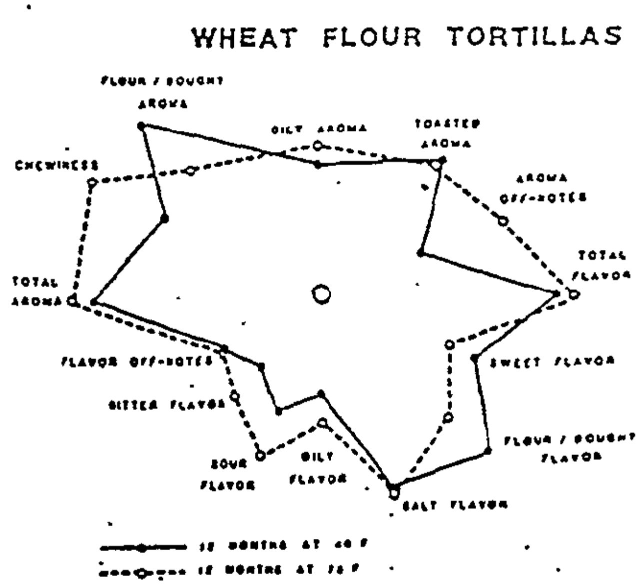


Figure 5. *Aroma and flavor profiles for wheat flour tortillas stored for 12 months at 40 or 75°F.*

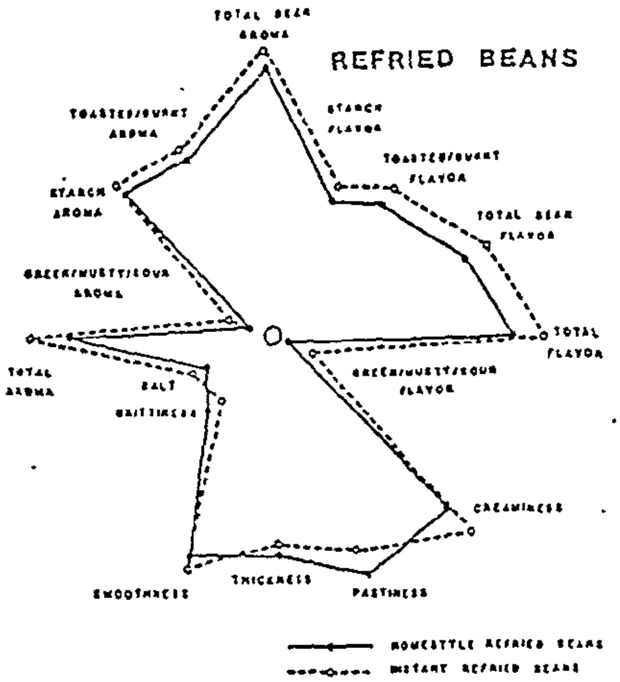


Figure 10. *Aroma and flavor profiles for homestyle refried beans and for an instant refried bean mix.*

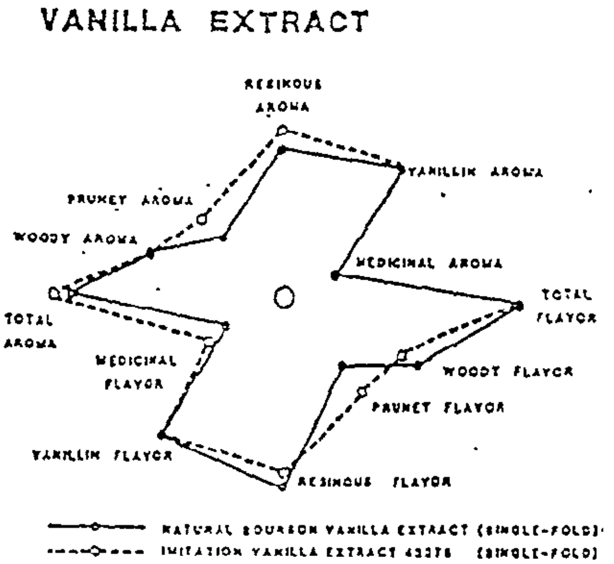


Figure 11. *Aroma and flavor profiles for natural and imitation vanilla extracts.*

scriptive panel can provide an accurate assessment regarding whether a product really has "improved" in one or more qualities. Does the "new improved" laundry detergent significantly increase the softness and color of laundered items? Does the new imitation vanilla extract smell and taste exactly like the natural vanilla extract (Fig. 11)?

In a similar vein, there is a significant amount of interest in reducing calories, fat, sodium, and sugar in food products today. Difference (i.e., triangle, duo-trio, paired comparison) testing will indicate whether or not the "reduced" product is different from the original product, but often it is virtually impossible to reduce product constituents significantly without altering flavor, texture, or appearance. Differences in flavor between an original product and a "reduced sodium" formulation may be perfectly acceptable as long as the reduced sodium product maintains a good flavor/texture balance with no undesirable characteristics. Again, to accurately test for such integrity, an acceptance panel is required. However, during the reformulations, excellent and cost-conscious guidance can be provided through a trained descriptive panel.

Additionally, the descriptive panel can help explain the results of a consumer tests. For example, if a consumer test indicated that a reduced fat product was "too thin," a trained descriptive panel could provide insight to the product development team regarding what "too thin" meant. The original and the reduced fat products might be equally viscous, equally chewy, but the reduced fat product might leave less oily mouthcoating; giving the impression of " thinness."

QUALITY CONTROL

A descriptive panel is sometimes used for routine monitoring of production output. The descriptive panel can compare the production samples to quality reference standards and determine whether any noteworthy differences exist. The major benefit of using a descriptive panel for quality control is the "committee decision" that the panel can offer. The expert panel can offer well-trained advice on whether or not to reject a questionable production batch based upon sensory qualities alone. For quality control work, a full-scale descriptive panel of 10 members is not always necessary; for routine monitoring, a smaller-scale panel of 4–5 experts may be adequate.

Sensory quality control can also be monitored by routine difference testing. When a significant difference is found, the descriptive panel can be used to define the exact nature of the differences.

When a triangle panel found a statistically significant difference between a production lot of dehydrated onions and the quality reference standard, a trained descriptive panel was able to reveal that the production sample was slightly too "toasted," in fact it was slightly burnt (Fig. 12)! The panel recommended rejection of the lot.

QUALITY ASSURANCE

Defining "quality" is always a difficult task, but a descriptive panel can help in this regard. If for instance, a "high quality" pair of blue jeans needed to be defined, a descriptive panel can review the various forms, types, styles, etc. of blue jeans (ignoring cost) and reach a consensus on what exactly constitutes a "quality" pair of jeans. This process can, of course, be applied equally well to food products. Figures 13, 14 and 15 illustrate quality standards chosen for onion, black pepper, and vanilla extract.

Once "quality" has been defined, the panel can select reference standards for future training and testing. The panel can monitor incoming raw materials and determine whether or not they meet the quality specifications for

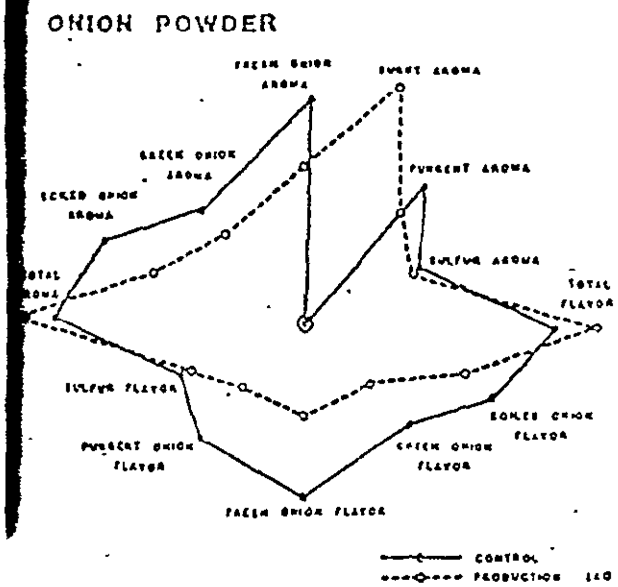


Figure 12. *Aroma and flavor profiles for the control and production samples of onion powder.*

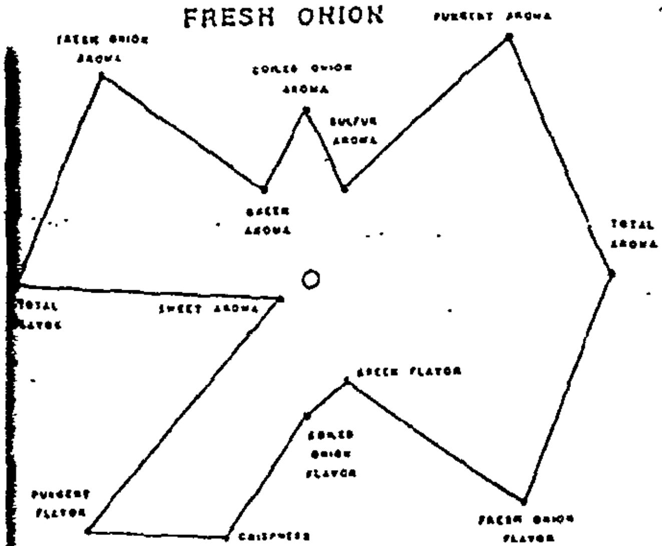


Figure 13. *Aroma and flavor profile for a fresh spanish onion.*

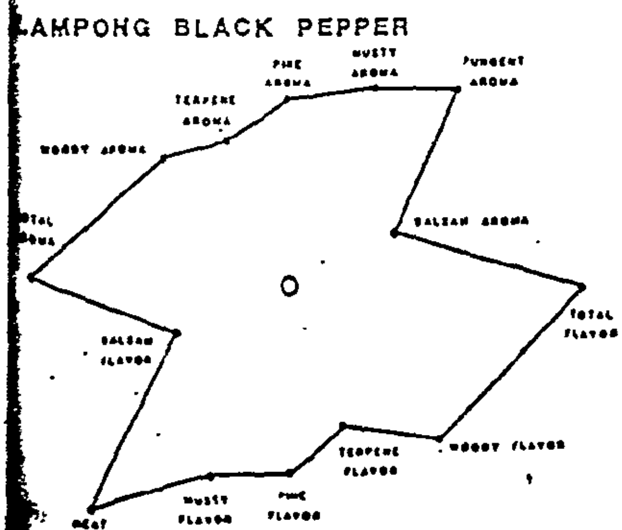


Figure 14. *Aroma and flavor profile for Lampong black pepper.*

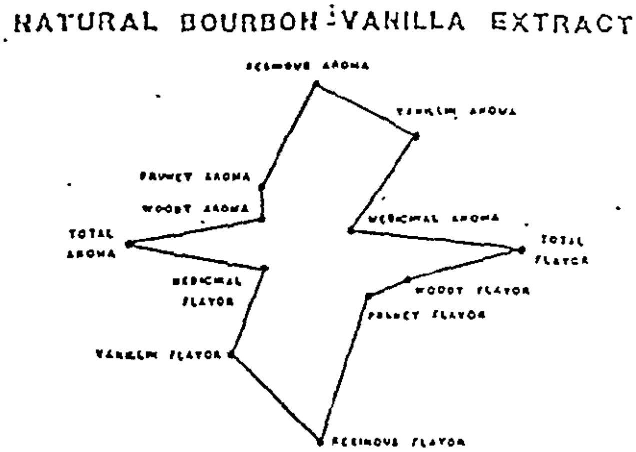


Figure 15. *Aroma and flavor profile for natural bourbon vanilla extract.*

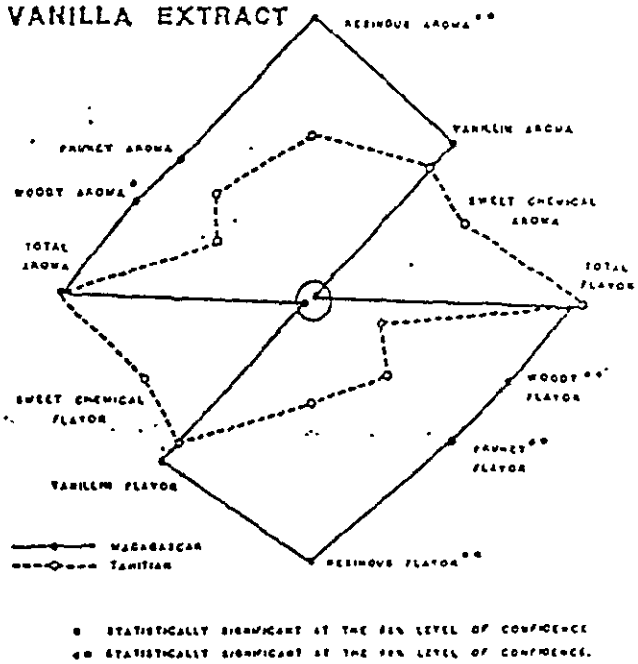


Figure 16. *Aroma and flavor profiles for Madagascar and Tahitian vanilla extracts.*

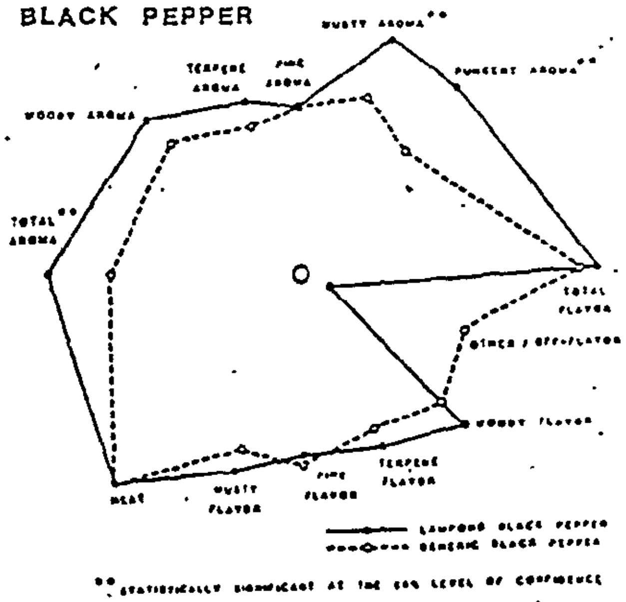


Figure 17. *Aroma and flavor profiles for Lampong black pepper and for a generic label black pepper.*

sensory characteristics (Fig. 16). A descriptive panel can also be used to monitor the quality of competitive products (Fig. 17). For routine monitoring, a small-scale descriptive panel of 4-6 members may be adequate and more desirable. The role of sensory evaluation in product quality assurance was reviewed at the annual meeting of the Institute of Food Technologists. (*2*,*7*,*10*,*16*)

SENSORY - INSTRUMENTAL CORRELATIONS

The use of instrumental methods for supplementing or replacing sensory methods has proven a valuable practice. Sensory panels are expensive and not always available, and panelists can become fatigued with routine testing. A wide variety of objective methods, including gas and liquid chromatographs, texturometers, and chemical assays have been used to compliment sensory panels.

To calibrate the instrument, strong correlations need to be determined between the sensory and the objective methods. To achieve strong correlations, a panel trained to be precise and accurate is necessary (*14*).

Very good results have been achieved in correlating sensory and instrumental methods for texture evaluation, using such instruments as the Instron to predict chewiness, cohesiveness, brittleness, etc. and the Brookfield viscometer to predict viscosity (*8*). Commendable results have also been obtained in correlating sensory aroma and flavor assessments to gas and liquid chromatographic data, color (i.e., Hunter Colorimeter), moisture, and volatile oil content (*4*,*6*,*8*).

Trained panels can also be used for providing a scaled attribute for maximization or minimization in Response Surface Methodology (*5*) or as a criterion for Discriminant Analysis (*9*). For example, based upon their chemical profiles, five varieties of an herb were classified by discriminant analysis (Fig. 18). A trained panel was then used to determine which classified groups had similar sensory properties (Fig. 19). New samples can then be classified, based upon their chemical profiles. If the sample falls into a region where flavor is acceptable, the new sample can be accepted; if not, it is rejected. This method provides a comprehensive and clear-cut basis for purchasing and quality assurance decisions.

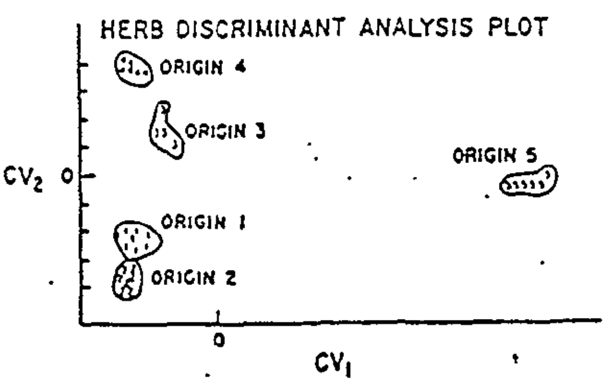


Figure 18. *Discriminant analysis plot classifying 5 origins of an herb based upon their chemical profiles.*

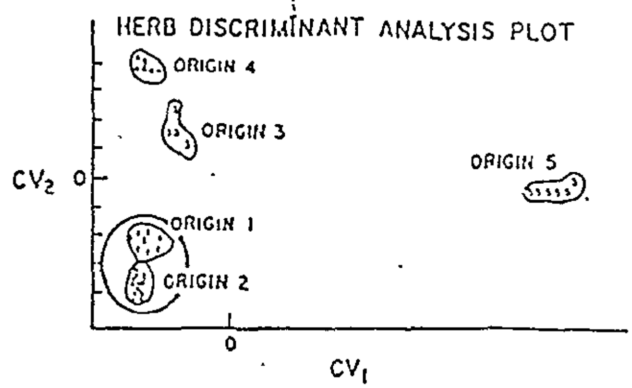


Figure 19. *Discriminant analysis plot classifying 5 origins of an herb; origins 1 and 2 taste the same.*

The sensory heat in red pepper can be predicted using the linear relationship between sensory heat and level of capsaicinoids (determined by high pressure liquid chromatography) in the red pepper (Fig. 20). This relationship can be used for purchasing, quality assurance, quality control, and research work.

