THE ROLE OF SENSORY ANALYSIS IN QUALITY CONTROL

JUNE E. YANTIS
EDITOR

ASTM

The Role of Sensory Analysis in Quality Control

June E. Yantis, editor

ASTM Manual Series: MNL 14

1916 Race Street ■ Philadelphia, PA 19103

Library of Congress Cataloging-in-Publication Data

The Role of sensory analysis in quality control/June E. Yantis, editor.
(ASTM manual series ; MNL 14)
"ASTM publication code number (PCN): 28-014092-36"—T.p. verso.
Includes bibliographical references and index.
ISBN 0-8031-1486-9
1. Engineering inspection. 2. Quality control. 3. Sensory evaluation. I. Yantis, June E. II. Series.
TS156.2.R64 1992
670.42'5—dc20 92-23451
CIP

ASTM Publication Code Number (PCN): 28-014092-36

Printed in Baltimore, September 1992

Foreword

This publication on the sensory evaluation function within manufacturing quality assurance/quality control (QA/QC) programs was sponsored by ASTM Subcommittee E18.07, a subcommittee on Quality Assurance, of ASTM Committee E-18 on Sensory Evaluation of Materials and Products. The scope of the subcommittee was to identify and recommend procedures for using sensory analysis in QA/QC functions.

Many people have contributed ideas for this manual from their actual work experience. Sensory programs vary with the diversity of products, size, and needs of individual plants and companies. Because there are few "standard procedures" for sensory testing in QC programs, the information presented in this document is intended to be used as reference material for developing sensory test programs appropriate to the ideas and needs of individual QC groups. Supplemental references related to QA procedures and sensory methods are listed in the bibliography.

Acknowledgments

Special thanks go the editorial contributions of Dick Jones, John Powers, and Morten Meilgaard. Many members of ASTM Committee E-18 contributed ideas and examples, many of which were used in this publication. Contributors over the years of developing this guide include Marjorie Albright, Joan Stouffer, Barbara Booth, Gail Civille, Margery Einstein, Vivienne Drimmer, Suzzanne Whitlock, Gregg Wilcove, Carol Gardze, Darla Simpson Hall, Jeanne Speight, Sally Hite, and Mark Land.

Related ASTM Publications

Manual of Sensory Testing Methods, STP 434 (1968), 04–434000–36
Guidelines for the Selection and Training of Sensory Panel Members, STP 758 (1981), 04–758000–30
Physical Requirement Guidelines for Sensory Evaluation Laboratories, STP 913 (1986), 04–9130001–36
ASTM Manual on Presentation of Data and Control Chart Analysis, MNL 7 (1990), 28–007089–34

Contents

Preface

This manual on *The Role of Sensory Analysis in Quality Control* describes general procedures and gives background information regarding the use of sensory testing as part of a quality control function in a manufacturing plant. Chapters 2 through 4 are intended to help those readers who are new to the sensory field, with a need for background material on establishing sensory testing in a plant. Those experienced with plant quality control may wish to go directly to Chapter 5 for sensory testing applications.

TERMINOLOGY AND DEFINITIONS USED IN THIS PUBLICATION

1. *Quality Assurance*—All those planned or systematic actions necessary to provide adequate confidence that a product or a service will satisfy given needs [*1*].

   Discussion

   - As a function of corporate management, quality assurance sets the policies, systems, programs, and procedures to be carried out by quality control.
   - Quality assurance defines the function of quality control and its programs and procedures.
   - Historically, quality control has permitted certain percent defectives. QA aims at achieving lower defect levels.

2. *Quality Control*—The operational techniques and the activities that sustain quality of product or a service that will satisfy given needs; also the use of such techniques and activities [*2*].

   Discussion

   - Quality control, as a function closely aligned to the manufacturing process, implements the quality specifications for raw materials, intermediate products and finished products as established by quality assurance.

3. *Sensory Evaluation*—The analysis of a substance(s) through the use of any or all of the senses [*3*]. 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.

   Discussion

   - Sensory evaluation measures perceived product characteristics, using one or more people as measuring devices.

REFERENCES

[*1*] ANSI/ASQC Standard A3–1978, American National Standards Institute, Inc., 1430 Broadway, New York.
[*2*] American Society for Quality Control, 230 W. Wells St., Milwaukee, WI.
[*3*] "Minutes of Sensory Evaluation Division Business Meeting at 35th Annual Meeting," Institute of Food Technologists, Chicago, IL, 1975.

Introduction/Program Objectives for Plant Sensory Function in QA/QC Programs

1

INTRODUCTION

Consumers are becoming more aware of taste as well as nutritional qualities of the foods they consume. As a result, perceived quality has had a growing influence on product marketing in recent years. The responsibility for creating highly acceptable flavorful foods is being transferred from the person doing the meal planning to the manufacturer of the food, as more prepared microwavable meals become available. The person doing the cooking is now part food preparer and part purchasing agent for the family meals.

Branded food and personal care products are often differentiated on the basis of certain sensory product characteristics, and these characteristics may determine whether or not a product is repurchased. Extensive and expensive market research and consumer testing is sometimes conducted before introducing a new product on the market. Predictions of market share are based on the acceptance results of the product that was tested. All of these factors make it is easy to understand why maintaining key product characteristics of manufactured products has been gaining in importance in the past few years. As a result, finished product specifications may now or may in the future include sensory characteristics in addition to chemical and physical measurements.

The quality assurance (QA), marketing, and research and development (R&D) groups in most successful companies know the sensory characteristics of their products. They also know which of these characteristics are important to their customers. Key product characteristics are usually identified during the development, testing and marketing of a new product. With the increased emphasis on knowing customers' wants and needs, defining the important characteristics of manufactured products is an ever growing responsibility of management, marketing, market research, and R&D.

Once these characteristics have been established, they must be maintained. The maintenance of product quality and uniformity is the responsibility of the quality assurance/quality control (QA/QC) groups within a company. These groups have mainly dealt with physical or chemical testing or with product safety in the past. The idea of tasting or smelling incoming ingredients, intermediates, and finished products is relatively new to manufacturing quality control. The idea of using procedures for measuring perceptions by human subjects may also be foreign to people accustomed to chemical or physical testing. Once these hurdles are overcome, the value of sensory testing in a QC program can be appreciated by the QC manager.

The sensory testing part of a QA/QC program will vary with the manufacturing process. It might cover only incoming and stored raw materials. It might also include in-process and finished goods. Like any other objective physical or chemical measurement, it can be useful

in identifying problem areas and averting or coping with product recalls. It can also contribute to profitability by avoiding manufacturing costs of products in which off-flavored, or otherwise unacceptable raw materials, are used.

This manual is intended for use in planning and implementing a plant sensory analysis program oriented to providing data appropriate for QA/QC applications. It is not intended to be a sensory methods or statistics manual. Rather it is a guide to adapting sensory methods to every day quality control situations. Procedures that might be appropriate for a large manufacturing facility producing the same product for weeks or months on end will be quite different from procedures used by small plants making short runs of multiple products.

Information on starting a sensory program where one does not exist is presented in Chapters 2 through 5. Chapters 6 through 8 contain information for the reader that already has a sensory test program that only needs reorientation so numeric data become part of the QC format.

PROGRAM OBJECTIVE

The primary objective of a sensory program included as part of the QA/QC function should be to measure the degree of conformance of sensory characteristics of products to target sensory specifications or quality standards. In other words, how does the product meet the company expectations with respect to sensory qualities?

A secondary objective of a sensory program included as a part of some QC programs might be to assist in total quality management (TQM) programs, so that products in nonconformance can be detected in-process, not as finished products. Along with the TQM programs, the sensory group may also participate in product trouble shooting.

BASIS OF A SENSORY PROGRAM FOR QA/QC WORK

An effective sensory program for QA/QC work in manufacturing plants should be based upon the following:

1. Sampling programs and test procedures that are cost effective, feasible within personnel and time constraints, and integrated with other QC sampling.
2. Test methods that measure critical product attributes and provide results that can be presented in a QC format.
3. Product attributes identified in well-written, clear sensory specifications.
4. A clear understanding of the product attributes by the personnel testing them.

Note the emphasis on the word "attribute." Using product attributes and having attribute target levels identified takes the sensory quality test program out of the realm of subjectivity to judging whether the product is "good" or "bad." It is important to position sensory testing for QC as an objective scientific measurement, similar to moisture or sugar analysis, with appropriate care taken in performing the test.

The sensory test program should provide actionable sensory test results presented in an easily understood format, reported in conjunction with other QC analyses and production data if possible. The test results may or may not have statistical support, depending on the test program, size of the plant, and decisions made with the results.

There are often two levels of sensory evaluation in a plant situation, especially when

online evaluations are conducted. Online evaluations are usually done by one or more people who have been trained to look for key product characteristics and are familiar with the product because they see it on a daily basis. This online evaluation is a "safety net" resulting in a similar to target, OK/not OK, place on hold, or reject type of decision.