Manual hardness can be a problem with tubed decorator icings. To monitor the hardness of filled tubes, an instrumental method was desired. A trained panel rated "squeezability" of tubes filled with decorator icings of different consistency. An Ottawa Texturometer also measured the consistency of these tubes. Results of the sensory and instrumental measurements were highly correlated, ($r=0.94$), so that "squeezability" can be predicted using the Texturometer (Fig. 21) (*15*).

Despite the many successes in correlating sensory and instrumental relationships, we must remind ourselves that only the human being can truly measure the full experience of flavor. No matter how precise and varied, instruments will never be able to fully quantitate the human experience.

OTHER BENEFITS

A descriptive panel activity has benefits other than providing sensory data. Participation on sensory panels provides panel members with greater understanding of the

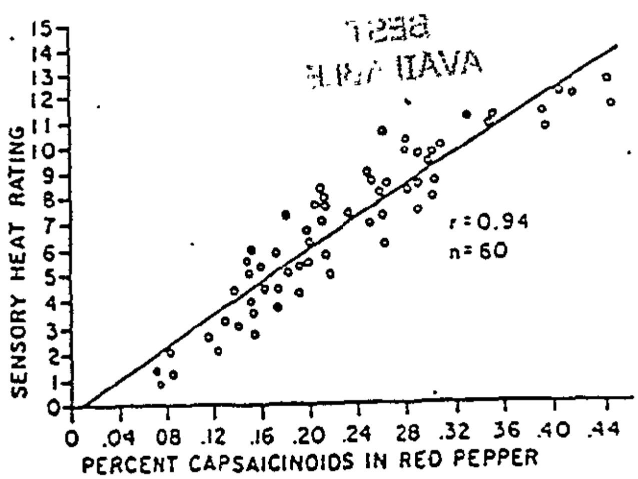


Figure 20. *Linear relationship between sensory heat rating and percent capsaicinoids in ground red pepper.*

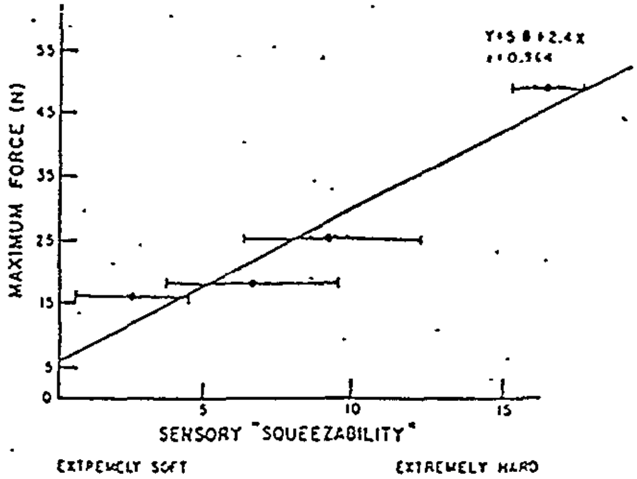


Figure 21. *Linear relationship between sensory "squeezability" and texturometer extrusion force (N) of tubed icings.*

company's business, of departmental projects, and of their own projects. Communications between food technologists and the sensory group, between technical and nontechnical personnel, as well as between customers and suppliers are improved. Trained panelists make better employees; not only do they better understand the company's business, but they also have a unique ability to provide an expert opinion and offer a valuable voice in decision making. Descriptive panels can be considered a form of participatory management, thus reaping many of the benefits of "quality circle" type institutions (*3*).

ACKNOWLEDGMENT

The author thanks Mr. Charles Prodey for his assistance in preparing the figures in this paper.

REFERENCES

1. Caul, J. F. 1957. The profile method of flavor analysis. Adv. Food Res. 7:1-40.
2. Dethmers, A. E. 1979. Utilizing sensory evaluation to determine product shelf life. Food Technol. 33(9):40-42.
3. Donnelly, J. F. 1971. Increasing productivity by involving people in their total job. Personnel Admin. 34(5):8-13.
4. Galetto, W. G., and A. A. Bednarczyk. 1975. Relative flavor contribution of individual volatile components of the oil of onion (*Allium cepa*). J. Food Sci. 40:1165-1167.
5. Giovanni, M. 1983. Response surface methodology and product optimization. Food Technol. 37:41-45.
6. Moore, L. J., and C. F. Shoemaker. 1981. Sensory textural properties of stabilized ice cream. J. Food Sci. 46:399-402, 409.
7. Nakayama, M., and C. Wessman. 1979. Application of sensory evaluation to the routine maintenance of product quality. Food Technol. 33(9):38-39, 44.
8. Noble, A. C. 1975. Instrumental analysis of the sensory properties of food. Food Technol. 29(12):56-60.
9. Powers, J. J., and E. S. Keith. 1968. Stepwise discriminant analysis of gas chromatographic data as an aid in classifying the flavor quality of foods. J. Food Sci. 33:207-213.
10. Reece, R. N. 1979. A quality assurance perspective of sensory evaluation. Food Technol. 33(9):37.
11. Stone, H., J. L. Sidel, and J. Bloomquist. 1980. Quantitative descriptive analysis. Cereal Foods World 25(10):642-644.
12. Stone, H., J. Sidel, S. Oliver, A. Woolsey, and R. C. Singleton. 1974. Sensory evaluation by quantitative descriptive analysis. Food Technol. 28(11):24-34.
13. Thomas, S. M., F. Caporaso, G. A. Miller, and R. M. Kiral. 1981. Importance of sensory evaluation in the stability testing of a medical food. J. Food Sci. 46:435-439.
14. Trant, A. S., R. M. Pangborn, and A. C. Little. 1981. Potential fallacy of correlating hedonic responses with physical and chemical measurements. J. Food Sci. 46:583-588.
15. Voisey, P. W., M. Klock, K. Summers, and M. Gillette. 1979. A method for testing the ease of extrusion of icing marketed in plastic tubes. J. Texture Studies 10:435-448.
16. Wolfe, K. A. 1979. Use of reference standards for sensory evaluation of product quality. Food Technol. 33(9):43-44.

THE SELECTION AND USE OF JUDGES FOR DESCRIPTIVE PANELS

BEST AVAILABLE

KATHERINE ZOOK and COLLEEN WESSMAN

SENSORY DESCRIPTIVE DATA from experienced individuals have been used as an adjunct to laboratory taste tests at Quaker Oats for some period of time. For the past 2½-3 years, however, a more tightly controlled method of handling descriptive sensory data, the Quantitative Descriptive Analysis (QDA), has been continuously in use. During that time, ten major descriptive panels have been developed to meet different objectives. All of these groups have had two things in common: They have been selected by a series of triangle difference tests, and they have received an intensive period of training before beginning their descriptive work.

The purpose of this article is to examine this method of selecting and training as a means of producing good descriptive judges and to look at some of the types of questions which have been answered with QDA data from such judges.

BASICS OF THE QDA METHOD

The QDA method of descriptive analysis was described in detail by its developers (Stone et al., 1974) several months after it was implemented by them at Quaker Oats. Early usage with beer at the Schlitz Beer Co. was described at the same time by Macredy et al. (1974).

The method as used at Quaker Oats followed this sequence of events:

- Screening of 24-36 prospective judges by means of 12 triangle difference tests, each administered twice.
- Selection of 10-12 of the most discriminating judges. Availability for the time involved in training was a prime requisite for judge use.
- Training for around 10 hr, usually 1 hr/day. During this period, the terminology was developed to describe the appearance, flavor, and texture of the products. Judges were provided with a broad assortment of training products as they modified and perfected the evaluation sheets.
- A Series of 4 replicated judgments on training products. After grading and statistical analysis, one or more correction sessions were conducted to clarify any confusion in the use of terms. The basic unit of evaluation in the QDA method is an unstructured line 6 in long anchored ½ in from either end by pairs of terms (Fig. 1). The judge evaluates the intensity of each sensory attribute by placing a vertical line across the unstructured line. This is translated into a score from 0 to 60 for statistical analysis. All data reported in this article are in terms of this scale, which is unstructured when the judge uses it and is assigned numerical values later.
- Replicated Judgments, using 10-12 replications on the first use of the panel immediately after training. With repeated usage of the group, 4-6 replications with 8 judges have been found to produce reliable data.
- Analysis of Variance of each sensory attribute separately (see Table 1), together with development of correlations between attributes and between attributes and an overall acceptance term.
- Examination and Interpretation of data.
- Preparation of QDA Configurations describing products, and visual or written presentation. A typical QDA configuration (Fig. 2) has lines radiating outward from a center point. Each line represents a particular descriptive term, and the average intensity for that term is plotted on that line (the center represents an intensity of 0, and the outer point a value of 35-60). Connecting the average intensities for all the terms provides a product profile.

Judges who made up the trained QDA groups were all employees of the Quaker Oats Research and Development Laboratory. A total of 279 screenings were conducted, from which 110 judges were selected for membership in the groups. About half of the judges were in food-related jobs where they might have experience in tasting food; the other half were in non food-related positions, but were found to be able to discriminate well by the triangle testing procedure. All of the groups developed were mixtures of the two types, as shown in Table 2.

PRACTICAL CONSIDERATIONS

There are a number of considerations in selecting and using judges for QDA, or any other type of continuing descriptive work, which have nothing to do with the technical aspects of selecting a good judge but are critical to the success of the program:

1. First, and most important, the whole program must have upper management support.

2. This must be communicated to middle management personnel who will be giving permission to release employees for the time required for the training and development of the configurations. It may be necessary to hold brief seminars or educational meetings to acquaint these people with examples of what QDA can do and how it is accomplished.

3. The time requirements must be clearly spelled out to both the participant and his immediate superior before participation begins. Securing the panelist for one hour per day during the training period is the crucial factor, the screening and repeated evaluations requiring much less time.

4. If two people from a small department both qualify, it is wiser to select only one for training. Thus, when any one group meets, participation is spread among a number of departments and no one department is suddenly denuded of a large part of its work force.

5. After management support, motivation and interest on the part of the panel member is the single most important factor in obtaining on-time, conscientious presence at all panel sessions. This is enhanced by giving the panel some idea of the purpose or importance of the project to the company. There is a very fine line between providing the panel with information on the overall aims of the work and giving clues which may bias their descriptive work.

6. As a morale booster, luncheon together at a restaurant of their choice for the group after the completion of a set or series of samples has proved to be effective not only as a reward but as a means of building "esprit de corps."

7. Occasionally, one person will qualify and serve on more than one group, but this has not been encouraged because of the potential conflicting demands of the two groups.

TECHNICAL ABILITIES NEEDED

The qualifications for expert descriptive judges have been thoroughly described in various papers. Some of the more important are:

The authors are with the Product Evaluation Dept., The Quaker Oats Research Laboratory, Barrington, IL 60010

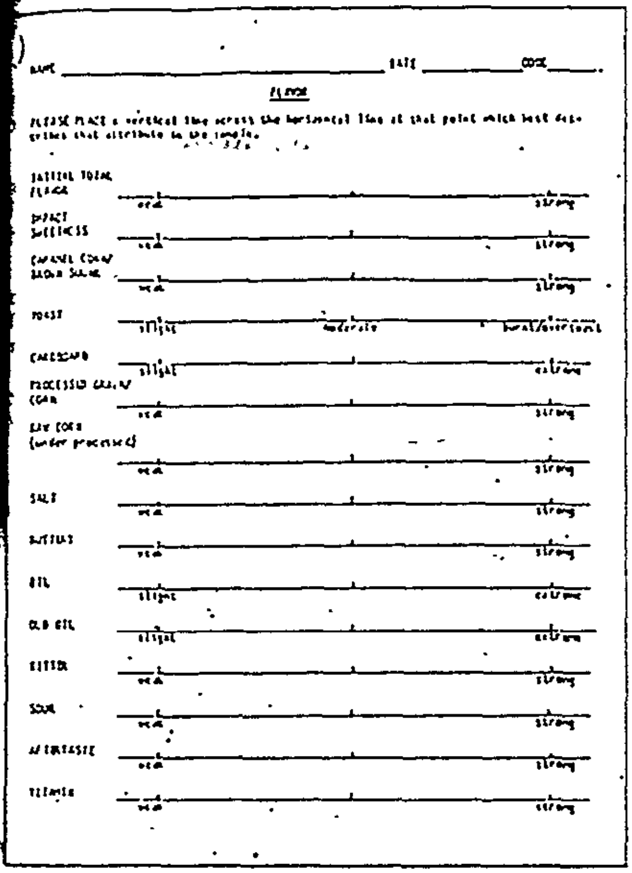


Fig. 1—SCORE SHEET *for quantitative descriptive analysis; the judge evaluates the intensity of each sensory attribute by placing a vertical line across the unstructured line. This intensity is later translated into a numerical score for statistical analysis by use of a template which divides the line into values ranging from 0 at the left end to 60 at the right end*

- Taste Acuity or ability to duplicate a difference judgment (Amerine et al., 1965). Normal gustatory and olfactory ability is presupposed.
- Ability to Deal Analytically with a complex test situation. Girardot et al. (1951) believed that it was impossible to test independently for all the factors underlying this unitary skill but that the test situation must be set up to require acts of discrimination and judgment, such as would be used later in the experiments. (These skills include flavor memory, ability to deal logically with flavor perceptions, and general adjustment to the test situation).
- Good Health and freedom from allergy, frequent head colds, and sickness (Amerine et al., 1965).
- Stable Personality, neither overly passive nor overly dominant (Caul, 1957). From some recent studies with personality trait scaling (Henderson and Vaisey, 1970), it has been suggested that individuals would continue to operate at a higher level of performance if they score high in the "need for achievement." Our observation is that an individual who is not easily distracted from the task at hand, who can perform no matter what the emotional upsets in his life, usually makes an excellent judge.
- Ability to Verbalize or describe what they taste.
- Interest and Motivation.
- Availability.

SCREENING BY TRIANGLE TEST

The triangle test is not a new test, nor is its use as a means of selecting judges new. This difference test, first suggested by Bengtsson (1943), was used by Helm and [illegible] beer-tasters.

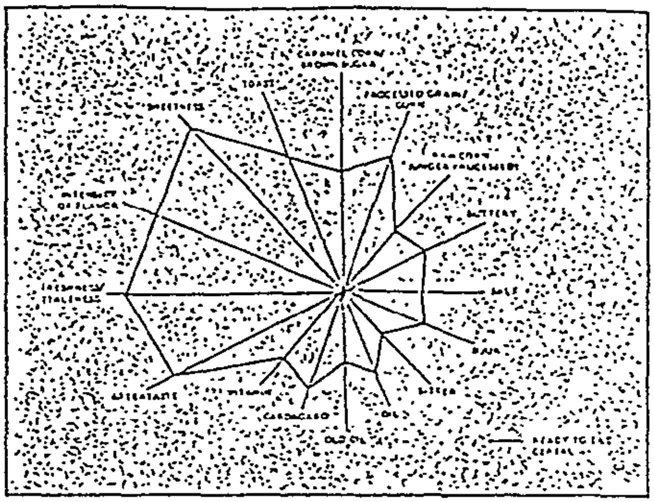


Fig. 2—A TYPICAL QDA CONFIGURATION *for the flavor of a ready-to-eat cereal, dry. The average intensities for the various attributes are graphed on lines radiating outward from a value of 0 at the center point to a value of 35 at the outer perimeter. For example, the average intensity of sweetness is located at a point representing 32.14*

Table 1—AVERAGE INTENSITIES *of various attributes for the flavor of a single sample of Cap'n Crunch® ready-to-eat cereal, dry*

Flavor attribute	Average intensity[a]
Intensity of flavor	29.69
Sweetness	32.14
Toast	19.55
Caramel corn/brown sugar	18.28
Processed grain/corn	20.05
Raw corn (underprocessed)	10.88
Buttery	12.60
Salt	11.99
Sour	12.05
Bitter	7.22
Oil	11.07
Old oil	9.26
Cardboard	14.36
Vitamin	12.06
Aftertaste	27.68
Fresh/stale	32.05

[a] 11 judges × 10 replications = 110 evaluations, intensities are from analysis of variance of panel as a whole, analysis is also made on the data of each individual judge, to monitor judge performance.

Table 2—MAKEUP OF QDA DESCRIPTIVE PANELS

Product evaluated	No. of panelists: Food-related personnel	Non-food-related personnel	Total
Breakfast item	8	2	10
Natural Cereal®	6	6	11
Syrup	7	4	11
Pizza	3	8	11
Life® cereal	6	5	11
Shredded Wheat	3	7	10
Cap'n Crunch® cereal	5	6	11
Frozen pancakes	7	5	12
Oatmeal cookies	3	9	12
Instant grits	8	3	11
Total	55 (50%)	55 (50%)	110 (100%)

Peryam and Swartz (1950) reported that the triangle test had been developed independently in the laboratories of Joseph E. Seagram and Sons in 1941. Amerine et al. (1965) states that "because of its extensive application the test has been the most thoroughly studied and criticized of all test designs."