When a product is placed on hold, there are more decisions to be made. Quite likely, it will be looked at by a larger group of people. If the problem involves a large quantity of product, or even a small quantity of high-value product, it may ultimately be tested by a larger group of judges or by a sensory panel at the corporate research center. The results of retesting, either at the plant or elsewhere, will be used to decide whether the product is to be sold, re-worked, or scrapped.

For online testing, efforts should be made to record the results in some numeric form. Recording results, whether in numeric form or as verbal comments, can provide the following benefits:

1. More care in evaluation is usually taken if the results are logged into some form of record.
2. Retention of online evaluations can be useful in providing information on product quality "drifts."
3. Examining evaluation results can provide a record of the performance of the personnel judging the products, providing motivation to the line people.

Providing a check form can serve as a reminder to test the product and as a guide for key characteristics to be evaluated.

Whether obtained from line personnel or taste panels, numeric sensory data form the basis for making decisions on how far a product is from target and whether or not it should be sold. QC formats usually require quantitative data, and moreover if sensory analysis is to be part of a QC program, the tests should be designed to provide some form of quantitative numeric data.

In considering numeric data from plant sensory testing as part of a QC program, it is important to remember that the values are often the results of evaluations by as few as four to six people. As a result, sensory data will differ from engineering data or even chemical analysis data, which often include multiple measurements of many samples. The results of the chemical and physical analyses will often have relatively smaller standard deviations than the results of small sensory test groups. The size of the sensory panel and other available resources must be considered before attempting to adopt recommended practices in QC reference books, especially as they relate to statistical sampling procedures and sample numbers.

The objective of sensory testing in a plant is to provide a range of information to meet the needs of the QC program. This information can include accept/reject results and data from many kinds of sensory tests on raw materials, intermediates, and finished products. The data can be presented in tables or graphs, either as a separate sensory report, or included with other chemical or physical test results. Because plant situations vary greatly, the person responsible for sensory testing must consider his or her own situation in selecting sampling and test methods. Chapters 6, 7, and 8 address QA/QC sensory data product specifications for plant sensory testing and suggest ways to adapt sensory test methods for plant QC purposes.

2 Background Information for Setting Up New Plant Sensory Programs

WHERE A PLANT SENSORY PROGRAM FITS WITHIN A COMPANY

The information presented here is a summary of several discussion groups comprised of ASTM members working in plant sensory programs, either as sensory test leaders in manufacturing plants or as coordinators of plant sensory as part of corporate sensory or QA. Actual experiences and practices were shared with the objective of helping someone starting a new program in a manufacturing plant.

Company structures vary greatly, both in size and numbers of plants manufacturing different products. As a result, plant sensory programs may have several different reporting relationships. Examples of some groups to which plant sensory programs might report are as follows:

1. Quality Assurance/Quality Control (QA/QC).
2. Research and Development.
3. Corporate Sensory.
4. Operations/Plant Management.

Each of these reporting relationships are discussed below.

Reporting to QA/QC

The scope of quality assurance and quality control varies in different companies. In large companies they are two separate units. In smaller companies the two functions are often combined into one group. To clarify where plant sensory might fit into these groups, QC and QA are defined.

QC is responsible for executing procedures to measure the specified quality of raw materials and in-process or finished products. These procedures are part of the written product specifications. QC is responsible for rejecting or placing on hold those products and materials that are not in compliance with established specifications. Sensory quality may be part of the product specification and may be a basis for rejecting or placing on hold. The sensory specifications and test procedures may be written by R&D or QA. The plant sensory group implements the test procedures and may have input into the writing of the test procedures.

QA is responsible for developing policies and programs to assure that the product is manufactured to a standard of uniformity and sets compliance standards for the uniformity

of the target or standard. QA, or QA in conjunction with R&D, often identifies the target and acceptable ranges for key product characteristics. The QC test procedures may be written by QA or by other groups, such as R&D.

QC in a plant is generally responsible for the physical, chemical, and if appropriate, the sensory testing of raw materials and finished products. The size of the QC staff, or the specific sensory group if there is one, generally determines the extent of sensory testing of ingredients and finished products. The sensory test activity within the QC group will also vary relative to the sensitivity of the ingredients and how they affect the finished product. Another influencing factor is the extent to which sensory attribute variation affects consumer acceptance.

The QC group within the plant may also be active in training TQM teams in spot testing for chemical, physical, and sensory attributes. This work is also often coordinated between the QA and QC groups.

Summary of Advantages and Disadvantages of Reporting to QA/QC

Having sensory testing as part of QC in a plant has several advantages. The reports on a product can combine all chemical, physical and sensory test results in a single document, which can facilitate communication of results. Sampling can be more efficient if both chemical, physical, and sensory samples can be collected at one time. Often QC personnel are assigned to a product rather than a specific chemical or physical test. In these cases, the sensory testing can be coordinated with the other QC tests on the product. As part of QC, sensory tests can be effectively used with other testing in trouble shooting.

Sensory testing as part of QC can have disadvantages. It is generally less well understood than chemical and physical test procedures and can be relegated to the last test performed. With limited time, this can result in the testing not being done at all. QC labs are generally small and can be crowded with instruments for chemical and physical testing. Size limitations can impose severe limits on space for serving samples to panelists.

Plant Sensory Reporting to Research and Development

The sensory plant QC program may relate to R&D in the areas of new product development, scale-up to production, cost reduction, and process optimization. This is especially true if R&D is located in close proximity to the plant. Feedback to R&D and purchasing regarding ingredient variation or performance can be valuable as a follow-up to an initial plant run, which is usually manufactured with a single ingredient sample. Sensory evaluation of ingredients at the plant may be useful in revision of specifications if needed.

Advantages and Disadvantages of Reporting to Research and Development

There are advantages to having the sensory program in the plant reporting to R&D. By reporting to R&D, especially with a new product, the plant sensory team can develop an understanding of the product by working with the research group. The plant sensory group can become familiar with the product as it was originally manufactured under R&D supervision, which will ensure that the sensory characteristics will be maintained. The plant sensory team can also track ingredient variations and report these to the R&D and purchasing teams.

There are also disadvantages to having the sensory program in the plant reporting to R&D. Unless there is a close relationship between R&D and plant operations, the sensory

group reporting to R&D but located in the plant may inadvertently not be included in the total picture.

Plant Sensory Reporting to Corporate Sensory Group

The plant sensory group might report to the corporate sensory group, especially in large multiplant companies. The corporate sensory group can provide technical support, training, and new methods research for the plant group. Especially in multiplant situations, the corporate sensory group can coordinate testing among plants to improve uniformity in testing among plants manufacturing the same product.

Advantages and Disadvantages of Reporting to the Corporate Sensory Group

There are advantages of having the plant sensory person/group reporting to a corporate sensory group. A person hired to conduct sensory testing in a manufacturing plant can feel overwhelmed or lost with limited reference material or technical support. Having a group to call when quick answers are needed is helpful. It is also an advantage to have the same tests conducted on products manufactured in several plants to facilitate summary of information by corporate QA.

There are also disadvantages. Unless the coordinator in the corporate sensory group has had experience in plant sensory work, there can be resentment on the part of the plant group being told what to do by someone who is "going by the book." The corporate sensory group must be conscious of the possibility of being viewed as "high and mighty," especially by a relatively junior person in an entry level QC position in the plant.

Plant Sensory Reporting to Operations/Plant Management

A plant sensory program might report to plant management or operations, especially in small to midsize companies. This reporting relationship is especially effective when sensory characteristics are key factors in product quality and test results are used by operations on a continuing basis. Being part of production, plant sensory can play a useful role in product troubleshooting.

Advantages and Disadvantages of Reporting to Operations/Plant Management

A reporting relationship to operations and plant management can be very effective, provided that a "team" relationship is maintained. If the sensory group is differentiated from the QC group by reporting directly to plant management, there is a better chance of establishing sensory testing as part of manufacturing and not being put into the background of a larger QC program.

There are disadvantages to the sensory program reporting to Operations or Plant Management. As with any QC activity, there are usually pressures placed on groups reporting to Operations to approve the shipment of products that are of marginal quality.

In summary, the ideal reporting relationship will result in the plant sensory group being independent and able to present unbiased test results. Companies and people managing QC/QA, R&D, and plant operations vary greatly. Where plant sensory will best fit and how it can be successful is difficult to generalize.

HOW SENSORY TESTS CAN FIT INTO CURRENT QC PRACTICES

Sensory testing can play an important role in maintaining product quality, but it may be one of the last programs to be established for QC. It is often initiated because a product problem has occurred or a need has arisen for control of a critical flavor or other product characteristic. In any case, objectives and ways of achieving them must be identified before a sensory program is initiated. For example, a critical flavor or texture may have to be controlled to maintain either customer (in the case of the product being an ingredient sold to a manufacturer) or consumer acceptance. To establish a control program, methods of measuring the characteristic must be identified and limits of variability around the target level must be set. If sensory testing is being done in response to a product problem, it is important to establish what product problems exist and what kind of information is expected from a plant sensory program. Sources of this type of information are plant, marketing and sales personnel, company management, and possibly customers (especially review customer complaint letters if available).