There are several advantages in screening with triangle tests:

1. It is obvious that those who do best on the series of iangle tests have the underlying abilities to make flavor d texture judgments on that particular product.

2. The results of the triangle tests are *objective*. Panelists now that if they are selected for training, they have a milar level of discriminatory ability to the remainder of he panel.

3. Panel members work directly with the product under onsideration. Included in the triangle testing can be some f the same variables which will be present in the test roducts.

4. The preparation of the products for the triangle tests y the panel leader gives this individual a great deal of xperience in what the product and its variations are like.

5. By the time the participants have completed the riangle tests, they also have a background of tasting xperience with the product even before they are trained.

6. If special products are prepared for the triangle tests uch as a special plant run of undercooked or overcooked oduct), they also make excellent training products for ter use.

However, screening by triangle test also has several sadvantages:

1. It does not tell you anything about the ability of the ndidate to verbalize what he tastes or what kind of a ersonality he will have for group interaction.

2. The triangle tests themselves must be quite carefully repared. If you have set up a test set to measure a certain avor difference, you do not want the judge to make a correct pairing from some appearance difference inadvertently present in the samples. Masking of appearance by educed or minimal lighting is very often necessary. Colored lights which may mask color often leave other clues which the participant can pick up.

3. The general plan of arranging the tests is to give the easier ones first and the more difficult ones later in the series. It is very difficult to get a perfect progression from easy to difficult. However, even without this, the judge gets exposure to a broad range of tasks.

4. To be fair to the candidates, tests must be given twice at the same sitting; once with a set containing two A samples and once with a set containing two B samples. This doubles the number of test sets used for screening.

RESULTS OF SCREENING

The samples for the 12 triangle tests were all pre-screened before the tests were administered, and an attempt was made to start with fairly easy tests and make them progressively more difficult. The aim was to expose the judges to a broad range of discriminatory tasks and to include in the series some of the types of variables which would later be described by the panel. Tests were administered twice, once with two A samples and once with two B samples, because it was felt that certain combinations might be easier to pair correctly than others. Table 3 shows a sample set of 12 triangle tests as administered for Cap'n Crunch ready-to-eat cereal, from which a group of 11 of the most discriminating judges were selected for training.

An overview of the 10 different groups screened by triangle tests shows that:

- Not all people are equally discriminating on all products. Thus, it proved worthwhile to test a significant percentage of candidates on more than one product type (Table 4).

- Some products were more difficult to discriminate with (pair correctly) than others. This is shown in Table 5, which

able 3—RESULTS OF SCREENING by triangle difference test, using variations of Cap'n Crunch ready-to-eat cereal, dry, when tested using two A amples and two B samples (X indicates an incorrect identification)

Judge	Day 1 Control vs Low sweetness with 2As	2Bs	Day 2 Control vs High sweetness with 2As	2Bs	Day 3 Control vs No salt with 2As	2Bs	Day 4 Normal vs Undercook with 2As	2Bs	Day 5 Control vs No oil with 2As	2Bs	Day 6 Normal vs Underdried with 2As	2Bs	Day 7 Normal vs Overdried with 2As	2Bs	Day 8 Control vs High salt with 2As	2Bs	Day 9 Normal vs Process change with 2As	2Bs	Day 10 Plant 1 vs Plant 2 with 2As	2Bs	Day 11 Lab product vs Plant 1 with 2As	2Bs	Day 12 Lab product vs Plant 2 with 2As	2Bs	Total incorrect out of 24 for each judge
													X		X		X	X			X		X		11
1				X	X	X			X		X		X	X											3
2						X							X	X		X				X					10
3				X	X	X			X	X	X														
4			X	X					X	X			X		X			X							7
5	X			X	X				X		X		X	X						X					8
6			X	X						X			X	X	X	X		X	X						9
7	X		X	X		X				X				X						X		X			8
8									X	X			X	X	X	X	X	X	X	X			X	X	12
9	X		X	X		X				X			X	X	X		X		X	X	X		X		13
10		X								X	X	X	X	X	X				X	X	X	X			10
11			X		X	X				X		X	X		X				X		X		X	X	11
12									X	X					X	X									4
13				X		X			X							X									4
14		X			X	X							X	X										X	6
15			X						X					X	X	X									5
16	X		X	X	X	X					X		X	X	X	X		X	X	X	X		X		15
17		X	X	X					X	X					X	X	X	X	X	X	X		X		13
18	X		X	X					X					X					X	X	X		X		9
19		X	X			X			X					X		X									6
20	X	X		X	X									X	X	X		X		X					9
21	X				X	X				X	X				X		X				X			X	9
Total incorrect for total panel of 21 judges	7	5	10	12	8	11	0	0	11	11	6	2	12	14	13	10	5	7	8	10	8	2	7	4	

ives the record of the 10-12 most discriminating members who were chosen for training on each product type. For 7 out of the 10 products, panelists averaged better than 75% correct identifications; for 2 products, the level was about 70%; for one, slightly lower.

• The level of discrimination attained by the groups tested appeared to vary within: (a) The complexity of the product—pizza contained many ingredients which often masked the real variable. (b) The variability of the product—shredded wheat biscuits showed unavoidable differences between individual units. (c) The difficulty of the test sets—although the same general plan was set up for all products, the tests on Life cereal were based on small variations in plant-made product which proved extremely difficult to identify.

• The level of discrimination attained for *all* individuals on *all* products in 279 screenings (Table 6) showed that: (a) 28.7% of those tested were very discriminating, picking the odd sample 75% of the time. Another 24.7% picked the odd sample 66.7% of the time. (b) 36.6% of those tested discriminated at a level in which they identified 50-62% of samples correctly. (c) Around 10% of the group did little better at discriminating than the level which could be attained by guessing on a triangle test (33⅓% correct).

Thus, generally speaking, if a level of discrimination of 67-75% correct identifications is desired, about 2-3 times as many candidates would have to be screened as would be selected for final training.

TRAINING OF JUDGES IMPORTANT

The importance of the training which is conducted with the judges after they are selected cannot be emphasized too much. For the training to be successful, an experienced and perceptive leader is necessary and several basic points must be observed:

• The panel leader does not take part in the descriptions of the products. His or her role is to keep the group functioning, provide standards and training samples as needed, prepare trial score sheets from the terms suggested, think of ways to clarify confusion, and test and monitor the judges.

• The terminology used to describe the products comes from the panel members themselves. All must understand and feel comfortable with the descriptive terms to be able to use them effectively in grading. This is why it is so important that members not miss sessions.

• Members of the group must feel on an even footing so that *all* will make contributions to the general pool of knowledge about the sensory characteristics of the product.

• The ingredients from which the product is made are not identified at first but only after a need for them is indicated, since they could possibly influence the expectations of the judges in their first impressions of the product.

• The physical surroundings of the group are important; these include adequate privacy, ventilation, and lighting, and a conference table around which all may sit to take part.

USES FOR QDA DATA

What are some of the uses of an expert panel developing descriptive data by the QDA method? There are undoubtedly many uses for a trained panel, but I will confine myself to examples of projects on which we have accumulated data and have results:

1. As an Aid in Product Development:

a. To describe the sensory characteristics of competitive products and one or more experimental prototypes (Figs. 3 and 4).

b. To document that the scaled-up product is like that originally developed.

c. To optimize a formula by describing several levels of several variables in more complex designs.

Table 4—NUMBER AND PERCENTAGE *of candidates screened by triangle test on one or more types of products; 170 persons participated in 279 screening tests to obtain 110 judges*

No. of times tested, each time on a different type of product	No. of persons tested	% of persons tested
1	97	57.1
2	51	30.0
3	12	7.1
4	6	3.5
5	2	1.2
7	1	0.1

Table 5—AVERAGE LEVEL OF DISCRIMINATION *for judges selected for QDA panels*

Product	No. of persons screened	No. of judges selected	% correct identifications in 24 tests for judges selected
Breakfast item	33	10	71.3
Natural Cereal®	35	11	79.6
Syrup	30	11	76.3
Pizza	31	11	64.6
Life® cereal	28	11	62.8
Shredded Wheat cereal	26	10	56.7
Cap'n Crunch® cereal	21	11	72.5
Frozen pancakes	25	12	71.7
Oatmeal cookies	23	12	78.3
Instant grits	27	11	85.8

Table 6—LEVEL OF DISCRIMINATION *(% correct identifications) shown by candidates for expert QDA descriptive panels in screening by triangle difference tests on various food products*

	No. of judges who paired correctly at the following levels of discrimination (% correct identifications):								No. of judges	
Product	75% or better	71%	67%	62%	58%	54%	50%	Less than 50%	Screened	Trained
Breakfast item	10	8	4	3	5	0	1	2	33	10
Natural Cereal®	17	2	3	3	6	3	0	1	35	11
Syrup	7	8	2	4	1	1	6	1	30	11
Pizza	2	1	3	3	2	8	4	8	31	11
Life® cereal	1	2	5	0	5	4	7	4	28	11
Shredded Wheat cereal	1	0	3	2	4	8	3	5	26	10
Cap'n Crunch® cereal	5	2	2	4	1	1	3	3	21	11
Frozen pancakes	5	9	4	1	1	1	1	3	25	12
Oatmeal cookies	11	4	2	1	1	2	1	1	23	12
Instant Grits	21	2	2	1	1	0	0	0	27	11
Total	80 (28.7%)	39	30 (24.7%)	22	22 (17.6%)	28	25 (19.0%)	28 (10%)	279	110

d. To provide language for description of an *ideal* or favorite product by consumers.

e. To describe changes in storage.

2. As an Aid in the Maintenance or Improvement of an Established Product. Here the QDA data furnished by a descriptive panel are used as an analytical tool and can function by themselves or in conjunction with objective measurements, or even consumer acceptance data. The following are several examples of these kinds of applications relating to typical problems which arise in continued manufacture of a product:

a. Can the variability be reduced? In one case, it was desired to reduce color variability of a cereal by setting color ranges on a colorimeter. Here, QDA data described the changes which took place so that realistic cut-off ranges could be set (Fig. 5). After the first description, product was made again, described a second time by QDA, and also submitted to children, the consumers of this cereal. Note how closely the results of the second QDA description duplicated those of the first, and also how the children reacted to the products described.

b. Can the manufacturing process be changed, for reasons of improvement or cost cutting? Descriptive data have been successfully used in changing extrusion conditions for a ready-to-eat cereal, as one example.

c. Can an ingredient be changed? Figure 6 shows that a desired change in processing of a particular ingredient could be used with negligible results on the product. Figure 7 shows that a 25% reduction in level of one ingredient will probably cause problems with the product at the low end of the permitted pouch weight; the projected decrease in the ingredient was dropped.

3. As a Diagnostic Tool when a product is slowly losing its accustomed share of the market. A detailed sensory description, together with that of the competitive product, is the first step in determining whether the problem is in the sensory area or in some other area, such as marketing. Figure 8 indicates that a shredded cereal biscuit possesses negative notes that the competitive product doesn't. Packaging and processing can now be examined to find the source of the critical off-notes.

4. As a Quality Control Measure. This depends upon the development and maintenance of a master configuration or QDA profile of the product against which later profiles of the product can be compared to measure "drift." This presupposes that the samples described are truly representative of most of the product that is being made at the point in time when the measurement is taken or that there is a historical knowledge of the amount of normal variability in manufacture. Figure 9 shows a ready-to-eat cereal as presently manufactured vs the same product from two years previous.

These are but a few illustrations of how the data from trained descriptive judges, when used in conjunction with statistical analysis to judge variability and reliability, can answer questions and give information.

The question inevitably arises as to whether sensory descriptive data can be used to predict *acceptance*. The experience with the 10 groups over a period of 2½-3 years

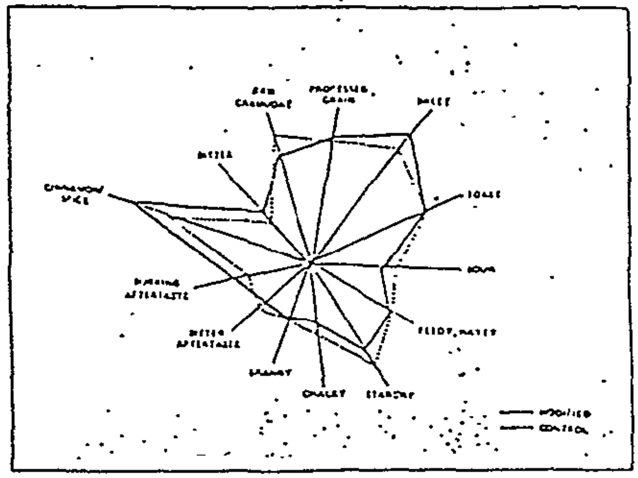


Fig. 3—QDA CONFIGURATIONS *for a prototype cinnamon-flavored cereal and a modification. Increasing the "cinnamon spice" and "sweetness" slightly resulted in a product with more impact and less "starchy" and "raw grain/oat" character. This modification was found to be significantly more acceptable in consumer testing and went on to become a successful market introduction. This is an example of how small but significant changes in key sensory attributes can be crucial to a product's success*

suggests that detailed sensory data can be used very effectively in combination with acceptance information obtained from larger or consumer groups.

In summary, each of the 10 descriptive sensory groups discussed in this article was developed by (1) screening 21-33 individuals to obtain the 10-12 most discriminating and (2) giving these individuals intensive training. They *all* went on to produce sensory descriptive data which met the standards of the particular technique for sensory descriptions which we were using (QDA). This was true even though there was some variation in the level of discrimination attained in the original screening on the various products.

REFERENCES

Amerine, M.A., Pangborn, R.M., and Roessler, E.B. 1965. "Principles of Sensory Evaluation of Food." Academic Press, New York.

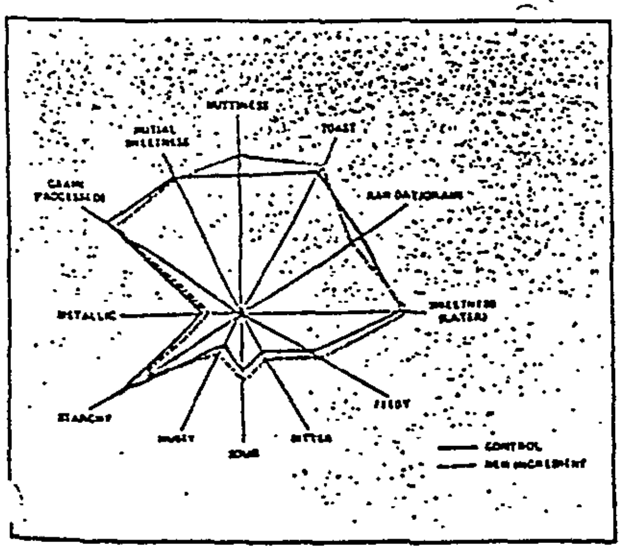


Fig. 6—EFFECT OF INGREDIENT REPLACEMENT *can be determined by QDA. Here, a sample of ready-to-eat oat cereal with milk was compared with another sample containing an ingredient prepared by a new process. The almost identical configurations (not significantly different) indicate that little flavor change need be expected if*

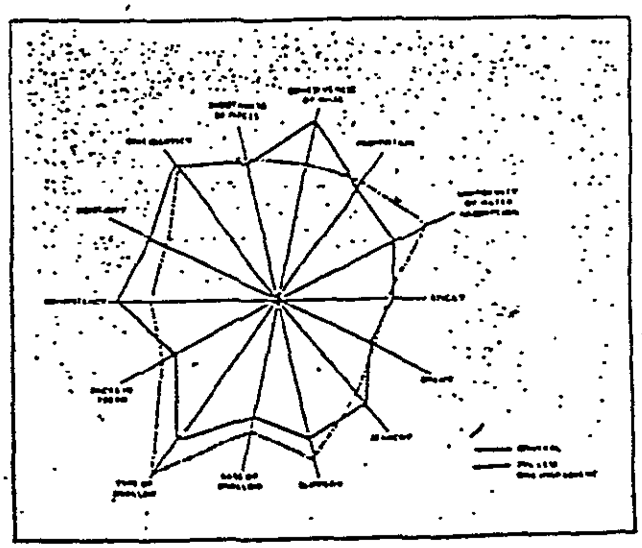


Fig. 7—EFFECT OF REDUCTION *in level of one ingredient can also be determined by QDA. Here, a comparison of a control sample of an instant cereal and a sample containing 25% less of one ingredient shows that the reduction would significantly alter the consistency and cohesiveness of the product*