It is important to remember that sensory testing differs from chemical or physical testing and is often poorly understood by people responsible for overall quality control. Unless specific objectives and priorities are set, the role of sensory testing can disintegrate into casual like/dislike tasting by plant management personnel with minimal sensory science input. It is also important to remember that sensory testing must be cost justified and reported in an easily understood manner to plant management.

CONSIDERATIONS IN SETTING UP A SENSORY QC PROGRAM

Assessment of Existing QA/QC Practices

A review of QA/QC practices currently used for chemical/physical testing is essential in setting up a sensory program. Such a review might include the following:

1. What sampling procedures exist for raw materials, in-process products, finished products, and packaging materials.
2. How priorities are established for sampling and testing critical materials.
3. What conformance and testing criteria have been established between the company and raw material suppliers.
4. How references and product retain samples are maintained.
5. What criteria and procedures exist for rejection of raw or intermediate materials or placing finished products on hold.
6. What critical time has been established for the rejection/on-hold decisions resulting from Step 5.
7. How many product specifications exist, and how many of these have sensory specifications.
8. What sensory testing has been conducted in the past, how were data reported and recorded, and how were the data used.

Assessment of Product Quality Records

A history of the kind of product defects and their frequency may help to target the proposed sensory program. Suggested information sources are as follows:

1. Past customer complaint records.
2. Plant reject experience with raw materials and finished products placed on hold, including why they did not comply with current QC specifications.

The customer complaint records can tell the sensory person whether there might be a pattern or trend with sensory problems, such as seasonal variations in water sources, or a humidity change in the packaging room. Information on what raw materials may have been rejected in the past, and for what reasons, can help in establishing priorities regarding what should be reviewed first. If raw materials have been rejected for broken containers, then the issue is not directly related to sensory, even though the material may have had off flavor characteristics. If the raw materials are within chemical and physical specifications, but the finished product still does not taste or smell right, then there is a need for a sensory test to be identified and implemented. Sensory testing is most successful for QC programs when it can provide answers not readily available from physical or chemical tests.

Assessment of Manufacturing Process

A review of the manufacturing process will provide necessary background information for starting a sensory program. Some questions to serve as a guide in this review are as follows:

1. Are the processes batch or continuous?
2. What are the control capabilities of the processes?
3. How are incoming raw materials planned, purchased, and delivered?
4. How are incoming raw materials recorded, stored, and rotated?
5. What are the possibilities for contamination or changes caused by physical conditions, such as odor transfer or heat/cold conditions, during storage before use?
6. How are in-process intermediates handled and stored?
7. How is the finished product tested and certified before shipment?

An understanding of the manufacturing process is essential to a person responsible for plant sensory testing. Testing ingredients is important, but equally important is how they affect the final product, and how the process changes the ingredients. An understanding of the sensory properties of the final product manufactured at target and at the outer limits of the process can provide suggestions to the sensory person as to whether the incoming ingredients or in-process products should be given priority in the test program.

Examining the history of the use of ingredients, assuring that they have been rotated properly and not damaged by freezing/thawing or overheating, is another important area to review. Looking at an ingredient warehouse may not show obvious problems, but an understanding of routine handling and storage of raw materials can be helpful if a problem does arise. An example of this is odor transfer from printed packaging materials, or the use of old raw materials because of a mistake in stock rotation. The manufacturing process has many random variables, and the more known about it, the more effective a sensory person can be, especially in trouble shooting.

Assessment of Current Sensory QC Program Dimensions

There may have been a sensory program in the plant at one time, or one may exist at the present time. In reviewing a program, the first thing to determine is whether the size of the

program matches the requirements for sensory testing in the plant. The size of the program will depend on many factors, some of which are as follows:

1. The sensitivity of the raw materials used, including packaging materials.
2. Intermediates produced or received from another supplier.
3. The production schedule, which may need testing over three shifts for a critical product, especially one with continuous production of intermediates.
4. The number of stock keeping units (SKUs) of finished goods produced at the plant.
5. Space and personnel available.

With current emphasis on being competitive and reducing manufacturing costs, the space and personnel available will usually be limiting factors, without even arguing the need. There are never enough resources to do everything that is needed. As with most other testing programs, the sensory testing will be expected to provide a payback in decreased reject product, and expansion of testing will have to be cost justified.

SUMMARY

The information gathered on the raw materials, manufacturing, and QC practices should form the context for the development of a sensory program. Some key questions to ask while developing the program are as follows:

1. Is there a physical or chemical test that answers the same quality questions? If so, use it instead of a sensory test. Why? Because more samples can be tested within a given amount of time and generally the testing will be less costly. For example, if an incoming batch of oil is darker than the color specification limit, there is little point in setting up a panel to taste it. However, gas chromatography cannot always point out small peaks that might be from flavors that could alter product quality. These could be noted by a sensory panel.
2. Is there a reasonable way to maintain a standard that is representative of expected quality of either an ingredient or finished product? If so, by all means use it to anchor the sensory test results to something that everyone understands. Written descriptive sensory specifications are meaningful to the person writing them. There must be a good understanding of those written descriptors by the person at the plant for plant sensory testing to be effective. To avoid misunderstanding, participation in the writing of the descriptors or providing a physical standard to the plant sensory person can be helpful. Standards are mainly useful if the product is reasonably shelf stable, such as dry packaged mixes, canned products, and cereal products. One example of using a standard with a target color and flavor and with acceptable ranges is a product such as chocolate ice cream topping.
3. Is there a way that the sensory person can become familiar with how quality management decisions are made? If so, give serious consideration to producing test data that will be useful to these decisions. Understanding how the data will be used is a key factor in designing the sensory test program.

3

Resource Assessment

INTRODUCTION

Just as chemical testing requires people, reagents, equipment, and a physical space to conduct the work, sensory testing has resource requirements as well. These requirements include people and a physical space in which to work. The physical space usually includes a need for a quiet, odor free place for people to taste samples and a place to prepare them. Instead of reagents and analytical equipment, sensory testing requires uniform sample preparation and presentation, and it needs a supply of panelists to taste the samples. Like chemical or physical testing, sensory testing also requires people trained in the methods to be used. The staff, the panelists, and the physical space are all resources for the test program, and will be discussed in this chapter.

SENSORY STAFF REQUIREMENTS

The education and experience required for a person doing sensory testing at a manufacturing site will vary with the size and complexity of the plant and products. Additional staffing needs are determined by the number of tests per shift, the extent of product preparation required, and the number of panelists used per test. In small plants, the sensory testing is part of plant QC and is often conducted by the same people who do the chemical and physical analyses. This is true even in some larger multiproduct plants, in which personnel are assigned broad responsibility for one product, rather than conducting a narrow group of tests on many products. The sensory staff must be cost justified, and it is best to start small and demonstrate the benefits of sensory testing, rather than start with a more ambitious program that will be cut if funds become short.

It is usually desirable with multiplant programs to have a sensory person or staff at each location, with a coordinator to manage the overall program. As indicated in Chapter 2, having a coordinator (usually part of a corporate sensory group) has the advantage of a uniform test program with technical support for the plant sensory staff. There is a potential for disagreements between plants and the plant sensory coordinator. The coordinator's tasks include monitoring performance uniformity across plants and, to perform these tasks successfully, the person should have exceptional interpersonal and communications skills.

SENSORY PANEL JUDGES

The type of panel judge needed will vary from plant to plant and product to product. The extent of preparation required will influence where and how the product is served, and who can be used as panelists. For example, if the product must be prepared and served within a prescribed length of time because of changes on standing (cooked pasta, for example), special arrangements must be made to have judges available at a specified time.

Panel judges can be recruited from two primary sources: employees of the company or people outside of the company.

Employees of the Company

The use of all plant personnel for testing provides a large, multishift pool of panelists. Plant personnel can be used if there is flexibility in scheduling. Using union personnel as panelists may require that product testing be defined as part of their tasks. Panels can be arranged near shift changes for convenience of the workers. A factor to remember, if the plant process has a high odor level, it is best to have the people come to panel before starting their shift.

Plant line personnel can also be trained to make meaningful observations online, including "safety net" sensory testing. A vital part of any sensory program is involvement of the company personnel. Reported observations by line employees should be encouraged. Their observations may be more useful and actionable than observations of expert panels or graphs and statistics. These activities can be a valuable part of a total quality program.

Often only the people on the plant support staff (such as accounting, QC, purchasing management, and so on) are available as panelists. The use of the support staff may limit the testing to the day shift. The size of the staff will also influence the kinds of testing that can be done. One of the benefits of conducting sensory tests with office or plant employees is more personal involvement with the manufactured product resulting in heightened awareness of product variability and the importance of product quality.

Participation by the sensory staff and technicians in QC should be limited to tests in which they have no knowledge of the sample identities. Often, the sensory and QC staff are trained to make go/no go decisions on certain incoming ingredients.

Nonemployee Panelists

Another source of sensory judges is outside (nonemployee) panelists. The recruiting of noncompany panelists is recommended if the sensory facility is already off site. Outside panelists may also be needed if the testing requires larger numbers of tasters than are available from within the plant. Outside panelists can also be used to supplement plant personnel if the sensory facility is at the plant. Recruiting can be conducted through advertisements in local papers or through civic and church groups. Safety and security are issues to be considered in bringing nonemployees into a plant facility. Payment of outside panelists is another factor. While employees are not "free," their costs may be more transparent than the need to establish a separate budget to pay panelists.

While less convenient than working with support staff or plant employees, the results of tests from outside panelists do have more face validity because employees are often seen as having a bias toward the company product.

An assessment of people who might serve as panelists must be done in the planning stages. Matching numbers of people with tasks under consideration is essential. For example, if only an office staff of 8 to 10 people is available, then the methods chosen must be slanted

toward descriptive or quality scaling rather than difference or other quantitative tests. If factory personnel are available, larger tests can be conducted. Another consideration is the number of people on the staff. It might not be feasible for a small staff to serve samples to 100 panelists, even if they are readily available. Thus the panel size, sample preparation, sample serving, data handling, and reporting must all be balanced before a sensory test program can be planned.

Considerations in Screening and Establishing a Pool of Panelists

There appear to be many misconceptions about requirements for judges in sensory tests, both from the side of those setting up tests and from those being recruited as judges. Some people believe that judges must be screened for taste and smell acuity, and that thresholds must be established for each judge. While this may be true for certain specialized descriptive panels, it is not the case for most plant QC testing. The most important factors in selecting judges for QC test panels are interest, willingness, and availability to participate. More people are lacking in interest than in the ability to differentiate between sensory characteristics of a set of samples.

Most newly recruited people come to the first panel convinced that they cannot taste at all, and that they would be totally useless in evaluating a series of samples. The first task of the sensory staff is to build interest and confidence in the panelist pool.

It is important that the sensory staff use common sense in recruiting judges. Their interest and alertness should be considered more important than whether they appear to be "gourmet" cooks, wine connoisseurs, or otherwise "superior tasters." The staff should conduct introductory sessions to build confidence among the prospective panelists. These sessions should include tasting samples of different brands of packaged products, such as cookies or crackers, and building a vocabulary to be used by the panelists to express their sensory perceptions.

Judge Training

Usually the training of panelists for plant sensory work is task oriented. In contrast to corporate sensory groups, who often have some form of formal, somewhat standardized profile or descriptive training, plant sensory judges are trained with the target product and with acceptable and unacceptable variations. There are two general areas to address:

1. Familiarization with the Target Product—The judges must be assisted in recognizing target product attributes, both qualitatively and quantitatively, and in developing a vocabulary to describe them.
2. Familiarization with the Test Questionnaire—The judges must understand the questions and the use of any scales utilized in the test.

Most panelists will become very proficient after a brief training period and some practice. It is advisable to continuously add a few people to the judge pool because of the inevitable loss of panelists over time. People given an opportunity to taste products with more experienced panelists on several occasions tend to "self train," and will perceive more differences with each experience. By adding a few people and allowing them to participate and practice without using their data for a few sessions, a stable panel for testing can be maintained without the need for extensive training sessions.

The panel is the analytical instrument used for sensory testing. As with any analytical instrument, the preparation and fine tuning of the panel must be balanced with the complexity and detail desired in the result. Panelist preparation can be relatively simple if the task is a simple pass/fail test of a final product, or it can be extensive if a small descriptive panel is required.

People new to sensory testing are usually apprehensive and need some basic orientation to the task of tasting samples. Inexperienced panelists usually feel that they "cannot taste." Providing them with simple sugar and salt solutions at two or three levels can demonstrate that they can perceive the tastes and that they can perceive differences in levels. Many other available foods can demonstrate differences in flavors, textures and tastes appropriate for orienting people to the specific task to be done. At the least, a panel should practice with rating scales for degree of difference and attribute scaling tests which will be discussed in Chapter 7. They should also be provided with some idea of the range of variation they might expect to see.

SENSORY TEST FACILITY

In addition to the sensory personnel and panelists, a place to conduct the tests must be selected. The size and location of the test facility will influence the scope of the sensory program. The primary consideration is the ability to provide a controlled, comfortable, quiet, and odor-free environment for the tests. In small plants, some space in the QC lab is often all that is available. Tests can be conducted in very small areas, assuming care is taken to provide adequate lighting and privacy for the individual assessments of samples to be done. In some plants, a conference room can be scheduled for a specified period of time for conducting tests. A suggested source of information on sensory test facilities is *STP 913, Physical Requirements Guidelines for Sensory Evaluation Laboratories* [*1*] (also, refer to list of related ASTM publications on p. iv of this manual.)

4 Program Implementation

INTRODUCTION

Often plant sensory programs are initiated to "put out a fire." If the program is shown to be effective, it can be maintained so that it can "prevent fires." This chapter addresses what to do when assessment of the situation, study of the manufacturing process, training the staff, and arranging for judges and space for testing is near completion.

FACTORS TO CONSIDER IN SELECTING WHAT TO DO FIRST

If the program was started to "put out a fire," the task is predefined. If it is not predefined, or the program is new, the following are ways to select tasks to accomplish the objectives of the program:

1. Meet the most important needs first. This may not be the easiest, but it will demonstrate effectiveness of the program most clearly. Ask whether the question can be answered as readily by a chemical or physical test. Success will be more likely if a sensory test is the only way to detect conveniently and reliably the attribute flaw in question.
2. Establish some success by doing easier tasks first. Suggestions might be evaluation of an ingredient against an identified target or establishing a record of differences in key attributes. Having an identified target as a "crutch" can help in building the confidence of the panel.
3. Select a method that fits the resources and apply it to the most important question that can be answered by the method. It often is more practical to start small and demonstrate the value of plant sensory testing than to start with a grand plan that conforms to a textbook, but is too costly to support.

PROGRAM REVIEW

Several months after implementation, the program should be reviewed to determine how effective it has been. The defined program objectives can be used to measure how it is working:

1. Have the consumer complaints been reduced?
2. Has rework been reduced?
3. Has the cost been in line with what was projected?
4. Have the results been reported in a timely, useful fashion?

The importance of sensory information is related to effectiveness, which can be measured by decreased product variability, increase in perceived quality, more product shipped, less product on hold, and fewer complaints. A valuable, but less tangible result, is increased employee awareness of product quality.

SUMMARY

The background information in the first four chapters has been provided to prepare the reader for the next four chapters, which cover selecting tests and ways of reporting results for use in plant quality control programs.

5 Data Presentation for QC Programs

INTRODUCTION

An understanding of data presentation for QA/QC use will be helpful when considering and selecting methods for plant sensory applications. For this reason, data and ways of reporting results are being presented before application of methods. Adapting sensory methods to provide numeric data will be covered in Chapter 7.

Compatibility with accepted QC formats requires that sensory tests provide quick actionable results and be oriented to producing numeric data, that is, numbers rather than words. For sensory data to be expressed in numeric form, the tests must be designed so the results are obtained as numbers, rather than pass/fail, or a paragraph of descriptions.

This chapter will address ways of presenting data, ranging from very simple graphs to control charts. Control charts are widely accepted as a means of reporting ongoing quality control. They have been adopted for reporting yields, percent defectives, package weights, and other manufacturing records across many industries. Even sensory data can be presented in control chart form with proper planning. Simple control chart analysis will be briefly described here, relative to sensory data presentation. An excellent and very readable reference on control charts is the ASTM *Manual on Presentation of Data and Control Chart Analysis* (*MNL 7*) [2]. The reader is referred to Refs *3* to *10* for further information on calculating upper and lower control limits for control charts. This manual is not intended to be a treatise on statistical methods or applications. It is intended to introduce some simple approaches to data handling and reporting for a person working in plant QC with limited staff, judges, samples, and data.

DATA PRESENTATION

A great deal of sensory testing conducted at a plant level is done using small trained descriptive groups, with or without replication. Much of the data reported from these tests can be reported as means, with ranges or standard deviations or standard errors. These data can be plotted as points on a line graph across time or batches, or as bar graphs.

A consensus summary can also be presented. Consensus summaries are produced by having the panel members evaluate samples individually on some sort of quantitative scale, and then, following discussion, an objective panel leader develops a consensus value with the panel participants.

Whether the sensory data are means and standard errors or consensus scores, these values

can be plotted as ratings over time. Maintaining data over time, rather than as individual test results, puts sensory testing within the realm of other QC tests.

Example of Hypothetical Panel Results

An example of hypothetical panel results of testing an array of samples, relative to a target or reference sample, is presented in Table 5.1. These results are graphed in Fig. 5.1. The relationship has been defined as 1 = not at all similar and 10 =, a perfect match. A target of 7 to 10 in similarity could be defined as an acceptable zone for the product, with zones of marginal and unacceptable also defined, if desired. Note that target score similarities do not have values above and below a mean value, because the scale is unipolar.