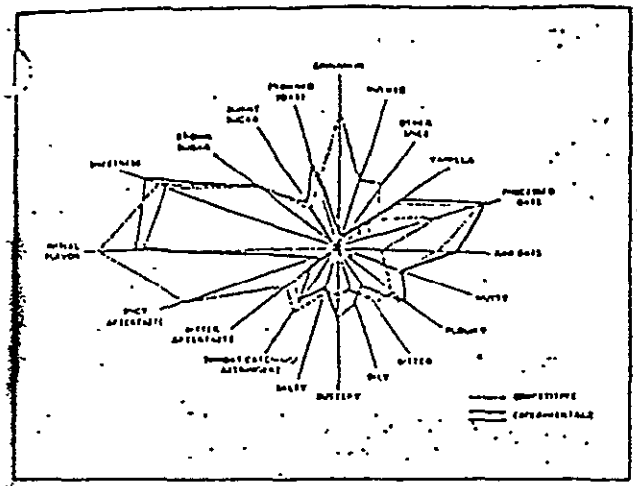
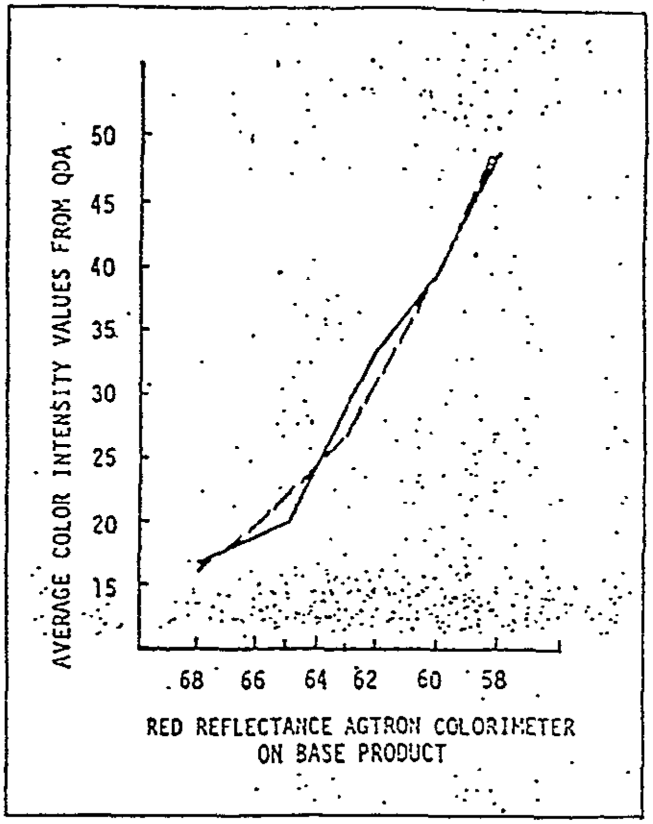


Fig. 4—CONFIGURATIONS OF FLAVOR ATTRIBUTES of a competitive brand of oatmeal cookies and two experimental prototypes (scale = 0 at center to 40 at perimeter)

Fig. 5—QDA DATA can be used to establish acceptable ranges of objective data. For example, QDA averages for color correlate with color readings from an Agtron colorimeter. Results are shown for two separate QDA descriptions. When product from the second was tested among 162 children (consumers of the cereal), the children significantly preferred product from the light end of the scale (Agtron value 68) and rejected the product from the dark end (Agtron value 58)



engtsson, K. 1943. Provsmakning som analysmetod. Statistisk behandling av resultaten. Svenska Bryggareforen Manadsblad 58.

Caul, J.F. 1957a. The profile method of flavor analysis. Adv. in Food Res. 7: 1.

Girardot, N.F., Peryam, D.R., and Shapiro, R. 1952. Selection of sensory testing panels. Food Technol 6: 140.

Helm, E. and Trolle, B. 1946. Selection of a taste panel. Wallerstein Lab. Commun. 9: 181.

Henderson, D. and Vaisey, M. 1970. Personality traits related to performance in a repeated sensory task. J. Food Sci. 35: 407.

Mecredy, J.M., Sonnemann, J.C., and Lehmann, S.J. 1974. Sensory profiling of beer by a modified QDA method. Food Technol. 28: 36.

Peryam, D.R. and Swartz, V.W. 1950. Measurement of sensory differences. Food Technol. 4: 390.

Stone, H., Sidel, J., Oliver, S., Woolsey, A., and Singleton, R.C. 1974. Sensory evaluation by Quantitative Descriptive Analysis. Food Technol. 28: 24.

Based on a paper presented during the Sensory Evaluation Division symposium on "The Selection of Judges for Sensory Testing," at the 37th Annual Meeting of the Institute of Food Technologists, Philadelphia, Pa., June 5-8, 1977.

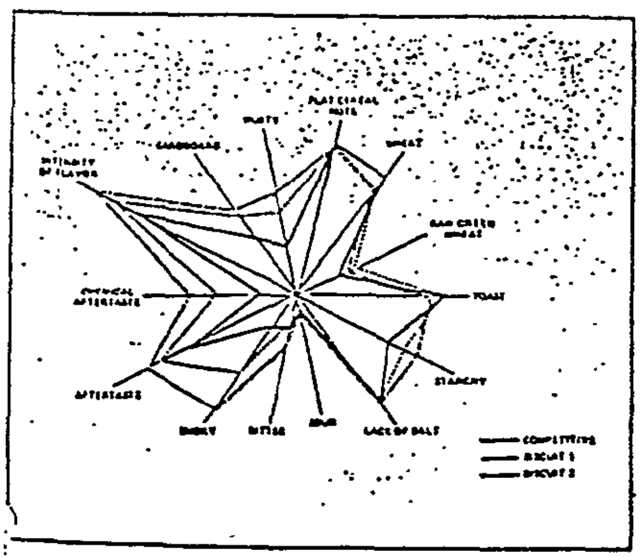


Fig. 8—AS A DIAGNOSTIC TOOL, QDA can indicate why a particular product might be less acceptable than a competitive product. Here, an unwrapped shredded wheat biscuit (Biscuit 1) and a wrapped biscuit (Biscuit 2) are compared to a competitive product. The comparison shows that the wrapped and unwrapped biscuits have negative notes that may be due to packaging or processing

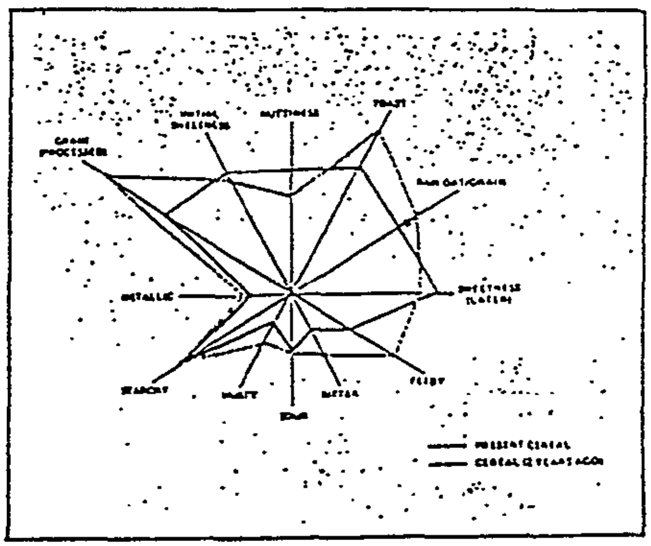


Fig. 9—AS A QUALITY CONTROL TOOL, QDA can determine whether a product has changed in character with time. Here, a profile of a dry ready-to-eat cereal is compared to a profile of the same cereal drawn two years previous. Note the increase in "nuttiness" (a positive attribute) and decrease in "feediness" (a negative attribute), as well as other changes

LABORATORIO No. 4

PRUEBAS DE PREFERENCIA UTILIZANDO PANEL INTERNO

OBJETIVO

Utilizar diferentes escalas para determinar aceptabilidad de los productos que se presentarán.

INSTRUCCIONES

1. Los participantes serán familiarizados con el uso de escalas hedónica de 9 puntos facial, y de ordenamiento.

2. Un problema será planteado a los participantes, ellos definirán las características de interés a ser evaluadas, formularán las preguntas y seleccionarán las escalas a ser evaluadas (hedónica 9 puntos, adecuado - inadecuado, ordenamiento).

LABORATORIO No. 4

PARTE A

ESCALA HEDONICA DE 9-PUNTOS

Por favor marque la frase que mejor describa cuanto le gusta a usted este producto:

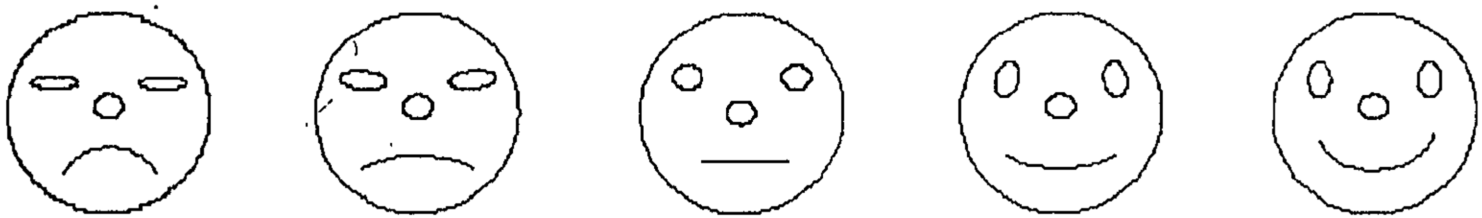
Código _____	Código _____	Código _____	Código _____
___ Gusta muchísimo	___ Gusta muchísimo	___ Gusta muchísimo	___ Gusta muchísimo
___ Gusta mucho	___ Gusta mucho	___ Gusta mucho	___ Gusta mucho
___ Gusta moderadamente	___ Gusta moderadamente	___ Gusta moderadamente	___ Gusta moderadamente
___ Gusta ligeramente	___ Gusta ligeramente	___ Gusta ligeramente	___ Gusta ligeramente
___ No gusta ni disgusta	___ No gusta ni disgusta	___ No gusta ni disgusta	___ No gusta ni disgusta
___ Disgusta ligeramente	___ Disgusta ligeramente	___ Disgusta ligeramente	___ Disgusta ligeramente
___ Disgusta moderadamente	___ Disgusta moderadamente	___ Disgusta moderadamente	___ Disgusta moderadamente
___ Disgusta mucho	___ Disgusta mucho	___ Disgusta mucho	___ Disgusta mucho
___ Disgusta muchísimo	___ Disgusta muchísimo	___ Disgusta muchísimo	___ Disgusta muchísimo
Comentario	Comentario	Comentario	Comentario

LABORATORIO No. 4

PARTE B

BOLETA PARA ESCALA FACIAL

Por favor marque bajo la figura que mejor describa cómo se siente usted respecto a cada producto.



Código

CUESTIONARIO

Prueba de aceptabilidad y Preferencia

PRUEBA DE ACEPTABILIDAD Y PREFERENCIA

Producto:_______________________________

El día de hoy, usted evaluará 3 muestras de pudines de chocolate, una muestra a la vez.

Hale el cartón rojo para recibir su primera muestra.

Antes de probar la muestra, por favor responda a la pregunta de Apariencia.

Código________

1. ¿Cuánto le gusta la apariencia de esta muestra?

___ Gusta extremadamente
___ Gusta mucho
___ Gusta moderadamente
___ Gusta ligeramente
___ No gusta ni disgusta
___ Disgusta ligeramente
___ Disgusta moderadamente
___ Disgusta mucho
___ Disgusta extremadamente

Ahora por favor pruebe suficiente muestra para hacer una evaluación adecuada.

2. ¿Cuánto le gusta esta muestra en general?
(tomando en cuenta aroma, sabor, textura)

___ Gusta extremadamente
___ Gusta mucho
___ Gusta moderadamente
___ Gusta ligeramente
___ No gusta ni disgusta
___ Disgusta ligeramente
___ Disgusta moderadamente
___ Disgusta mucho
___ Disgusta extremadamente

Antes de continuar con la siguiente pregunta por favor escriba sus comentarios sobre aspectos que le gustan y no le gustan de esta muestra en la hoja final del cuestionario.

3. ¿Cuánto le gusta el sabor de esta muestra?

___ Gusta extremadamente
___ Gusta mucho
___ Gusta moderadamente
___ Gusta ligeramente
___ No gusta ni disgusta
___ Disgusta ligeramente
___ Disgusta moderadamente
___ Disgusta mucho
___ Disgusta extremadamente

4. Por favor evalúe la dulzura de esta muestra.

___ Demasiado dulce
___ Ligeramente dulce
___ Adecuada
___ Poco dulce
___ Muy poco dulce

5. Por favor evalúe la intensidad de sabor a chocolate.

___ Demasiado fuerte
___ Ligeramente fuerte
___ Adecuada
___ Ligeramente débil
___ Demasiado débil

6. ¿Cómo evalúa la calidad del chocolate?

___ Excelente
___ Muy buena
___ Buena
___ Regular
___ Mala

7. Por favor evalúe el sabor amargo.

___ Extremadamente amargo
___ Muy amargo
___ Moderadamente amargo
___ Ligeramente amargo
___ No amargo

8. ¿Cuánto le gusta la textura de esta muestra?

___ Gusta extremadamente
___ Gusta mucho
___ Gusta moderadamente
___ Gusta ligeramente
___ No gusta ni disgusta
___ Disgusta ligeramente
___ Disgusta moderadamente
___ Disgusta mucho
___ Disgusta extremadamente

9. ¿Cómo evaluaría la cremosidad de esta muestra?

___ Demasiado cremosa
___ Ligeramente muy cremosa
___ Adecuada
___ Poco cremosa
___ Muy poco cremosa

10. ¿Compraría usted este producto?

___ Definitivamente lo compraría
___ Probablemente lo compraría
___ Tal vez lo compraría o no lo compraría
___ Probablemente no lo compraría
___ Definitivamente no lo compraría

Al finalizar la evaluación de las 3 muestras, por favor ordénelas en orden de preferencia.

1. ___________

2. ___________

3. ___________

Por favor escriba las razones por las que prefirió la muestra del primer lugar en la hoja adjunta.

HOJA DE COMENTARIOS

Nombre: ______________________________ Fecha: ______________

INSTRUCCIONES

El número de su muestra está escrito sobre el comentario en el orden que usted lo recibirá. Hay espacio para sus comentarios y para razones de preferencia. Solamente siga las instrucciones.

Razones específicas de gusto o disgusto para muestra No. ________

Gusto ____
____
____

Disgusto ____
____
____

Razones específicas de gusto o disgusto para muestra No. ________

Gusto ____
____
____

Disgusto ____
____
____

Razones específicas de gusto o disgusto para muestra No. ________

Gusto ____
____
____

Disgusto ____
____
____

Reprinted from FOOD TECHNOLOGY, Vol. 22, No. 5, Pages: 49-56

MEASURE OF Q

A Critical Review of Recent Literature on

preference testing methodology — PART 1 —

B. H. ELLIS

THE PURPOSE of this paper is to review recent publications concerned with preference testing methodology. Part One deals with definition, background and new concepts and methods. Part Two will cover the analysis, interpretation and current status of this important segment of the sensory evaluation field.

DEFINITION

PSYCHOLOGICAL measurement tests can be classified according to the kind of psychological functioning required of the subject as

(a) Affective—tests based on preference, pleasure-displeasure, like-dislike;
(b) Discriminative—tests based on judgment which may be concerned with difference per se or difference on a specified dimension;
(c) Descriptive—tests requiring the sorting out of the many separable qualitative dimensions which explain people's behavior toward products (Peryam, 1964).

Preference testing as used in this paper refers to *all affective tests based on a measurement of preference, or a measure from which relative preference may be determined (pleasure-displeasure, like-dislike)*. Preference as ... eva... defined by Amerine *et al.* (1965) as follows:

> (1) Expression of higher degree of liking; (2) Choice of one object over others; (3) Psychological continuum of affectivity (pleasantness-unpleasantness) on which such choices are based. This continuum is also referred to as that of degree of liking or disliking.

Preference is sometimes used interchangeably with acceptance. The two terms are related, but they are not the same. Acceptance has been defined by Amerine *et al.* (1965) as follows:

> (1) An experience, or feature of experience characterized by a positive (approach in a pleasant) attitude; (2) Actual utilization (purchase, eating). May be measured by preference or liking for specific food item.

Preference is only one of many factors involved in acceptability, but it is an important factor. Peryam *et al.* (1960) determined the relative preference of various foods by hedonic tests and measured the acceptability of these same foods by quantitative methods such as determining the percent of men taking a normal proportion of food, or the proportion of a normal portion consumed. Correlation between preference and acceptability was demonstrated. Peryam estimated that preference accounts for 35–60 percent of the variation in consumption, depending on the test conditions, and he thought it unlikely that any other single variable would be found as effective in predicting acceptability. It has since been determined that 75 percent of the variation in acceptance of army food can be predicted from preference data if factors of satiety, amount of fat and amount of protein are also taken into account (Pilgrim *et al.*, 1963).

Inaccurate use of the word "preference" sometimes causes confusion as to the method used. It is not uncommon for an investigator to report that "there was a preference" for a certain sample when in reality he has assumed this fact by assessing data from a test which does not measure preference per se (Wilson *et al.*, 1960).

A serious problem arises with those who recommend a preference method to determine difference (Shukis, 1967). • Preference methods can be used to determine differences in preference, but NOT differences per se; discriminative tests based on judgment (triangle, duo-trio, paired comparison) must be used for this purpose. • Difference tests should PRECEDE preference tests; *if there is no difference, there can be no preference*. Increased recognition of this fact might lead to fewer "break-evens" in preference testing.

METHODOLOGY—BACKGROUND

THE MOST COMMON methods of measuring preference are ranking, paired comparison and rating scales (Peryam *et al.*, 1960). The latter is also called the method of single stimulus or absolute judgment (no external standard), depending upon its use (Pilgrim *et al.*, 1955).