TABLE 5.1—*Test data from a panel rating samples as similar to a reference on a 10 point scale.*

	Sample			
Judge	Week 1	Week 2	Week 3	Week 4
1	8	9	6	7
2	8	7	7	6
3	9	7	7	5
4	8	7	6	7
5	8	7	6	7
6	8	8	6	7
Mean	8.2	7.5	6.3	6.5
s	0.4	0.8	0.5	0.8

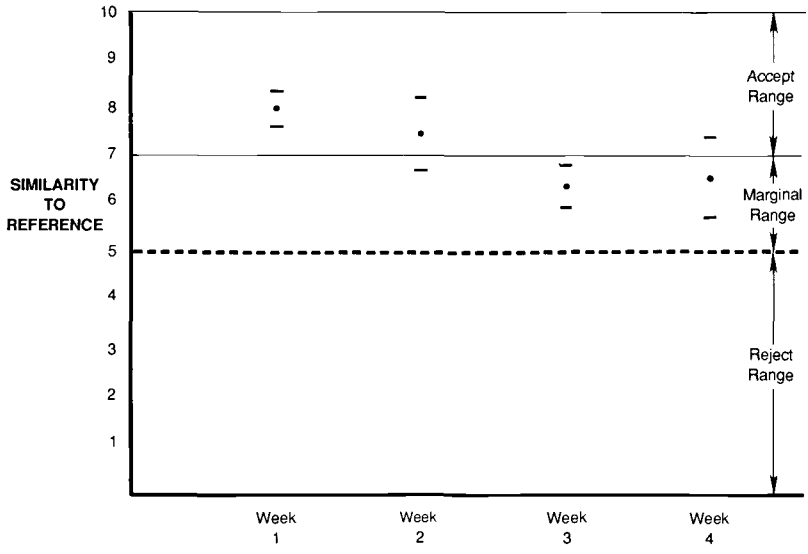


FIG. 5.1—*Graphic presentation of data from Table 5.1.*

A comparison of the data in Table 5.1 and Fig. 5.1 illustrates how sensory data can be more readily communicated if presented in graphic format, rather than in tables or as individual data points entered in a notebook. Graphic presentation need not be complex and need not require a computer. A simple graph can be prepared, and new data entered by hand with pencil and graph paper. Graphic or summary table presentation does require a reorientation of thinking about designing sensory tests so that the results are numeric instead of "it tastes OK."

Note that the type of graph in Fig. 5.1 shows the relationship of the sample to a reference over a series of samples or a given time period. Next, we will discuss how sensory data can be reported in control charts.

Adapting Plant Sensory Data for Control Chart Reporting

A control chart, or X chart, is a plot of the arithmetic mean of sample measurements. The chart is composed of a center line, or X line, which represents the mean values from evaluations of "target" product, and two other lines, one above and one below the X line. These two other lines in standard quality control chart analysis represent limits set by QA. Most reference books use three standard errors above and below the target value. With the variance often encountered in sensory testing, three standard errors could be out of the limit of acceptability. Training and experience will greatly reduce this variance, and must be taken into account and reviewed, especially with a new program. For sensory testing in QC programs, the upper and lower lines, or what are called the "upper" and "lower" control limits, can be established by QA, based on the following:

1. Marketing information, if available, on how much variation can be tolerated before consumer acceptance is adversely affected.
2. Using panel and product variance from a series of tests conducted on target product.

Depending on the product and objective of the manufacturer, determination can be made on how much more variance above panel and product variability can be tolerated based on hedonic testing of a range of products.

These "upper" and "lower" control limits can then become a part of the sensory specification for the product tested. X charts can be used for tasks, such as tracking incoming ingredient quality, following process control, or monitoring ingredients and finished products. Examples of applying X charts for expressing sensory data are presented next.

Preparing X *Charts*

An X chart for presenting sensory data presents the sample series (code dates, hour, shift, lot number, or other identification) on the x axis and the attribute intensity scale on the y axis. Figure 5.2 shows some theoretical panel data from panelists rating flavor strength of six samples on a 10-point scale. The data points are plotted around the X line, which was established at a flavor strength of 4. We see that all samples were within the control limits set, with the exception of Sample 3, which was above the upper limit.

The following two examples illustrate setting the limits by knowledge of what is acceptable and by using the variance about the mean of a set of judgments.

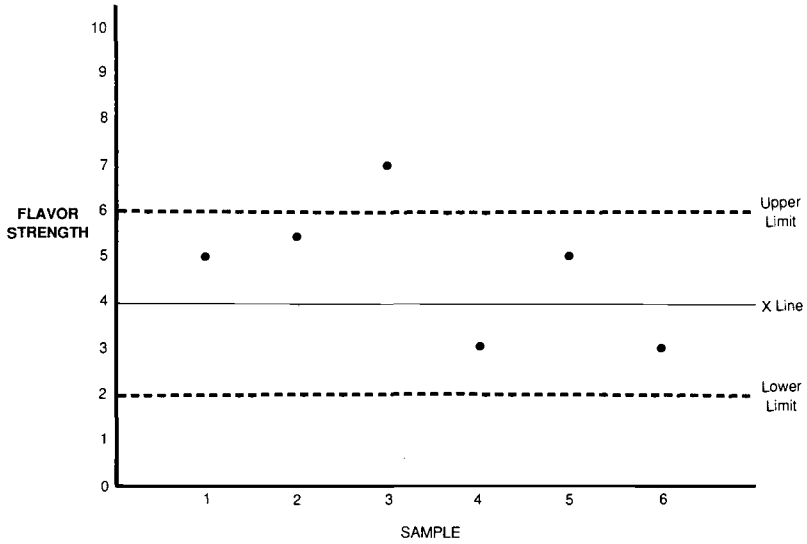


FIG. 5.2—*Example of simple control chart presentation of data.*

Fixed Control Limits

Control limits can be set by the company, based on consumer test and plant production capability, or other information. The limits may be set by marketing, QA, or other groups. The following example illustrates this approach:

A food company sells a finished product to a large restaurant chain. The QC group of the restaurant chain has set, with concurrence from their marketing management groups, an acceptable range of quality (similarity to their target product) for the product they are willing to accept to sell to their customers. They have a target product, identified as their ideal, that they supply to all of their vendors as an example of the product they wish to purchase to sell in their shops. Because they recognize that not all products will be an exact match for the "ideal," they also supply examples of products they consider to be close enough to be acceptable, and examples of products that are unacceptable. These examples of the quality range, supplied by the QC group of the retail chain, are used by the vendor's plant QC group for panel training and to set attribute target limits and ranges.

The vendor's QC panelists then rate production samples on a 1–10 scale for overall similarity to the target. Because the panel has been trained, they use a "mental" target but have a sample of the target product available for reference if they wish to use it. A series of three to four production samples can be evaluated in a single session. An example of a scoresheet for this type of sensory evaluation is shown in Fig. 5.3.

Note that the scoresheet addresses the similarity question and provides numeric data that can be presented as a panel mean. It also provides a place for the panelist to describe why the sample was not similar. These comments can be summarized and presented with the graphed results. The results of a hypothetical evaluation of a set of six samples is shown in Fig. 5.4. The graph has a fixed line, represented by the dotted line corresponding to the similarity value of 5 on the 1–10 scale. It also shows the area not sufficiently similar to the target. The graph shows that four samples were close to target: one was a "match" and one

NAME_____ **DATE**_____

PRODUCT_____

Draw a circle around the number that you feel best describes how well the sample matches the target.

How close is the product to target?

Reject	**Not close**	**Close**	**Match**
1 2	3 4 5	6 7 8	9 10

If the sample is scored a 9 or 10, stop here.

If scored 8 or less, please indicate WHY below:

Not Nearly Enough	Not Enough	Too Much	Much Too Much
❑	❑	❑	❑
❑	❑	❑	❑
❑	❑	❑	❑
❑	❑	❑	❑
❑	❑	❑	❑
❑	❑	❑	❑

COMMENTS / OTHER_____

FIG. 5.3—*Questionnaire for similarity to target test.*

was not close to target because of a burnt note. As discussed previously, with similarity to target tests there is only a lower control limit, since a perfect match would be 10, the top of the closeness to target scale.

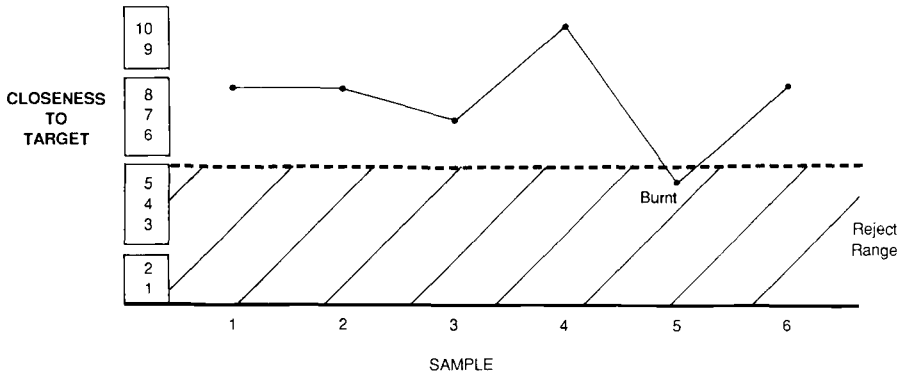


FIG. 5.4—*Graphic presentation of a set of samples relative to an established target.*

Limits Set by Statistics

Setting limits based on the variation of panel data obtained from evaluation of a number of within specification products is another way of preparing a modified "control chart" for sensory data. The following is an example of this approach.