Ranking methods are easily applied and interpreted. For this reason, and also because they antedated other sensory evaluation techniques, preference ranking tests have enjoyed tremendous popularity.

The main disadvantage of these tests is that *they do not measure the degree* of preference difference between samples. Recent uses of the ranking method include that reported by Moncrieff (1966) with 132 odorants and 12 judges to determine trends in odor preferences; and with 10 odorants and 500 judges to determine general preference findings.

The paired comparison test is widely used probably because of its simplicity and ease of application.

It has come to be a more accepted practice than evaluating samples one at a time (Carroll, 1963). It has been used to determine preference with as many as 1,008 judges (Simone *et al.*, 1957).

Scheffe's method (1952) of paired comparisons provides a means of taking advantage of the interrelationship of the samples in order to build up precision on certain contrasts (Carroll, 1963). It includes a scoring system for degrees of preference; e.g., strongly "3", moderately "2", slightly "1", and no preference "0" which extended to the minus side becomes a seven-point scale. A plus value is used when the second sample is preferred. Gridgeman (1961) used a variation of Scheffe's method in which he dropped the neutral point.

Various rating scales have been developed for preference testing.

The best known rating scale is the nine-point hedonic scale (from "like extremely" to "dislike extremely") developed at the U.S. Army's Quartermaster Corps for the purpose of determining preferences as predictors of army food acceptability (Peryam *et al.*, 1952). A rating scale method was selected for this purpose because the ranking test was considered too impractical and the paired comparison test too involved for a large number of samples (Peryam *et al.*, 1960).

Variations of the hedonic scale include five, six, seven and eight-point scales. Jones *et al.* (1955a) determined

(1) that longer scales up to nine intervals tend to be more sensitive to preference differences;

(2) that elimination of the "neutral" category seemed to be beneficial; and

(3) that balance (an equal number of positive and negative intervals) is not an essential feature of a rating scale.

The QM continued to used the nine-point scale for the sake of uniformity of results. Palmer *et al.* (1966) reported the use of the eight-point scale with no neutral point for the evaluation of toughness differences in chickens in terms of consumer reaction.

Elimination of the neutral point in a preference rating scale precludes the possibility of the "no preference" vote and forces a choice. Proponents of the forced-choice state the following advantages of this approach:

(1) increased transmitted information (Jones, *et al.*, 1955a);

(2) discouragement of laziness (Gridgeman, 1961);

(3) allowance of an equal chance in favor of either sample (Gridgeman, 1961);

(4) use of the desirable equal interval scale (Gridgeman, 1961); and

(5) presence of potential subthreshold influences (Baker *et al.*, 1960; Pangborn *et al.*, 1964).

Arguments against a forced choice include the lack of regard for the integrity of the respondent who may wish to vote "no preference" and the danger of determining a significant preference for one sample over another when there is no true preference for either sample.

Ferris (1960) reported it a well recognized phenomenon (in no-forced-choice preference tests) that many respondents indicate a preference even though they really have none and should be indicating "no preference." This phenomenon was attributed to the fact that the respondents were in a test situation and therefore tried a little too hard. In situations like these, Ferris suggested they are influenced by external factors such as order of tasting effect, or bias in coding of the samples, to indicate a pseudo-preference for one or other of the samples.

Comparative Studies of Methods

Comparative studies of paired comparison and rating scale methods have been reported from several sources (Wever, *et al.*, 1928; Pfaffman, 1935; Barnhart, 1936; Hedlund *et al.*, 1954; Pilgrim *et al.*, 1955; Simone *et al.*, 1957; Murphy *et al.*, 1957; Gridgeman, 1961).

In general, the evidence indicates that the paired comparison method is more sensitive than the rating scale method. Rating scale methods employed in these studies included single stimulus hedonic, paired stimuli hedonic and quality scales.

There was some indication that the single stimulus hedonic method was more sensitive than the paired stimuli hedonic method. Simone *et al.* (1957) reported the single stimulus hedonic approximated the precision of the paired comparison, although their final conclusion was that the paired comparison was superior to the hedonic method.

Hedlund *et al.* (1954) observed that the degree of difference between two products might influence the results when both are served simultaneously. Pilgrim *et al.* (1955) determined the hedonic scale method and the paired comparison to be equally sensitive whether the difference in preference was large or small.

Laboratory Preference Panels vs. Consumer Panels

Mixed views have been published on the difficulties of correlating small-scale laboratory panels with large scale consumer tests.

Raffensperger *et al.* (1956) report that the usual difficulty is not that laboratory panels are too sensitive and find differences that consumers cannot appreciate, but more often that small panels cannot find differences that have been decidedly demonstrated by consumer preferences. Amerine *et al.* (1965) report that the laboratory group is carefully selected, highly trained, and hypercritical as compared to the general consumer.

Pangborn *et al.* (1964) stated that preferences of a small panel (11 judges) of milk experts were of little value in predicting preferences of milk-drinking consumers. Ellis (1963) reported similar findings when comparing the preference ratings of six experts with a larger consumer panel of 206 persons evaluating brands of beer.

There is general agreement that comparisons of small laboratory or pilot consumer groups agree fairly well with larger consumer studies in direction, but not in magnitude (Amerine *et al.*, 1965). (Small laboratory or pilot consumer groups would usually consist of a minimum of 24 selected persons rather than the 6 or 11 experts referred to above.)

Murphy *et al.* (1958) reported that ranking methods more nearly predicted consumer preference of sardines than did paired presentation.

Calvin *et al.* (1959) found good agreement between a student laboratory panel and a household consumer panel with the hedonic scale for various products. Peryam *et al.* (1957) reached the same conclusion when comparing a laboratory panel and a soldier consumer panel.

Caul *et al.* (1964) reported a study on beef soup with inosinate in which ninety percent of the consumer panel also discerned the flavor differences between the control soup and the inosinate soup which had been observed by the laboratory profile panel.

Comparison of Two Consumer Panels

Weckel *et al.* (1962) included a comparison of two panels in a study on the effect of added sugar in canned corn. A Madison, Wisconsin, panel of 814 families was compared with a Kroger panel (15 central states) of 590 families. The Madison panel demonstrated a greater percentage of preference votes for the lowest sugar level than did the Kroger panel. However, the Madison panel was asked *not* to make any additions to the corn, while the Kroger panel was allowed to add salt, pepper or butter in serving the samples. Addition of seasonings may have lessened the flavor effect of added sweetener. The panels agree fairly well on the drop-off point of preference votes for increasing amounts of sugar.

Caul *et al.* (1954) reported good agreement between two different panels with different instructions and questionnaires when judging beef soup with inosinate.

Place of Testing

Preference testing generally takes place in booths in stores, lobbies, fairs, transportation centers; in food markets; in church, school or club dining rooms; or in mobile units which operate in well-populated and travelled areas (Simone *et al.*, 1957; Coleman, 1964; Williams *et al.*, 1963).

Advantages of testing in such public places (rather than in the home) include the following:

- the researcher can maintain control over product handling, preparation, serving and questioning of the consumer;
- only the flavor characteristics under study can influence the consumer; and
- the chance of incorrect entries on questionnaires is minimized.

Group testing allows the stimulating effect of conversational recall among the consumers in sharing experiences and opinions and discussions can be tape-recorded in addition to filling out questionnaires by the participants. Recorded comments are of great importance in cases of borderline differences or where secondary taste characteristics are of interest. The *tone of voice can be more illuminating* than all the written comments (Coleman, 1964).

Home testing may be used for impartial appraisal of products when no labels or other company identification is used. In development guidance testing, pre-selected stratified family panels of fifty each are usually large enough to give directional results. Instructions for preparation and use of the products and questionnaires for each family member to evaluate the products are standard procedure in home testing.

An advantage of the home testing method is that factors (such as the package, appearance of the uncooked product, the cooking aroma, and behavior in handling) which may influence acceptance before the product is tasted can be evaluated.

Disadvantages include (1) loss of control in preparation so that improper handling or misunderstanding of directions may cause poor acceptability of a product on tasting, and (2) misunderstanding of the questionnaire may cause incorrect entries and thus lead to wrong conclusions (Coleman, 1964).

Preference Testing Techniques

Motivation and personality techniques are sometimes used to assess preference-determining factors in consumers.

The motivation technique of the depth interview—a free and flexible conversation between interviewer and correspondent—can be used in studying consumer preferences. The urgency or apathy of her attitudes and what part they play in final product preference are revealed through the respondent's interview. Patterns that set the consumer in action rather than isolated attitudes are sought (Anonymous, 1964).

The depth interview is generally supplemented with other techniques such as the Figure Drawing Test—a projective personality test in which a respondent is given a blank sheet of paper and a pencil and asked to draw a person of the same sex. Respondents reveal themselves in a variety of ways: the size of the figure drawn reveals how big and important, or how small and insignificant the respondent feels; and the placement of the figure reveals whether she is introverted or extroverted (Anonymous, 1964).

Devices developed for measuring consumer preference of packaging include the

- Stereo-Rater (determines which of two stimuli dominates awareness by presenting simultaneously and monocularly);
- a Visuo-Value Rater (measures speed and ease with which any visual display can be recognized and read);
- the Size-Distance Rater (measures apparent size, apparent distance, size constancy and emergence of features with changing distance);
- the Van Rosen Visiometric Comparator (measures various perceptual concepts of package design which make the package effective in the market place); and
- the automatic poll taker (eliminates the interviewer).

NEW CONCEPTS & METHODS

NEW concepts and methods reported for the determination of preference include: a new type of pilot consumer panel; the Food Action Rating Scale (FACT); the facial hedonic scale; the semantic differential; the trend-rating method; and pupil response measurement.

A New Concept In Pilot Consumer Testing

A new concept in pilot consumer testing is based on the premise that the food product itself—as represented by its use and sensory properties—is of first concern, taking priority over merchandising tactics, packaging and convenience (Caul *et al.*, 1959). Such a test may be an adequate predecessor to a market test provided that it is carefully designed. It should be designed to determine whether or not the product achieves the concept the producer had in mind when he set out to develop it.

The pilot consumer panel should be chosen according to the following test objectives:

(1) a modification or improvement of the current product;

(2) a replacement for an existing one;

(3) a challenger to a competitive product; or

(4) a new product which has no prototype.

Panelists are selected on the basis of the following qualifications: interest, communicativeness, discriminatory food buying habits and intelligence. The philosophy in selecting such a panel is that "if the product pleases them, it likely will please others whose quality standards may be lower and who are followers rather than leaders" (Caul *et al.*, 1959).

This new type of pilot consumer panel has been defined for very specific objectives directed toward the food product rather than toward the consumers themselves. It should not be confused with the usual small consumer test in which an effort is made to select a population representative of the intended users.

Food Action Rating Scale (FACT)

Three types of successive category-rating scales used for measurement of food attitudes are: quality judgment; like-dislike affect; and action.

A nine-point successive-category food action rating scale for measuring food acceptance was developed by Schutz (1965). The FACT scale can be seen in Figure 1. It requires the individual to be very specific about what actions he would take in terms of the number of times he would be interested in eating a food product in a given period.

A scale using nine-points was chosen to permit direct comparison of the new scale with a hedonic scale in order to test usefulness as a tool in a laboratory and in a survey questionnaire.

Comparison of the results showed a high correlation. The distribution of

FOOD ATTITUDE RATING FORM

NAME	DEPT.	BOOTH #	DATE

CODES: ____ ____ ____ ____

I WOULD EAT THIS EVERY OPPORTUNITY I HAD

I WOULD EAT THIS VERY OFTEN

I WOULD FREQUENTLY EAT THIS

I LIKE THIS AND WOULD EAT IT NOW AND THEN

I WOULD EAT THIS IF AVAILABLE BUT WOULD NOT GO OUT OF MY WAY

I DON'T LIKE IT BUT WOULD EAT IT ON AN OCCASION

I WOULD HARDLY EVER EAT THIS

I WOULD EAT THIS ONLY IF THERE WERE NO OTHER FOOD CHOICES

I WOULD EAT THIS IF I WERE FORCED TO

COMMENTS:

CODE______

CODE______

CODE______

CODE______

FIG. 1. *Food Attitude Rating Form for FACT Method. Reference: Schutz, 1965. Reprinted by kind permission of the Institute of Food Technologists.*

ratings was less skewed for the FACT scale than for the hedonic scale, and the FACT scale was more sensitive to food differences than the hedonic scale.

The superiority of the FACT scale could be due in part to the more realistic attitude that one has when one makes an action statement rather than simply an affective one concerning food products (Schutz, 1965).

The FACT scale is not meant as a replacement for the hedonic scale or other affective methods. The others can be used with effectiveness in evaluating the specific characteristics of food such as appearance or texture. The FACT scale can only be used as an over-all measurement of food acceptance (Schutz, 1965).

The FACT scale is being used at Hunt-Wesson Foods (where it was developed) as well as by independent market research groups.

Facial Hedonic Scale

Descriptive phrases may differ greatly in ambiguity (Jones *et al.*, 1955b) and problems in semantics have arisen with the use of descriptive rating scales.

Peryam *et al.* (1960) reported that judges drawn from the military tended to avoid the response "dislike moderately" in the hedonic scale. Jones *et al.* (1955b) in a study of psychophysical semantics found that there was confusion among such soldier subjects about the meaning of this phrase. Some placed it in its proper relationship to the other terms; others rated it on the "like" side of the scale. They concluded that this was related to the colloquial use of "moderate" as an expression of mild appreciation in the sense of "good" or "favorable." Thus a contradiction would arise for some people when the word was joined with "Dislike," and they would resolve the conflict by avoiding the category.

Simone *et al.* (1957) observed that paired hedonic test participants were frustrated when they "liked" the two

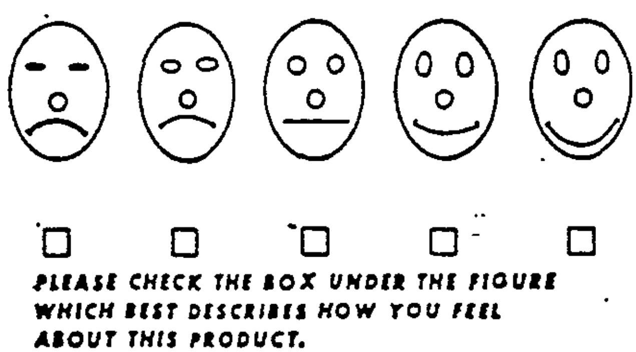


FIG. 2. *Facial Hedonic Scale Developed by the Continental Can Company, Inc. Reference: Ellis, 1964. Reprinted by kind permission of the Society of Soft Drink Technologists.*

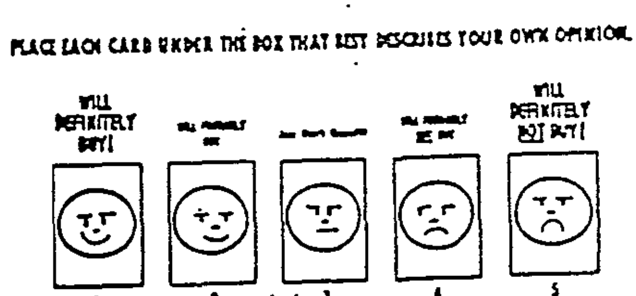


FIG. 3. *Facial Hedonic Scale Developed by the Oscar Mayer Company. Reference: Twedt, 1966. Reprinted by kind permission of the Oscar Mayer Company.*

samples equally, but had a preference for one—and found no place on the score sheet where this information could be recorded. They concluded that this method caused respondents to place more emphasis on completion of the response sheet than on precision in noting a difference in degree of likeness.

A modified hedonic method which serves to minimize confusion due to terminology is the *facial* hedonic or "Smiley" rating scale. In this method, faces depict the degree of pleasure, or displeasure, experienced by the subject. A neutral face is the median interval. These scales may consist of five, seven or nine faces. Figures 2, 3, 4 and 5 show examples of the *facial* hedonic scales. Data is treated in the same manner as for the *word* hedonic scale.

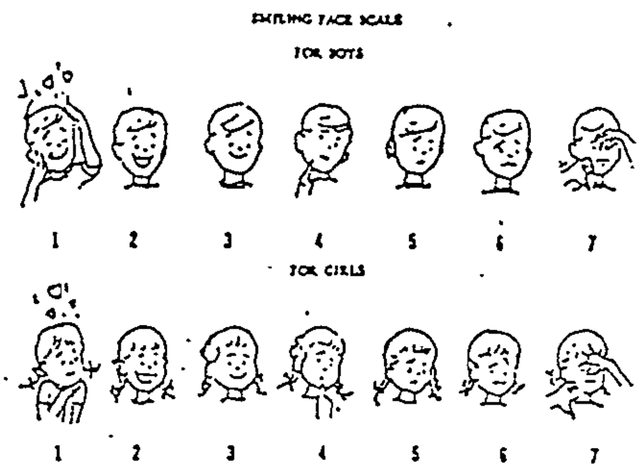


FIG. 4. *Facial Hedonic Scale For Children. Reference: Wells, 1965. Reprinted by kind permission of the* Journal of Advertising Research. *Copyright 1965 Advertising Research Foundation, Inc.*

Little published data is available to support the reliability of this method, but its fairly widespread use indicates that the method is considered to be both reliable and sensitive. General Foods, Oscar Mayer, Continental Can and Swift are among those who regularly use this method.