A baking company wishes to maintain a reasonably consistent flavor of apricots used in a cake. There is a browned flavor in the apricots, which at a high level, can be considered overripe or "off" in flavor. A certain level of this flavor is desirable, and if not present, the apricots may have insufficient ripe flavor. To establish the limits, a representative sampling of tins of apricots within chemical and physical specifications is presented to a panel previously oriented to tasting apricots. The panel training consisted of providing samples of acceptable apricots with the right amount of browned flavor, and unacceptable products, representing apricots with not enough and too much "browned flavor." Following the training, ten panelists are asked to score seven blind coded "target" samples for "brown" flavor on a 1–10 intensity scale, providing seven groups of ten data points, each with a range, a mean, and distribution about the mean. This establishes the panel variation and the perceived product variation to be expected for an array of products considered to be within specification based on chemical and physical testing.

The control limits are then calculated using these groups of data by a simple formula and tables available from most basic statistics books (reprinted *ASTM MNL 7* [*2*]). As previously stated, the objective of this manual is not to provide extensive statistical background and principles. These are simple approaches to start using graphic presentation on data. The reader is referred to Refs *3* to *10* for background and principles of the procedures given here.

The panel data from the ten judges tasting seven samples of apricots will be presented in Table 5.2, along with a simple summary of what will be needed to use the formula from *ASTM MNL 7* on control chart analysis.

TABLE 5.2—*Panel data from ten judgments of apricot browned flavor on a 10-point intensity scale.*

	Sample						
Judge	1	2	3	4	5	6	7
1	4	1	4	5	5	5	3
2	3	3	5	4	7	4	4
3	1	3	3	3	5	4	4
4	3	2	5	4	1	5	5
5	2	2	2	3	6	3	4
6	4	1	4	2	8	4	2
7	2	2	3	4	6	5	3
8	2	4	4	4	6	4	4
9	3	3	3	3	5	4	4
10	4	3	3	3	2	3	5
$\bar{R}$	3	3	3	3	7	2	3
$\bar{X}$	2.8	2.4	3.6	3.5	5.1	4.1	3.8
s	1.03	0.97	0.97	0.85	2.13	0.74	0.91

NOTE: $\bar{\bar{R}} = 3.43$.
$\bar{\bar{X}} = 3/61$.
$\bar{s} = 1.08$.

(For ease in understanding the tables from *ASTM MNL 7* [2], Table 5.3 shows symbols used and a brief explanation of their meaning.)

The control limits are then calculated by a simple formula and Table 5.4 reprinted from *ASTM MNL 7* [2].

Using the simplified formula for small samples of equal size, presented on p. 55 of *ASTM MNL 7*, we can create a control chart using either the overall average standard deviation or the overall average range. Looking at Table 5.4, the first two columns provide factors for control limits, factors A_2 and A_3. The simplified formula using Table 5.4 and the central line representing the X value is presented in Table 5.5.

Thus, for our example, our control limits using the average standard deviation would be calculated using the A_3 factor for seven samples, which is 1.182. This would give us (1.08)(1.182) = 1.28, above and below 3.61; so our X line would be 3.61, our upper control limit would be 4.89, and the lower control limit would be 2.43. If we wished to use the range, we would use factor A_2, 0.419 for seven samples. This would give us (3.43)(0.419) = 1.44; so our X line would be 3.61 with an upper control limit of 5.05 and a lower limit of 2.17. Other examples of calculated control limit lines are presented in *ASTM MNL 7*, Chapter 3 [2].

Looking at the data in Table 5.2, we see that there was quite a bit of disagreement about Sample 5 in the set. The data showed a range of 7, with two of the ten judges apparently

TABLE 5.3—*Control chart symbols.*

n =	The number of observations in a sample or subgroup.
s =	The standard deviation of the observations in a sample.
$s = \sqrt{n\dfrac{(x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^3 - (x_1 + x_2 + x_3 + \cdots x_n)^2}{n(n-1)}}$	
$\bar{s}$(*s*-bar) =	The arithmetic mean of the standard deviations of a group of k samples.
$\bar{s} = \dfrac{s_1 + s_2 + s_3 + \cdots s_k}{k}$	
X, x =	An observed value of some measurement.
$\bar{X}, \bar{x}$ (*x*-bar) =	The arithmetic mean of n observed values in a sample.
$\bar{x} = \dfrac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$	
$\bar{\bar{X}}, \bar{\bar{x}}$ (*x*-double bar) =	The arithmetic mean of a group of k sample means.
$\bar{\bar{x}} = \dfrac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \cdots + \bar{x}_k}{k}$	
R, r =	The range within a sample of observations.
$r = X_{\max} - X_{\min}$	
$\bar{R}, \bar{r}$ (*r*-bar) =	The arithmetic mean of a group of k samples ranges.
$\bar{r} = \dfrac{r_1 + r_2 + r_3 + \cdots + r_k}{k}$	

NOTE: Contributed by Richard M. Jones, Philip Morris Research Center, Richmond, VA.

TABLE 5.4—*Factors for computing control chart lines—no standard given* [2].

Observations in Sample n	Chart for Averages: Factors for Control Limits A_2	Chart for Averages: Factors for Control Limits A_3	Chart for Standard Deviations: Factors for Central Line c_4	Chart for Standard Deviations: Factors for Control Limits B_3	Chart for Standard Deviations: Factors for Control Limits B_4	Chart for Ranges: Factors for Central Line d_2	Chart for Ranges: Factors for Control Limits D_3	Chart for Ranges: Factors for Control Limits D_4
2	1.880	2.659	0.7979	0	3.267	1.128	0	3.267
3	1.023	1.954	0.8862	0	2.568	1.693	0	2.575
4	0.729	1.628	0.9213	0	2.266	2.059	0	2.282
5	0.577	1.427	0.9400	0	2.089	2.326	0	2.114
6	0.483	1.287	0.9515	0.030	1.970	2.534	0	2.004
7	0.419	1.182	0.9594	0.118	1.882	2.704	0.076	1.924
8	0.373	1.099	0.9650	0.185	1.815	2.847	0.136	1.864
9	0.337	1.032	0.9693	0.239	1.761	2.970	0.184	1.816
10	0.308	0.975	0.9727	0.284	1.716	3.078	0.223	1.777
11	0.285	0.927	0.9754	0.321	1.679	3.173	0.256	1.744
12	0.266	0.886	0.9776	0.354	1.646	3.258	0.283	1.717
13	0.249	0.850	0.9794	0.382	1.618	3.336	0.307	1.693
14	0.235	0.817	0.9810	0.406	1.594	3.407	0.328	1.672
15	0.223	0.789	0.9823	0.428	1.572	3.472	0.347	1.653
16	0.212	0.763	0.9835	0.448	1.552	3.532	0.363	1.637
17	0.203	0.739	0.9845	0.466	1.534	3.588	0.378	1.622
18	0.194	0.718	0.9854	0.482	1.518	3.640	0.391	1.609
19	0.187	0.698	0.9862	0.497	1.503	3.689	0.404	1.596
20	0.180	0.680	0.9869	0.510	1.490	3.735	0.415	1.585
21	0.173	0.663	0.9876	0.523	1.477	3.778	0.425	1.575
22	0.167	0.647	0.9882	0.534	1.466	3.819	0.435	1.565
23	0.162	0.633	0.9887	0.545	1.455	3.858	0.443	1.557
24	0.157	0.619	0.9892	0.555	1.445	3.895	0.452	1.548
25	0.153	0.606	0.9896	0.565	1.435	3.931	0.459	1.541
Over 25	. . .	[a]	[b]	[c]	[d]	. . .	. . .	. . .

[a] $3/\sqrt{n - 0.5}$.
[b] $(4n - 4)/(4n - 3)$.
[c] $1 - 3/\sqrt{2n - 2.5}$.
[d] $1 + 3/\sqrt{2n - 2.5}$.

TABLE 5.5—*Simplified formulas for using standard deviations or ranges for control charts.*

Central Line	Simplified Formula Using Table 5.4
$\bar{\bar{X}}$	$\bar{\bar{X}} \pm A_3\bar{s}$
$\bar{\bar{X}}$	$\bar{\bar{X}} \pm A_2\bar{R}$

not perceiving the "browned" flavor to any great extent. Figure 5.5 shows a control chart presentation of the data. Two samples are outside the limits, one above and one below. Because of the range of responses to Sample 5, the sensory person might wish to retest the sample or examine it for uniformity. It is always advisable to look at the data that compose

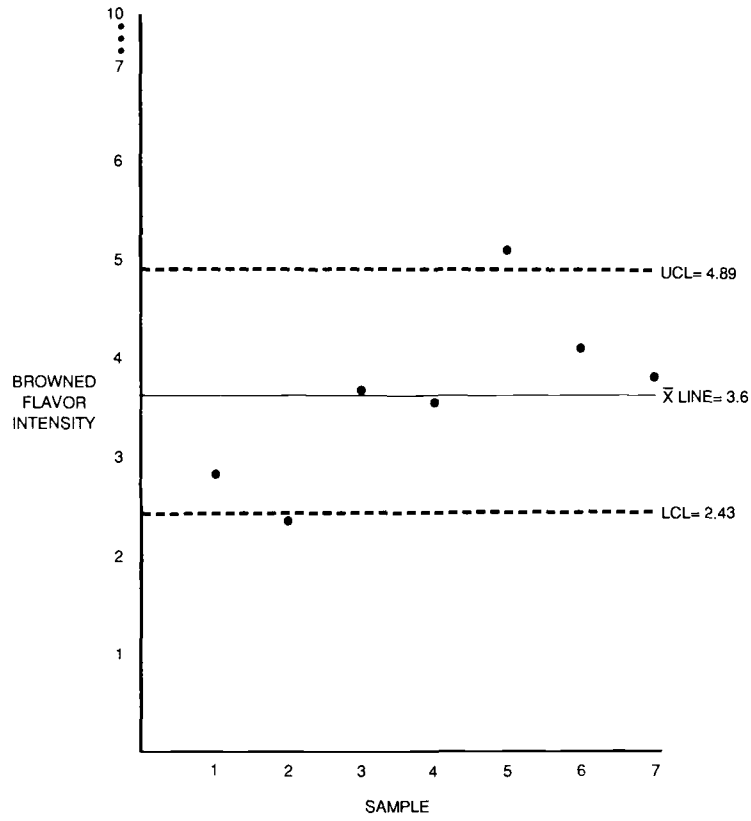


FIG. 5.5—*Control chart of apricot browned flavor intensity data from Table 5.2.*

the mean value. The data of the individual can provide valuable information on panelist performance, product uniformity, and possible confusion with the questions being asked.

In general, if the mean standard deviation is composed of five or fewer data points, it is advisable to use the range to calculate the control limits. Between six and ten, either can be used, but with over eleven data points, the standard deviation formula is preferred.

The objective of this guide is not to provide a background in statistics, but rather quick applications. The reader is urged to read *ASTM MNL 7* [2] or the references listed in the bibliography for in-depth information on control charts.

SUMMARY

Reporting the sensory data in a form similar to other QC analyses allows management to relate sensory and other QC measurements with product and process conditions and may also help in troubleshooting. When planning a test, think about the possibility of reporting and maintaining the sensory data in graph or chart form. Graphs or charts are easy to read and can show trends more readily than tables of numbers.

The Basis of Sensory Data for QC: Sensory Specifications

6

INTRODUCTION

Sensory methods used for QC should measure an attribute or group of attributes with limits identified in product specifications. Having a sensory specification as a guideline takes the sensory test out of the realm of OK/not OK and into a sensory QC program. Because the specification provides the basis for the sensory target, an understanding of specifications is presented next, before the sensory methods section.

SENSORY SPECIFICATIONS

A product specification typically defines the physical, chemical, and if appropriate, the sensory characteristics of a product. In contrast to physical (height, weight, bulk density) or chemical (moisture, pH, fat) measurements, which often are set within acceptable numeric ranges, sensory specifications have not typically been written with a quantitative orientation. Instead, they may contain a simple value orientation (good flavor, acceptable color), or if they were written in an attempt to be objective, they might have included vague descriptions such as "typical color" or "close to the standard." To be successfully applied to QC reporting formats such as those discussed in this manual, plant sensory testing should be based on quantitative specifications that define the critical sensory modalities and attributes. The best way to shift sensory testing away from liking or subjective judgments is to tell the panel what to look for and how to measure it. In writing a sensory specification, the panel should be viewed as an instrument, like a pH meter. Just as a pH meter is calibrated with known pH buffer solutions, the panel should be furnished with definitions and examples of what is target and what is not. After all, no one asks the pH meter if the pH is "typical" or "good." Why then should a sensory QC panel not deserve the same respect?

An ingredient specification should be made up of descriptive, physical, and chemical limits that are considered important to the product character and usage in the manufacturing process. If an ingredient is to be both a major part of a product and a source of flavor, it is important that the strength and quality of flavor characteristics of that ingredient be monitored as well as checking for "off" flavors. An example might be cherry juice to be used as a product ingredient, with a specification like this:

Product: Cherry Juice Concentrate

1. It shall be manufactured from sound, ripe cherries.
2. It shall be free of insect fragments, leaf and stem particles, and other contamination in accordance with good manufacturing practices.
3. The product shall meet all U.S. Food and Drug Administration (FDA) limits for pesticide and fumigant residues.
4. The flavor test should show intensities of 6–10 for cherry jam notes and 1–2 for burnt sugar notes on the 10-point weak-to-strong flavor scale.
5. A certificate of analysis shall be sent by the supplier to the QC manager at the plant receiving the shipment indicating a negative salmonella result and a brix of 38 minimum.

Tasting procedure:
Disperse 4-g concentrate, 4-g sugar, and 0.03-g citric acid in 100-g water. The aroma should be like cooked cherry. The flavor should be tart, moderately cooked, and jam like with no obvious burnt notes. Panel values should be between 6 and 10 on the 10-point cherry jam flavor scale, and not more than 2 on the browned/burnt sugar scale. If the value for cherry jam flavor is less than 6 or if the browned/burnt sugar flavor is more than 2, the shipment should be rejected or blended. An example of a graphic record of evaluations of cherry jam flavor levels in cherry juice concentrates by a trained panel is shown in Fig. 6.1.

If the flavor of a raw material is critical, or if the final manufactured product is produced by blending raw materials, some companies provide panel training and references to the supplier companies as well as their own QC labs. An example would be the blending of orange juices to produce a finished product with a target characteristic flavor profile.

An alternative to a descriptive specification is an agreed upon physical standard, which can be defined and reviewed on an appropriate basis by QA or by a panel. For example, a sample of strawberry preserves showing color, fruit piece size, and flavor at target levels

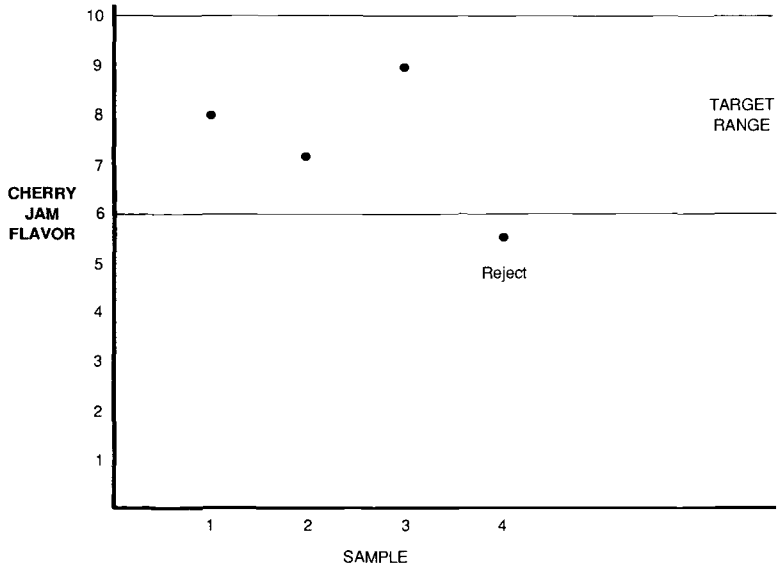


FIG. 6.1—*Panel evaluations of cherry juice concentrate for cherry jam flavor on a 10-point intensity scale.*

could be used instead of a descriptive specification for a product, with a selected acceptable range for color, piece size, and flavor. Target sample "drift" is a concern common to many companies. The "target" sample must be checked on a regular basis by a trained panel, the corporate sensory group, or other personnel familiar with the product to minimize drift away from the target in key characteristics.

Many QA groups have lists of product faults on which they base a quality "grade." If the product is lacking in faults, then it is assumed that it is satisfactory. Listing key differentiating characteristics is less common than listing product faults. Companies often depend on one for two "experts" to "know" a target product when they see it. One way to back up the "experts" is to have the quality team (management, marketing, QA, or other company personnel) taste the product, along with competitive products, asking "how does our company's product differ from the others?"

Differentiating in sensory characteristics does not necessarily relate to a quality or "better than" judgment. For example, the company might be manufacturing an economy grade product such as an inexpensive wine. To define their product sensory specification and then to maintain their target, the "target" sample can be tasted with other inexpensive competitive wines or even more costly wines of "finer quality." The important differentiating characteristics should then be listed, along with relative intensities. These characteristics, while not necessarily representative of the finest wines, are what the customers expect when they purchase the product. These are the characteristics that should be used in maintaining the target sample. Maintaining differentiating characteristics and keeping a record of the extent to which they are maintained can be helpful in answering marketing questions. If the product share is eroding, and the records show that the characteristics of the product are being maintained, then the answer might be marketing related. It is easier to define and tackle one problem than to try to answer both quality maintenance and competitive or pricing issues concurrently.

SUMMARY

The idea of a specification usually relates to the chemical and physical properties stated in a purchasing specification. A sensory specification defines the sensory parameters of an ingredient or finished product, such as flavor, aroma, texture, or absence of off flavors. The sensory specification defines the form of sensory data required, and the sensory data required will dictate the kind of test to be conducted. The more specific the wording of the specification with respect to sensory characteristics, the easier the selection of a test method and means of reporting results will be. The development of the specification requires understanding of product variation, effect of ingredient variation on the final product, characteristics of the product that drive acceptance, and the extent to which the differentiating characteristics can be consistently perceived by a trained or semitrained panel.