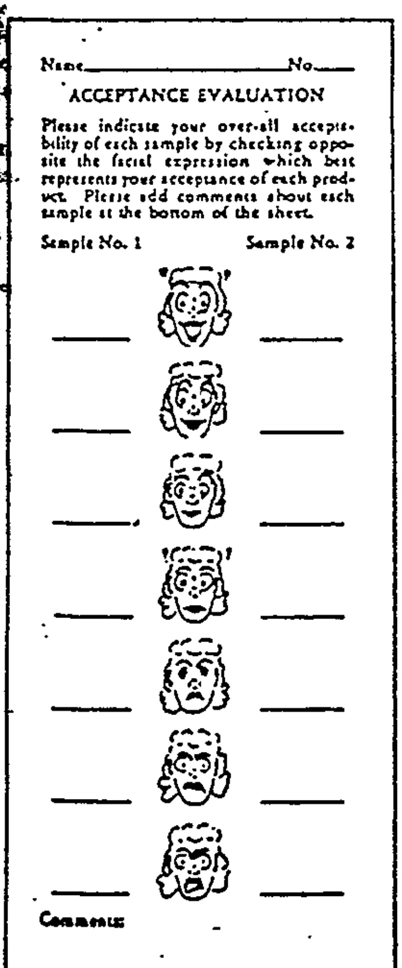


FIG. 5. *Facial Hedonic Scale Developed by Swift and Company. Reference: Schmalz, 1963. Reprinted by kind permission of Swift and Company.*

One investigator using a seven-point scale reports that there is a "play" on all faces and that this scale is more sensitive than the seven-point word hedonic scale.

Wells (1965) found that children 5–7 years old used this method well and that they are much readier to use the faces than they are to use words or numbers. The method appears to allow good communication with respondents who do not seem to have any difficulty understanding what is required of them regardless of age (5 years or older), intelligence, education or even ability to speak English. References for the facial hedonic method include Schmalz (1963); Ellis (1964); Wells (1965); and Twedt (1965).

Semantic Differential Method

Recently, semantic differential tests have been used to learn about consumer desires and reactions to flavors (Moores, 1966).

This type of test is designed to find out which descriptive measures really describe the flavor of the food. Consumers are given either a single list of words describing a flavor or two parallel lists of words with opposite meanings (see Fig. 6) and asked to give a relative numerical score of their opinions regarding each taste characteristic.

Osgood *et al.* (1957) provide a complete description of the logic, use and evaluation of this method as an objective measure of meaning. Some of the major advantages of the semantic differential are that many dimensions can be explored in a short time and that it is easy to administer (Twedt, 1966). Semantic differential tests can be conducted by mail. Figure 7 shows a scale for rating a concept of a product. Figure 8 shows profiles for five hypothetical variables judged on five different concept scales.

Hansen (1961) reported the Graphic Semantic Profile as a method for evaluating beer. A weighted average score is obtained for each flavor characteristic and these scores are plotted on graph paper.

The importance of the graph lies in the proximity of the two profile lines to one another. The arithmetic results shown by plotting the weighted averages have to be evaluated for statistical significance. (Hansen gives no details on how the statistical analysis is made.) Desirability of a significant difference depends on whether its profile line is to the left or right of the control beer's profile line. If the test beer's profile line is to the right, the taste tester reaction is decidedly unfavorable. If the two profile lines coincide or are not significantly different from one another, the two beers have registered

Mild	Bitter
Aged	Green
No Aftertaste	Aftertaste
No Afterthirst	Afterthirst
Not Gassy	Gassy
Pleasant Taste	Unpleasant Taste
Non-filling	Filling
Full-bodied	Watery

FIG. 6. *Parallel Lists of Words with Opposite Meanings Used in the Semantic Differential Method. Reference: Hansen, 1961. Reprinted by kind permission of* Modern Brewery Age.

BEST AVAILABLE

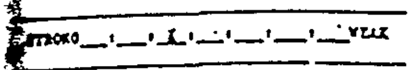


FIG. 7. *Scale for Rating a Product Concept in the Semantic Differential Method. Reference: Osgood et al., 1957. Reprinted by kind permission of the University of Illinois Press.*

...proximately the same taste sensation among testers for that particular ...ection. Testers might be called upon ...further describe the importance of flavor characteristics (Hansen, 1961).

Little has been published on the application of the semantic differential, possibly (1) because of the unwillingness of companies to publish consumer descriptions (profiles) of their products; or (2) because of problems involved in the analysis of data obtained.

There is a good deal of interest in ...e semantic differential approach. In ...dition to describing the flavor of ...d it could be useful in determining ...oduct or company image as well as ...evaluating package design. It is a ...thod which should be used with a ...at deal of common sense—it is NOT ...method which can be simply run off ...a computer.

Trend-Rating Method

The trend-rating method is concerned with repeated evaluations (use) of the same product. It was developed by Byer *et al.* (1961) for investigating consumer reaction to beer.

Earlier work by Byer *et al.* (1953) on discriminative methods (expert panels) indicated that multi-sample comparison may work to obscure the tasters' reactions. It was thought that this might be of particular importance with non-expert panels where the multi-sample evaluation is a marked change from the usual consumer situation in which only one product is consumed at a time, and that it might diminish a taster's sensitivity or otherwise alter his reaction to the flavor characteristics (Byer *et al.*, 1961).

The trend-rating method involves a single stimulus presentation with four sessions of evaluation for any one stimulus. An unstructured vertical scale (the best a beer could be—the worst a beer could be) is used.

The unstructured scale is used for several reasons:

(1) to provide more freedom in rating;

(2) to make it less likely that the taster will be influenced by his previous ratings;

(3) to avoid hedonic terms so that a person who has no special liking or disliking for a given product will not be asked to give an illogical rating of "very pleasing" or "very displeasing";

(4) to avoid misunderstanding of terms or numerical scales where the highest number sometimes corresponds to the best product and sometimes to the worst; and

(5) to allow a large part of the statistical operations to be done geometrically by plotting the ratings and trend lines on graph paper.

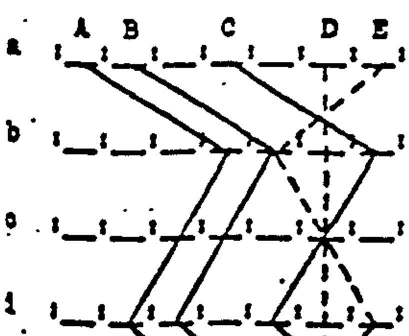


FIG. 8. *Semantic Differential Profiles for Five Hypothetical Variables Judged on Five Different Concepts. Reference: Osgood et al.,*

Graphing of each taster's performance indicates whether or not his rating is rising or falling. It also permits a comparison of taster performances. Such measurement of whether the product increases or decreases in acceptability with repeated consumption forms a novel feature of this method. The trend-rating method successfully indicated differences in acceptability between brands of beer (Byer *et al.*, 1961).

Pupil Response Measurement (Eye Camera)

The eye camera method is based on the measurement of an involuntary physiological response—the dilation or contraction of the eye pupil as a measure of acceptability. All other acceptability determining methods measure *voluntary* responses and are therefore subject to bias.

This method was developed by Hess at the University of Chicago (Hess, 1965; Hess *et al.*, 1960; 1964; 1966) and has been reviewed by Ellis (1967). It has been investigated mostly with ... stimulus picture is reverse projected by slides. A motion-picture camera records the reflected image of the eye. Size changes in the pupil are then measured on the film with a ruler.

Pleasant stimuli tend to cause dilation whereas unpleasant stimuli tend to cause constriction. Changes in pupil size are very sensitive, sometimes revealing response differences not apparent to subject or investigator at the verbal level.

Hess established a correlation between pupil response and preference by testing 64 persons with five pictures of food using the eye camera and also by asking them to rank the foods from "favorite" to "least preferred." The results do suggest that a technique for measuring a response ". . . that is not under the control of the person being tested may yield more accurate representations of an attitude than can be obtained with even a well-drawn questionnaire or with some devious projective technique in which a person's verbal or motor responses are recorded in an effort to uncover his real feelings" (Hess, 1965).

A limited amount of work has been done using the actual food stimulus rather than the visual stimulus. Flavor studies with liquids (carbonated drinks, chocolate drinks, milk, concentrated lemon juice and quinine solution) indicated that both pleasant and unpleasant drinks might cause an increase in pupil size. Of five orange drinks selected as being pleasant, one caused a significantly larger average increase in pupil size than did the other four. This data correlated with that obtained from a separate preference test in which the same orange beverage was preferred to the other four (Hess *et al.*, 1960).

Preliminary studies with the sense of smell and the eye camera have been conducted, but no published information is yet available. Effect of tactile stimuli on pupil response as concerned with textiles is also being investigated.

The pupil response method is currently being used to study responses to packages, products and advertising as well as the process of decision making, racial attitudes and the efficacy of different methods of problem-solving. Marplan, a communications-research organization in New York, and Allied Research Council in Chicago are two firms currently using the eye camera as a means of determining consumer attitudes.

REFERENCES

Evaluation of Food," p. 657; p. 540; p. 421. Academic Press, New York.

Anonymous. 1964. Some approaches to consumer acceptance. *Food Technol.* 18, 1712.

Baker, G. A., Amerine, M. A., Roessler, E. B. and Filipello, F. 1960. The nonspecificity of differences in taste testing for preference. *Food Res.* 25, 810.

Barnhart, E. N. 1936. A comparison of scaling methods for affective judgments. *Psychol. Rev.* 43, 387.

Byer, A. J. and Abrams, D. 1953. A comparison of the triangular and two-sample taste-test methods. *Food Technol.* 7(4), 185.

Byer, A. J. and Saletan, L. T. 1961. A new approach to flavor evaluation of beer. *Proc. Am. Soc. Brewing Chem. Ann.* 6.

Calvin, L. D. and Sather, L. A. 1959. A comparison of student preference panels with a household consumer panel. *Food Technol.* 13, 469.

Carroll, M. B. 1953. The use of subjective evaluation in product control and development. *Food Quality Control* III(5), 8.

Caul, J. F. and Sjostrom, L. B. 1959. Consumer food product acceptance. *Perfumery and Essential Oil Record*, November, 816.

Caul, J. F. and Raymond, S. A. 1964. Home-use test by consumers of the flavor effects of disodium inosinate in dried soup. *Food Technol.* 18, 353.

Coleman, J. A. 1964. Measuring consumer acceptance of foods and beverages. *Food Technol.* 18, 1711.

Ellis, B. H. 1963. Chicanery in the flavor room. *Proc. Am. Soc. Brewing Chemists*, 57.

Ellis, B. H. 1964. Flavor evaluation as a means of product evaluation. *Proc. 11th Ann. Meeting, Soc. Soft Drink Technologists*, 187.

Ellis, B. H. 1967. Pupil response measurement: new test cancels concealed product biases. *Food Product. Dev.* 1(1), 26.

Ferris, G. E. 1960. A new model in consumer testing. *Food Res.* 25, 802.

Gridgeman, N. T. 1961. A comparison of some taste test methods. *J. Food Sci.* 26, 171.

Hansen, W. F. 1961. Profiling beer tastes and brand images. *Modern Brewery Age*, December, 56.

Hedlund, G. J., Albrecht, B. and Wyatt, R. L. 1964. Single product or paired comparison consumer tests: which should you use? 9 pp. mimeo of paper presented at 24th Ann. Meeting, Inst. Food Technologists, Los Angeles.

Hess, E. H. 1965. Attitude and pupil size. *Scientific American* 212, 46.

Hess, E. H. and Polt, J. M. 1960. Pupil size as related to interest value of visual stimuli. *Science* 132, 349.

Hess, E. H. and Polt, J. M. 1964. Pupil size in relation to mental activity during simple problem solving. *Science* 143, 1190.

Hess, E. H. and Polt, J. M. 1966. Changes in pupil size as a measure of taste difference. *Perceptual and Motor Skills* 23, 451.

Jones, L. V., Peryam, D. R. and Thurstone, L. L. 1955a. Development of a scale for measuring soldiers' food preferences. *Food Res.* 20, 512.

Jones, L. V. and Thurstone, L. L. 1955b. The psychophysics of semantics. *J. Appl. Psychol.* 39, 31.

Moncrieff, R. W. 1966. "Odour Preferences" John Wiley, New York.

Moore, R. G. 1962. Flavor technology and fixation as influenced by instrumental and expert panel and consumer evaluation. Methodology of Flavor Evaluation Special Report 66-1, The Packaging Institute, Inc., New York, 7.

Murphy, E. F., Covell, M. R. and Dinsmore, J. S. 1957. An examination of three methods for testing palatability as illustrated by strawberry flavor differences. *Food Res.* 22, 423.

Murphy, E. F., Covell, M. R. and Berglund, R. M. 1958. A consumer survey versus panel testing for acceptance evaluation of Maine sardines. *Food Technol.* 12, 222.

Osgood, E. E., Suci, G. J. and Tannenbaum, P. H. 1957. "The Measurement of Meaning" University of Illinois Press, Urbana.

Palmer, H. H.; Klose, A. A., Smith, S. and Campbell, A. A. 1965. Evaluation of Toughness Differences in Chicken in Terms of Consumer Reaction. *J. Food Sci.* 30, 898.

Pangborn, R. M. and Dunkley, W. L. 1964. Difference-preference evaluation of milk by trained judges. *J. Dairy Sci.* 47, 1414.

Peryam, D. R. 1964. Psychological methods for measuring flavor effects. Proc. 1963 Conference on Advances in Flavor Research. USDA. 1: 72.

Peryam, D. R. and Girardot, N. F. 1952. Advanced taste-test method. *Food Eng.* July, 58.

Peryam, D. R. and Haynes, J. G. 1957. Prediction of soldiers' food preferences by laboratory methods. *J. Appl. Psychol.* 41, 2.

Peryam, D. R., Polemis, B. W., Kamen, J. M., Eindhoven, J. and Pilgrim, F. J. 1960. Food Preferences of Men in the U.S. Armed Forces. QM Food and Container Institute.

Pfaffman, C. 1935. An experimental comparison of the method of single stimuli and the method of constant stimuli in gustation. *Am. J. Psychol.* 47, 470.

Pilgrim, F. J. and Wood, K. R. 1955. Comparative sensitivity of rating and paired comparison methods for measuring consumer preference. *Food Technol.* 9(8), 385.

Pilgrim, F. J. and Kamen, J. M. 1963. Predictors of human food consumption. *Science* 139, 501.

Raffensperger, E. L. and Pilgrim, F. J. 1956. Knowledge of the stimulus variable as an aid in discrimination tests. *Food Technol.* 10(6), 254.

Robineau, D. N. 1961. Predicting consumer preference in packaging (mimeo). Presented at Packaging Institute's 23rd Ann. Nat. Packaging Forum, New York, 1961.

Scheffe, H. 1952. An analysis of variance for paired comparisons. *J. Am. Statistical Assn.* 47, 381.

Schmalz, P. E. 1965. The use of sensory evaluation in the meat industry. Food News and Views. Swift and Co., Chicago. Bull. 159, 6.

Schutz, H. G. 1965. A food action rating scale for measuring food acceptance. *J. Food Sci.* 30, 365.

Shukle, A. J. 1967. How to use a sensory panel to "taste-test" your products. *Bakers Weekly* January 23, 31.

Simone, M., and Pangborn, R. M. 1967. Consumer acceptance methodology: one vs. two samples. IFT Symposium 1, 17th Ann. Meeting. "The Methodology of Sensory Testing." p. 25.

Twedt, D. W. 1966. New developments in measurement and attitude scales (mimeo). Presented, Am. Marketing Assn., Chicago Chapter, Feb. 17, 1966.

Wackel, K. C., Martin, J. A., Lakamp, R., and Lyle, M. 1962. Effect of added sugar on consumer acceptance and physiochemical properties of canned cream-style corn. *Food Technol.* 16(6), 121.

Wells, W. D. 1965. Communicating with children. *J. Advertising Research* 5(2), 2.

Wever, E. G. and Zener, K. E. 1928. The method of absolute judgment in psychophysics. *Psychol. Rev.* 35, 466.

Wilson, G. D., Brown, P. D., Pohl, C., Weir, C. E. and Chesbro, W. R. 1960. A method for the rapid tenderization of beef carcasses. *Food Technol.* 14(4), 186.

Based on a paper presented in the symposium "Preference Testing Methodology" at the 27th IFT Annual Meeting in Minneapolis.

THE AUTHOR is a consultant. Address: 2205 Maple Ave., Evanston, Ill. 60201.

STATISTICAL ANALYSIS
Friedman Analysis for Ranked Data

The Friedman statistic is used to determine if differences exist between samples being ranked for a single attribute, such as sweetness, saltiness or preference. The data are ordinal and do not give the degree of difference between samples. The Friedman test, like the F-test, tells us whether the samples are different. To determine which samples are different a secondary multiple comparison procedure must be used.

(It should be noted that there is a set of tables developed by Kramer to determine if rank sums were statistically different. These tables published in 1973 in Food Technology are not longer considered correct. In 1988, Basker (Food Technology 42(2):79) published revised tables, which are more correct and are for > 20 panelists).

Ranking Tests:

For ranking tests where all samples are evaluated simultaneously, the procedure is as follows (see also Poste et al., pp. 26-29). This is a randomized complete block (RCB) design.

Example:

A ranking test was used to compare the saltiness of four soups made with NaCl, KCl, and two blends. Eight panelists rated the samples as shown in the table.

Rank scores of seasoning agents for saltiness (1 = most salty).

	Treatments			
Panelists	NaCl	KCl	Blend 1	Blend 2
1	1	3	2	4
2	2	3	1	4
3	1	2	4	3
4	1	2	3	4
5	1	4	2	3
6	1	3	2	4
7	2	4	1	3
8	1	2	3	4
Rank Sum	10	23	18	29

The Friedman's T statistic is calculated as follows

$$T = \left([12 / bt(t+1)] \sum_{j=1}^{t} X_{\bullet j}^2 \right) - 3b(t+1)$$

where

b =	no. of panelists
t =	no. of treatments
X.j =	rank sum of sample (column total)

For our example:

$$\begin{aligned}
T &= [12/8 \times 4 \times 5](10^2 + 23^2 + 18^2 + 29^2) - 3(8 \times 5) \\
&= (12/160)(100 + 529 + 324 + 841) - 3(40) \\
&= (0.075)(1794) - 120 \\
&= 134.55 - 120 \\
&= 14.55
\end{aligned}$$

Use table T5 to find the value of chi-square (χ^2) with 3 d.f. (t-1) at $\alpha = 0.05$. The value is 7.81. Since T is greater than 7.81 (14.55 > 7.81), the samples are significantly different in saltiness.

To determine which samples are significantly different use the HSD_{rank} test (Eq. 24 in Meilgaard, p. 268) for RCB design. (Note: Poste et al. on p. 28-29 call this test an LSD_{rank}. Meilgaard et al. uses a slightly different formula for LSD_{rank}: Eq. 15, p. 261. Be sure to specify which you are using.) In this case $\alpha = 0.05$, t = no. of treatments = 4), and d.f. always equals ∞ for the rank tests.

$$HSD_{Rank} = q_{\alpha,t,\infty} \sqrt{bt(t+1)/12}$$

From Table T14, $q_{\alpha,t,\infty}$=3.63

$$\begin{aligned}
HSD_{Rank} &= 3.63\sqrt{8(4)(4+1)/12} \\
&= 3.63\sqrt{13.33} \\
&= 13.3
\end{aligned}$$

Any two samples whose rank sums differ by more than 13.3 are significantly different.

Blend 2	- NaCl	= 29 - 10	= 19 > 13.3
KCl	- NaCl	= 23 - 10	= 13 < 13.3
Blend 1	- NaCl	= 18 - 10	= 8 < 13.3
KCl	- Blend 1	= 23 - 18	= 5 < 13.3
Blend 2	- Blend 1	= 29 - 18	= 11 < 13.3
Blend 2	- KCl	= 29 - 23	= 6 < 13.3

Therefore NaCl is significantly saltier than Blend 2, but no other significant differences exist in saltiness between samples.

However, if the $LSD_{Rank} = t_{\alpha/2,\infty}\sqrt{bt(t+1)/6}$ is used, the results are slightly different. (Use table T4 to calculate LSD_{rank}).

Pairwise Ranking

In this test, a series of samples are presented in pairs (multiple paired comparison), and the evaluation is done on the basis of a single attribute. This allows us to arrange the samples on an intensity scale of a given attribute, and gives a numerical indication of differences between samples. The significance is determined by the Friedman analysis. Since the panelists do not receive all the samples simultaneously, this is a balanced incomplete block design.

Example: The manufacturer of a vanilla cookie has the choice of 4 different vanilla extracts. The project objective is to determine which vanilla extract has the most potential for use. The test objective is to rank the vanilla cookies from the most to least intense vanilla flavor. Each cookie (vanillas A, B, C, D) is presented with every other cookie to 15 panelists. The number of times (out of 15) each (row) sample is rated as "more intense vanilla" than each (column) sample is shown in the table.

		Column Samples (less intense)			
		A	B	C	D
	A	---	1	0	3
Row	B	14	---	2	3
(more	C	15	13	---	0
intense)	D	12	12	15	---

First, calculate rank sums for the samples. Remember if A is ranked more intense it gets a 1, and if it is ranked less intense it gets a 2. So for sample A, add the sum of the row frequencies (0 + 1 + 2) to twice the column frequencies ([2(14 + 10 + 11)]).

Sample	A	B	C	D
Row frequency	(1+0+3)	(14+2+3)	(15+13+0)	(12+12+15)
2 x Column frequency	2(14+15+12)	2(1+13+12)	2(0+2+15)	2(3+3+0)
Rank Sum	86	71	62	51

The test statistic (Friedman's T) is calculated by the following equation, which is a special case of Eq. 13 in Meilgaard (t = treatments, p = panelists, R = rank sum).

$$T = \left((4/pt)\sum_{i=1}^{t} R^2 \right) - (9p[t-1]^2)$$

For this example:

$$T = (4/15 \times 4)(86^2 + 71^2 + 62^2 + 51^2) - (9 \times 15[4 - 1]^2)$$

$$= (0.067)(7396 + 5041 + 3844 + 2601) - 1215$$

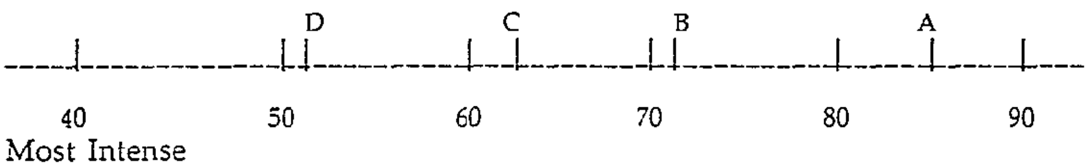
$$= 1265 - 1215$$

$$= 50.09$$

Compare the calculated value for T (50.09) with the critical value of c^2 with (t - 1) d.f. in Table T5.

α	0.05	0.01
Critical T	7.81	11.3

Since T is 50.09, it is greater than the critical T, so there is a significant difference in vanilla intensity.



To determine which are different, calculate the HSD value, using Table T_{14} for q. Note: This is a special case of Eq. 25.

$$HSD = q_{\alpha,t,\infty}\sqrt{pt/4}$$

$$HSD = 3.63\sqrt{15(4)/4}$$

$$HSD = 14.06$$

A	B	C	D
85	71	62	51

D has more intense vanilla flavor than A, B or C.

C is more intense than A, but no different from B.

A is significantly weaker than C or D.

Experimental Design and Analysis of Sensory Tests

Understanding experimental design is the first step in obtaining accurate results

JOEL L. SIDEL and HERBERT STONE

☐ THE MAJOR PURPOSE of any sensory evaluation study is to provide information regarding the effect of certain experimental treatments upon a particular population. That effect usually is described as changes or differences in a response which are measured and then analyzed using one or more mathematical operations. The accuracy of information provided by the sensory study will depend upon selection of an appropriate experimental design and appropriate analysis of the data.

DESIGN COMPONENTS

In a very broad sense, experimental design refers to all the primary components of a sensory experiment. These components are: product objective, test (hypothesis) objective, testing environment, samples, judges, response forms, serving procedure, and data analysis.

In a more-restricted sense, experimental design refers only to the particular sequence in which a set of samples is presented to a specified population of judges. Analysis refers to the specific mathematical operations applied to the collected responses. Selection of the appropriate design and analysis for a sensory experiment will depend on the information associated with these components. For this reason, it is appropriate to identify the contribution of each of these components to "use without abuse" in a sensory experiment.

• Objectives. The first stage in designing a sensory study is the planning; proper planning requires a complete and concise statement of both the project objective and the test objective, obtained through communication with the requester. Without adequate delineation of these objectives, one cannot be expected to design an appropriate experiment. The following are examples of appropriate and inappropriate statements of objectives:

Appropriate

Project Objective: To cost-reduce product X-Y-Z without altering its acceptability.
Test Objective: To determine whether current and cost-reduced X-Y-Z are detectably different.

Inappropriate

Project Objective: New formula X-Y-Z acceptance
Test Objective: Triangle difference test

A test objective may take the form of a hypothesis (which must be stated clearly and concisely during the planning stages). The hypothesis will have a considerable influence on selection of the statistical method for testing the hypothesis and, when appropriate, on the selection of the probabilities associated with use of a one- or a two-tail test. An example of a test objective requiring a one-tail test is: "To determine whether Sample A is rated sweeter than Sample B." Note that there is no concern whether Sample A is as sweet as or less sweet than Sample B. An example of a test objective requiring a two-tail test is: "To determine whether Sample A is rated equal in sweetness to Sample B." Note that there is concern whether Sample A is rated more or less sweet than Sample B.

• Testing Environment. The physical conditions of the facility are important because unanticipated problems can occur. The sensory evaluation literature is in good agreement regarding the most appropriate sensory test environment (Amerine et al., 1965; ASTM, 1968; Stahl and Einstein, 1973). One consideration of the test environment that deserves attention is the use of special lighting to mask visual product differences which could have a significant impact on the validity of test conclusions. This may occur (1) when the masking light does *not* successfully mask wavelength (or hue), purity (saturation), and lightness (luminance) differences, or (2) if the judge has not previously demonstrated an ability to perform sensory tasks under similar lighting conditions. The first is a physical consideration, the second is behavioral.

• Samples. The samples should be selected to provide an adequate test for both project and test objectives. Their type and number will influence most, if not all, components of the sensory experiment. For example, if the sample number exceeds that which a judge can comfortably evaluate in a single session, an incomplete block design may be required. Sample intensity will influence the interstimulus interval, as well as the procedure used to remove sample residue to minimize sensory fatigue. In addition, sample intensity may determine which test we select (e.g., single-sample, paired, or multisample test) and the serving sequence (e.g., simultaneous vs successive, random or balanced vs fixed, from weak to strong, etc.). In practice, sensory practitioners may choose a duo-trio test rather than a triangle difference test, simply to reduce judge exposure to intense samples.

• Judges. As with sample choice, judge type and

THE AUTHORS are with Tragon Corp., P.O. Box 783, Palo Alto, CA 94302

number are directly related to the test objective. Trained judges are used for descriptive and quality assessments. Experienced judges, selected on the basis of discrimination ability, are used in discrimination tests. And either experienced or inexperienced judges qualifying on the basis of acceptance attitudes ("likers" of a product class) are used for affective tests. Confsuion and problems can occur when the incorrect panelist is used, e.g., when untrained panelists are used in descriptive analysis or when trained panelists are used to obtain consumer acceptance data.

In most sensory tests, each judge receives more than one of the sample set, for reasons of test efficiency, thus dictating the use of designs and analyses for dependent rather than independent data. Often, however, over the course of a single study requiring many different test data, e.g., a storage study, the panel may consist of both new and "old" judges. This is especially true if we must draw on a limited number of judges from a fixed population at each withdrawal period of the storage study. We end up with both dependence and independence. Some solve this problem by choosing the more-conservative independence analyses; however, we would be more comfortable with better analyses designed specifically for this mixed situation.

Selection of the appropriate number of judges for an experiment poses additional difficulties. Use of too few judges may require large differences between samples for statistical significance, whereas use of too many judges may result in statistical significance with extremely small differences between samples. Both results may be valid from a statistical aspect, but they may be impractical and misleading to the product developer. We suggest consideration of both statistical and practical differences. Selection of the appropriate panel size will best bring together these differences. When in doubt, a preliminary study should be done from which the number of judges required for a specified level of significance can be calculated. Figure 1 presents data generated from a number of consumer test panels. From these data, we can determine the number of judges to use for the specified level of significance desired when designing similar types of panel tests. Note that we rapidly enter a situation of diminishing returns by adding panelists at around 30-40, and especially beyond 100 judges.

Fig. 1

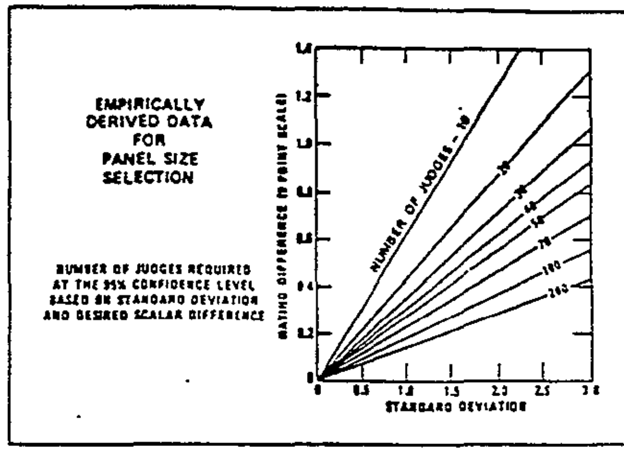


Table 1—EXPERIMENTAL DESIGNS *and test plans—a partial listing (source: Cochran and Cox, 1957)*

Design	Number of test plans
Completely randomized	—
Randomized block	—
Latin square	10
Complete	—
Incomplete	30
Quasi-Latin squares	14
Graeco-Latin squares	8
Crossover	—
Confounded	14
Split-plot	—
Factorial	2
Fractional factorial	—
Response surface	—
First-order	—
Second-order	—
Incomplete block	46
Balanced	—
Partially-balanced	—
Chain block	—
Lattice	—
Balanced	16
Partially-balanced	—
Rectangular	—
Cubic	—
Balanced lattice square	8

• **Response Form.** A primary consideration in the selection of an appropriate statistical test of the experiment's hypothesis is the form of response. Judges' responses may be classified into one or more of the following categories:

(1) *Open-end responses.* These refer to questions such as "What did you particularly like about Sample A?" or simply a request for "Comments." Upon obtaining these responses, the experimenter proceeds to establish categories of response reflecting the statements made by the judges.

(2) *Selection or choice responses.* These are frequently used in sensory evaluation. In each case, the judge is asked to select the sample from the set according to some stated criteria, such as preference, difference, a specific attribute, or some level of an attribute. Usually the judge is required to select the one sample which satisfies the specified criteria; occasionally he may be asked to select more than one from the set. The latter is a sorting task and, because of the additional complexities of assigning appropriate probabilities to sample choice (as well as the judge's need to have a good memory), is less popular.

(3) *Scaled responses.* Judges may be asked to scale or otherwise attach some number or value to each sample. They may be asked to rank each sample, score it on a rating scale (either numerical or graphic), or assign some ratio or magnitude (Stevens, 1956; 1962) relative to another sample or established reference. For each response form, there are appropriate procedures for statistical analysis. When judges are asked to perform multiple selections or score many aspects of a sample set, as with a multidimensional scaling task, additional complex data reduction

analyses are available (Green and Rao, 1972). Many of these procedures will be discussed below.

• **Serving Procedure.** Another major consideration in designing the experiment, and one which will influence selection of the statistical analyses, is the sample serving procedure. This procedure includes the physical conditions of the sample and the presentation sequence (i.e., the test plan). Table 1 is a partial list of experimental designs and test plans taken from a textbook on experimental design (Cochran and Cox, 1957). More than 20 design categories and 150 different test plans are listed. Awareness of design plans should help the sensory practitioner select or develop designs for which a satisfactory analysis is available. It is less than desirable to do an experiment to which no reasonable statistical analysis may be applied.

In sensory testing, we want to develop design plans that minimize the interactive effects of judge, product, and time, rather than test for them. For this reason, completely-randomized designs are often less desirable than balanced-block designs which have been completely balanced for judge, order, sample, and replication.

An example of completely-balanced, 4-sample,

Fig. 2

4—SAMPLE BALANCED BLOCK DESIGN

Judge	Order	Judge	Order	Judge	Order
1	1 2 3 4	5	2 3 4 1	9	3 4 1 2
2	2 4 1 3	6	3 1 2 4	10	4 2 3 1
3	3 1 4 2	7	4 2 1 3	11	1 3 2 4
4	4 3 2 1	8	1 4 3 2	12	2 1 4 3

Judge	Order	Judge	Order	Judge	Order
13	4 1 2 3	17	1 4 2 3	21	2 3 1 4
14	1 3 4 2	18	4 3 1 2	22	3 4 2 1
15	2 4 3 1	19	2 1 3 4	23	1 2 4 3
16	3 2 1 4	20	3 2 4 1	24	4 1 3 2

Fig. 3

4—SAMPLE, 3 REPLICATION, BALANCED BLOCK DESIGN

JUDGE	REPLICATIONS			
	Order of Presentation	Order of Presentation	Order of Presentation	Order of Presentation
1	4 3 1 2	3 2 4 1	2 1 3 4	1 4 2 3
2	2 1 3 4	1 4 2 3	4 3 1 2	3 2 4 1
3	1 2 3 4	2 3 1 4	3 4 2 1	4 1 3 2
4	3 4 2 1	4 1 3 2	1 2 4 3	2 3 1 4
5	4 1 2 3	1 3 4 2	3 2 1 4	2 4 3 1
6	3 2 1 4	2 4 3 1	4 1 2 3	1 3 4 2
7	4 2 3 1	2 1 4 3	1 3 2 4	3 4 1 2
8	1 3 2 4	3 4 1 2	4 2 3 1	2 1 4 3
9	2 3 4 1	3 1 2 4	1 4 3 2	4 2 1 3
10	1 4 3 2	4 2 1 3	2 3 4 1	3 1 2 4
11	2 4 1 3	1 2 3 4	3 1 4 2	4 3 2 1
12	3 1 4 2	4 3 2 1	2 4 1 3	1 2 3 4

Fig. 4

TEST PLAN FOR 4 SAMPLES NOT COMPLETELY BALANCED

JUDGE	ORDER
1	4 1 3 2
2	1 3 4 2
3	2 3 4 1
4	2 4 3 1
5	1 2 3 4
6	4 2 1 3
7	3 1 4 2
8	3 1 2 4
9	2 4 3 1
10	3 2 1 4
11	2 1 4 3
12	4 3 1 2
13	4 3 2 1
14	1 3 2 4
15	3 2 1 4
16	3 4 1 2
17	1 2 4 3
18	4 3 1 2
19	1 2 4 3
20	4 3 2 1
21	1 2 3 4
22	2 4 1 3
23	2 4 3 1
24	2 1 3 4

NOTES:

1. EACH JUDGE TESTS EACH SAMPLE

2. SAMPLE

Sample	Position				
	1	2	3	4	
1	6	5	7	6	24
2	7	7	4	6	24
3	5	7	7	5	24
4	6	5	6	7	24
	24	24	24	24	

3. PAIRS

Pairs	Position			
	1–2	2–3	3–4	
1 2	....	.	...	8
1 3	..	..	..	6
1 4	0	..	..	4
2 1	..	...	..	7
2 3	.	..	0	3
2 4	....	..	..	8
3 1	..	..	...	7
3 2	..	...	.	6
3 4	.	..	...	6
4 1	.	...	0	4
4 2	.	0	..	3
4 3	....	..	....	10
	24	24	24	

monadic, sequential tests is shown in Figure 2. Note that the design is not random, but is completely balanced. All possible permutations occur and are ordered in blocks of four, such that each sample is tested in each serving position before another block is begun. This may not entirely eliminate a sample × time interaction during a test; however, it should minimize it. This exact design need not be repeated each time a 4-sample test with 24 judges is planned. Other serving sequences are possible, as can be seen when replications are added (Fig. 3). However, note that the replicated design also is completely balanced such that each judge tests each sample in all possible serving orders. The 4-sample, 24-judge, no-replication-type design may be used in a laboratory acceptance test; the 4-sample, 12-judge, 3-replication-type design may be used in a trained-panel descriptive test.

For most sensory tests, replication is desirable for both statistical and behavioral reasons. Statistically, error-term selection in two-way designs (similar to those shown in Figs. 2 and 3) often can be substantially improved by adding a replication. Behaviorally, replication gives us an opportunity to study response consistency of both the panel and the individual judges. Although randomization over a large number of tests may be a more-desirable approach, random plans may introduce considerable order bias as related to an individual judge, sample, or period of time. An example of a random design in which the only restriction is that each judge test each sample is given in Figure 4. Note that all samples do not occur in all test positions an equal number of times, thus failing to minimize position bias which occurs in sensory tests (Eindhoven et al., 1964). Next, some sample pairs never occur, while others occur as many as 10 times, thus not equalizing possible contrast and convergence effects, again known to occur in sensory tests (Kamenetzky, 1959). Finally, all permutations do not occur an equal number of times.

In balanced-block designs, the complete permutation block should be presented before repeating any order in that block. Completing the permutation before replication should ensure that all samples were tested in all positions equally, both early and later in the study, thus minimizing possible time and sequence effects.

In discrimination tests, we may also wish to minimize the sample × time effect. Figure 5 shows how this may be achieved in the triangle test. There are six possible orders for serving in the triangle test. The first block of six permutations is given to the first six judges, the second block to the next six, and so on. Within each block of permutations, the experimenter may choose to rearrange or randomize the position of an order. If judges are to receive more than one set of the test samples, the experimenter may randomly assign the repeat sets, as in Figure 6, or again completely balance the design, as in Figure 7, such that judges will not get the odd sample in the same position (unless more than three sets are given to a judge). Random assigning of repeat sets is most appropriate when judges are frequent participants in triangle tests and may "learn" to expect that the odd sample position is *always* the same or different on replicate tests. When such learning is not possible or is minimal, the completely-balanced approach may be

Fig. 5

TEST PLAN FOR COMPLETELY BALANCED TRIANGLE TEST
(One Set per Judge)

THERE ARE 6 POSSIBLE ORDERS FOR SERVING A TRIANGLE TEST:
ABB BAB BBA
BAA ABA AAB

JUDGE	
1	ABB
2	BAB
3	BBA
4	BAA
5	ABA
6	AAB

JUDGE			
7	ABB	OR ➧	BBA
8	BAB		BAB
9	BBA		ABB
10	BAA		AAB
11	ABA		ABA
12	AAB		BAA

NOTES:
1. ALL PERMUTATIONS OCCUR AN EQUAL NUMBER OF TIMES
2. ONE COMPLETE PERMUTATION BLOCK OCCURS BEFORE A SECOND COMPLETE PERMUTATION BLOCK IS INTRODUCED
3. JUDGES RECEIVE THE A OR B SAMPLE AS THE ODD SAMPLE.

Fig. 6

TEST PLAN FOR PARTIALLY BALANCED TRIANGLE TEST
(Two Sets per Judge)

Judge	Set 1	Set 2
1	AAB	BBA
2	BBA	ABA
3	BAA	BAB
4	ABA	ABB
5	ABB	BAA
6	BAB	AAB

NOTES.
1. ALL PERMUTATIONS ARE USED IN EACH SET.
2. SELECTION OF SET 2 IS RANDOM EXCEPT THAT EACH JUDGE RECEIVES A DIFFERENT ODD SAMPLE THAN RECEIVED BY HIM/HER IN SET 1.
3. SOME JUDGES RECEIVE ODD SAMPLE IN SAME POSITION IN BOTH SETS.
4. EACH JUDGE RECEIVES THE A AND B SAMPLE AS THE ODD SAMPLE

Fig. 7

TEST PLAN FOR COMPLETELY BALANCED TRIANGLE TEST
(3 Sets per Judge)

	JUDGE					
	1	2	3	4	5	6
1st SET	A B B	A A B	B A B	B B A	A B A	B A A
2nd SET	A A B	B A B	B A A	A B B	B B A	A B A
3rd SET	B A B	A B B	A A B	A B A	B A A	B B A

A second design is obtained by substituting A for B and B for A.

NOTES
1. ALL PERMUTATIONS ARE USED IN EACH SET
2. SELECTION OF EACH SET IS FIXED SUCH THAT EACH JUDGE RECEIVES THE ODD SAMPLE IN A DIFFERENT POSITION ON EACH SET

• **Data Analysis.** Figure 8 lists some typical kinds of statistical tests used to analyze sensory data and the judge response form most appropriate for these analyses.

Open-end comments can be counted once they are categorized, and these frequencies can be converted [illegible] The cate[illegible]ry with the most comments may be of interest; seldom is further analysis warranted.

To analyze choice data, there are many special tables available which are based on the chi-square (χ^2) or binomial distribution. In paired tests, the data may be converted to percentages and a t-test for proportions performed. For rank data, one can refer to tables (such as those published by Kramer, 1960; 1963) that identify samples ranked significantly different. Other treatments of rank data include chi-square analysis of rank order and Spearman's correlation technique (Siegel, 1956; Snedecor, 1956; Hollander and Wolfe, 1973).

If data have been collected using a scalar device but the assumptions for use of metric statistics are not met, many non-parametric techniques are available for analysis (Siegel, 1956; Hollander and Wolfe, 1973). The non-parametric methods require conversion of the data to ranks and may be more powerful than metric statistics in cases where the assumptions for quantitative analysis have been seriously violated. Other procedures for analyzing data whose statistical assumptions are in question or violated include log and square-root transformations prior to conventional analyses. The important point is to select the appropriate analysis for the data.

PLAN AHEAD

There are many designs, plans, and analyses to choose from. The experimental design should be selected well in advance of the sensory test and should be chosen to at least provide: (1) an unbiased estimate of the effect to be measured, (2) a valid estimate of the variability of the estimated effect, (3) an opportunity to use a simple mathematical model in the analysis of data and for testing a specific hypothesis concerning the true effects, and (4) efficiency in terms of cost per unit of information.

Fig. 8

TYPICAL DATA ANALYSES AS RELATED TO JUDGE RESPONSE FORM

JUDGE'S TASK	TYPICAL DATA ANALYSES
OPEN END	[illegible]
CHOICE (e.g., [illegible] and paired preference)	SPECIAL TABLES BASED ON χ^2 [illegible]; BINOMIAL DISTRIBUTION t-TEST FOR PROPORTIONS
SCALE	
• ORDER (ranking)	RANK SUM TABLES, χ^2 ANALYSIS OF RANK ORDER, SPEARMAN'S
• INTERVAL — NON-PARAMETRIC	MANN-WHITNEY U TEST ([illegible]) [illegible]
	FISHER SIGN TEST ([illegible])
	WILCOXON MATCHED RANK TEST ([illegible])
	KRUSKAL-WALLIS ([illegible])
	FRIEDMAN ([illegible])
	HOLLANDER AND WOLFE ([illegible])
• INTERVAL — FULLY METRIC	MEANS, MODE, MEDIAN, t-STATISTIC, ANALYSIS OF VARIANCE, DUNCAN
	TUKEY [illegible] SIGNIFICANT DIFFERENCE, SCHEFFE, PRODUCT MOMENT CORRELATION
	REGRESSION, MULTIPLE REGRESSION, FACTOR ANALYSIS, CO-VARIANCE
• RATIO	ALL ABOVE, LOGARITHMS OF RATIOS, COEFFICIENT OF VARIATION
MULTIDIMENSIONAL SCALING	
• NON-METRIC	M-D SCAL V, TORSCA 9 ([illegible])
• PARTIALLY METRIC	TORSCA 9 ([illegible])
• FULLY METRIC	TORGERSON'S CLASSICAL ANALYSIS, [illegible], INDSCAL

Don't force-fit your data to an analysis because it is the only analysis you know or have. Plan your experiment ahead so that you can orchestrate properly all the components of experimental design to allow for the most powerful analysis possible.

Finally, a note of caution: Summary statistics are fine; however, procedures which allow for viewing individual judge response patterns may help you to avoid nonsense similar to looking for that one-half person in the average family of two and one-half.

REFERENCES

Amerine, M.A., Pangborn, R.M., and Roessler, E.B. 1965. "Principles of Sensory Evaluation of Food." Academic Press, New York.

ASTM 1968. "Manual on Sensory Testing Methods." STP 434. Am. Soc. for Testing and Materials, Committee E-18. Philadelphia, Pa.

Cochran, W.G. and Cox, G.M. 1957. "Experimental Design," 2nd ed. John Wiley & Sons, New York

Eindhoven, J., Peryam, D., Heiligman, F., and Hamman J.W. 1964. Effect of sample sequence on food preference. J. Food Sci. 29(4): 520

Green, P.E. and Rao, V.P. 1972. "Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms." Holt, Rinehart and Winston, Inc., New York.

Hollander, M. and Wolfe, D.A. 1973. "Nonparametric Statistical Methods." John Wiley & Sons, New York.

Kamenetzky, J. 1959 Contrast and convergence effects in ratings of foods. J. Applied Psychol. 43(1): 47

Kramer, A. 1960. A rapid method for determining significance of differences from rank sums. Food Technol. 14(11). 576.

Kramer, A. 1963. Revised tables for determining significance of differences. Food Technol. 17(3). 86.

Siegel, S. 1956. "Nonparametric Statistics for the Behavioral Sciences." McGraw-Hill, New York.

Snedecor, G.W. 1956. "Statistical Methods," 5th ed. Iowa State College Press, Ames, Iowa.

Stahl, W H. and Einstein, M.A. 1973. Sensory testing methods. In "Encyclopedia of Industrial Chemical Analysis," Vol. 17. John Wiley & Sons, New York.

Stevens, S.S. 1956. The direct estimation of sensory magnitudes: Loudness. Am. J. Psychol. 69(1).1.

Stevens, S.S. 1962. Mathematics, measurement, and psychophysics. In "Handbook of Experimental Psychology," S.S. Stevens, ed. John Wiley & Sons, New York.

Based on a paper presented during the Sensory Evaluation Division symposium, "Use Without Abuse of Sensory Measurements," during the 36th Annual Meeting of the Institute of Food Technologists, Anaheim, Calif., June 6-9, 1976.

APPENDIX VI

Questionnaire and Results

Nombre:____________________

Compañía:____________________

Fecha:____________________

EVALUACION DE CONOCIMIENTOS
ANALISIS SENSORIAL

I. Por favor consteste a cada una de las siguientes preguntas.

1. Defina qué es análisis Sensorial:

2. ¿Puede el análisis sensorial ser sustituido por pruebas instrumentales? ¿Por qué?

3. Describa brevemente el mecanismo de percepción de olores:

4. Mencione 2 aspectos importantes para llevar a cabo análisis sensorial:

5. Mencione 3 áreas en las que se utiliza análisis sensorial dentro de la industria de alimentos:

6. Describa por que es importante el orden de presentación de muestras en una prueba de análisis sensorial:

II. Selección Múltiple

Para cada pregunta, seleccione la letra que corresponde a la respuesta correcta y escríbala en la línea de la derecha.

7. El patron de inspiración y espiración a través de la nariz durante la evaluación de odorantes se llama?

   a. Respiración
   b. Husmeo
   c. Aromatización
   d. Sensibilidad olfatoria ____

8. La probabilidad de obtener una respuesta correcta al azar en la prueba de duo-trio es:

   a. 25%
   b. 33%
   c. 50%
   d. 62% ____

9. En las pruebas afectivas, los evaluadores deben ser:

   a. Semi-entrenados
   b. No entrenados (consumidores)
   c. Expertos
   d. Capaces de discriminar ____

10. La prueba que identifica y cuantifica las características sensoriales de un producto se llama:

   a. Prueba de ordenamiento
   b. Análisis descriptivo
   c. Prueba hedónica
   d. Ninguna de las anteriores ____

11. La prueba en la que el juez identifica el tipo de estímulo que esta evaluando, 50% de las veces, se llama?

a. Umbral de detección
b. Umbral de diferencia
c. Umbral de reconocimiento
d. Umbral términal ____

12. ¿El error de rechazar la hipótesis nula (Ho) cuando ésta es correcta, se conoce en estadística como?

a. Error tipo II (B)
b. Error tipo I ( )
c. Propabilidad 0.9 ( )
d. Ninguno de los anteriores ____

13. Describa brevemente cuáles son las más importantes pruebas de diferencia y las ventajas y desventajas de cada una de ellas:

14. Describa el tipo de pruebas sensoriales y tipo de jueces que usted utilizaría para resolver la siguiente situación:

La compañía desea utilizar un substituto de grasa en la elaboración del queso cheddar bajo en grasa. Este nuevo producto (con 1/3 menos de grasa comparado con el producto normal), está destinado a satisfacer la necesidad de un sector específico de consumidores.

Se ha preparado un prototipo en planta y aún no se conocen las características sensoriales del mismo. Desarrollo de nuevos productos, todavía no se sabe si este producto será aceptado por el consumidor. ¿Le gustará a la población?

Que tipo de pruebas sugeriría que se hicieran si usted fuera el analista sensorial.

RESULTS FROM SENSORY EVALUATION KNOWLEDGE TEST

Participant	Initial Score % correct	Final Score % correct
Patricia Bernal	27	52
Orlando Buitrago (was late)	-	80
Martha Caceres	26	71
Nidia Fajardo	26	61
Javier Fernadez	46	70
Nelson Gonzalez	25	75
Rosalina Lopez	12	35
Iris Lopez	23	71
Victoria Monje	27	93
Liz Murillo	42	76
Liza Pira	36	69
Gustavo Reyes	27	80
Indra de Reyes	28	75
Pedro odriguez	32	46
Sonia de Sarmiento	39	47
Frank Tosta	24	69
Marissa Umana	27	83
Ana Lucrecia Urizar	40	68
Lisbeth Villatoro	17	67
Paulina Wittkowski	36	92
Jeanny Zimeri	32	90
Mean score	29	70
Max	46	93
Min	12	35

APPENDIX VII

Course Evaluation

Table 1
Course Organization

Questions	Excellent %	Good %	Regular %	Bad %	Very Bad %	Total %
How was the coordination of the course?	55	40	5			100
How was the organization of the program activities?	50	45	5			100
How was the organization of the laboratory practices?	55	40	5			100

Table 2
Instruction and Course Content

Questions	Excellent %	Good %	Regular %	Bad %	Very Bad %	Total %
How was the presentation of the lectures?	65	35				100
How was the task developed by the instructor?	85	15				100
How would you rate the classroom facilities	20	65	15			100
How would you rate the technical information received (expected vs. actually received?	65	35				100

Table 3
Practical Sessions

Questions	Excellent %	Good %	Regular %	Bad %	Very Bad %	Total %
How was the relationship between lectures and laboratory practices?	75	25				100
How was the organization of the laboratory activities?	40	40	15			100
How helpful was the lab manual for the development of the laboratory practices?	60	40				100

Table 4
General Evaluation

Questions	Too Long/ Just Right %	Right %	Too Short/ Too Little %	Total %
How would you rate the duration of the course?		70	30	100
How would you rate the amount of information and material given?		95	5	100

NCBA
National Cooperative Business Association

The National Cooperative Business Association
1401 New York Avenue, N.W., Suite 1100
Washington, D.C. 20005-2160
Telephone: (202) 638-6222; Fax: (202) 628-6726