7

Methods

INTRODUCTION

Methods used for plant QC programs usually differ from those used in R&D or other kinds of consumer product testing. There are often two levels of sensory testing done in a plant quality program: (1) online checking and (2) informal to formal testing of ingredients, products in process, and finished products.

Online Checking

Online checking is being used with increased frequency, especially with the popularity of the TQM programs. Part of TQM is involving line personnel and training them to look for and report defects when they happen. The line people and the TQM team are often empowered to shut down a line if a problem is noted, thus saving the need for decisions being made as to what to do with a large quantity of out-of-specification product.

It is often the responsibility of the sensory staff to train and work with the line people to familiarize them with the characteristics of the products. To minimize distractions, the initial training should be conducted at a site away from the line, such as a conference room, with follow-up sessions online. Line personnel are sometimes hesitant to evaluate products because they have not been oriented to analytically tasting, smelling, or observing the sensory characteristics of products being manufactured. It is often helpful simply to purchase some familiar products like cookies or candy to orient them to the task of tasting or expressing their perceptions. Presenting two different brands of cookies or crackers of the same flavor can make them aware of differentiating characteristics of competitive products. Following training with products from the grocery store, the products made at the plant can be introduced. Some sensory groups make products with differences in ingredients, such as omitting a flavor, or mixing in a double weight of something, like salt or sugar, to demonstrate product faults. Several tasting sessions may be required to make untrained people aware of differences in sensory characteristics. A checklist of desired characteristics and product faults should be provided to the line personnel as a reminder after their training.

Training can be relatively simple if the product is a snack or a baked product. Training and subsequent evaluation can be more difficult with products requiring cooking, not because there are fewer perceptible differences, but because changes occur during and after cooking. Other products, such as salad dressings, require equilibration time to develop characteristic flavors and textures. These products may not lend themselves to online evaluations and may have to depend heavily on ingredient testing and process control to maintain quality.

Making plant workers aware of differences and able to see differentiating characteristics

of products can reinforce the idea that maintaining the sensory identity of a product can build brand loyalty, which can in turn enhance job security.

Sensory QC Testing of Ingredients, Intermediates, and Finished Products

The test methods used in a sensory QC program must fit the available resources. Test methods commonly used in sensory testing at a corporate research center might not be feasible at a plant level. There might be space limitations, too few people to serve samples and process the data, or too few people available to taste products.

In principle, most sensory testing conducted for plant QC programs is conducted to answer the question: Is the ingredient similar in sensory characteristics to what has been approved, and is the manufactured product similar in characteristics to what the company intends to manufacture? Referring to the previous chapter, how well does the product fit the sensory specifications set for it? These questions are most easily answered by comparing the sample in question to a standard or target product. This standard, or target, can be a physical retained sample of previously acceptable product and evaluated with the sample array, or it can be a mental profile of characteristics based on the product sensory specification.

Whether the sensory test is a difference test (such as, triangle or duo-trio), a measurement of degree of difference from a reference, a quality or grading scale, or a form of hedonic measurement, the test still measures how close current production is to some form of target product.

There are many variations of tests appropriate for sensory testing at a plant level. What testing works and does not work varies greatly from product to product and from company to company. Because there are few industry standards and such a diversity of products, it is unrealistic to prescribe "standard methods" for plant sensory evaluation. The approach taken for this guide is to provide examples of sensory testing used in hypothetical plant situations with varying testing capabilities and allow the reader to adapt the tests to his or her situation.

Sensory tests for plant quality control can range from simple pass/fail judgments to large consumer hedonic panels. In general, the tests requiring the fewest resources also have the lowest calculable confidence level. However, in actual practice, more small-scale than large-scale testing is conducted.

One of the most critical needs for sensory testing in a manufacturing plant is with incoming ingredients. If the ingredient has an off flavor or does not meet sensory specifications, the resulting finished product is likely to be adversely affected. Yet, few plants have the luxury of taking the time to conduct on the spot full-scale sensory panel testing of incoming ingredients. Many plants do not have the flexibility to reject incoming ingredients without shutting down the line. For these reasons, many companies rely on vendor certification of ingredient quality or require preshipment samples for evaluation. Methods for evaluating incoming ingredients and intermediates should be selected with consideration given to the effect they have on finished product quality and the resources available to test them. For example, corn syrup used in certain processes can have a range of flavor variations without adversely affecting the final product. It would not be appropriate to reject incoming shipments for flavor variations if they will not be detected in the final product. In other instances, where corn syrup is a major ingredient and variations may affect the flavor of a final product, such as a fruit-flavored beverage, it would be important to conduct a sensory test on the incoming corn syrup. Knowledge of the process and the product is very important in determining what ingredients must be tested, especially in small plants with limited resources. A summary of test methods based on the principles behind them is presented in Table 7.1.

TABLE 7.1—*Summary of methods for sensory QC testing.*

Method	What It Is	How It Works	Benefits	Concerns
Pass/fail/accept/ reject	a judgment(s) by one or more people to accept a sample	it is a comparison to a standard or what the product should be; the standard can be a real, physical reference, or a "mental picture"	minimal resource quick turn around widely recognized	no measurable confidence depends on individual/individuals being very dependable, consistent and objective
Simple difference	judgments picking an odd sample, as in a triangle or duo-trio test	samples are presented to judges blind coded, to see if products are perceived as similar or different	tests are well known tests can be supported statistically	more samples are required because of number of observations required for statistical support; inexperienced nondiscriminating panelists may fail to find a difference when a difference does exist
Degree of difference, difference from reference	a comparison of test sample to a reference indicating the magnitude of differences on a line scale	a reference is presented, along with a set of blind coded samples; an acceptable range of difference from the reference must be established	an integrated evaluation quantitative can be supported statistically can be presented in a QC format	difficult to use with untrained panelists; people weight attributes differently unless relative importance is agreed upon

Attribute scaling	1. concensus scaling—a group of panelists agrees on the magnitude of each attribute intensity 2. individuals scale the attribute intensity followed by statistical analyses	samples are accepted if responses are within range of limits, or within limits defined by reference samples	descriptive quantitative can monitor panelists	no estimate of variance performance expensive in resources and time data difficult to manage
Qualitative flavor	description of the sample by 1. a generally trained (flavor profile) or specifically trained panel 2. an individual group with no formal preparation or standard vocabulary	sample accepted on presence of typical attributes and rejected on the presences of a typical or absence of typical	descriptive, more information than simpler tests no data analysis cannot be presented in a QC data format	not quantitative; description may be dependent on group dynamics; less rapid turn around than pass/fail; highly or dependent on individuals
Quantitative flavor	description of the sample by a trained panel with replicated testings	samples evaluated in an array and compared to a known target	descriptive quantitative, supported statistically	requires training; maintenance of a panel for the product, must be cost justified
Quality scale	rating of quality on a numeric or word anchored scale	scale has a cutoff point for acceptance	integrated response fits QC data format	subjective importance of attributes must be agreed upon
Hedonics	consumers evaluate for overall acceptability or for end use functionality	product is routinely sent to a large panel or to a sample of the target population	most easily understood by marketing norm can change with competitive mix	image may change although acceptance has not measures more than conformance to specification

APPLICATIONS OF SENSORY METHODS

The following examples are presented to illustrate the methods listed in Table 7.1. The examples are in the form of case studies adapted from experiences of the contributing members. They include applications to realistic products, variations of the basic methods outlined in Table 7.1, and ways to use and report the resulting data.

Example 1: Pass/Fail Accept/Reject

Company A has several small plants located across the United States. Each plant has a QC lab. Ongoing finished product quality is monitored by retail audits conducted by an outside company. Company A management depends on vendors to maintain ingredient quality, and because their retail product is a dry pudding mix, they have committed minimal resources to a plant QC program.

A QC staff of one person conducts the required chemical tests of the mix before packaging. A final product check of each batch, after packaging, is made by the QC person by preparing the product according to package directions. This evaluation is a pass/fail, accept/reject decision.

The data in this case are a recorded pass/fail for the day, shift, or batch. This QC sensory check assures that the correct packaging is used (that is, chocolate mix is not packaged in a vanilla carton, and there are no obvious flaws such as failure to perform because of missing ingredients, presence of off flavors as a result of foreign material, or mishandling). The advantage of this procedure is minimal cost and rapid turnaround. Most likely, it is appropriate to the complexity of the process and product. Disadvantages are the possibility of an individual being "blind" to some off flavors that might be detected by some consumers and the weight of the pass/fail decision, which must be made by the QC person, with no gray area in between.

An alternative to the pass/fail test would be to make the evaluation in more of a controlled fashion to produce data suitable for presentation in chart or graphic form. This alternative would involve preparing the pudding according to a standardized procedure, then allowing it to stand for a specified time in a specified container. It then would be evaluated by a small panel, comparing it to a chosen "target" or "standard" for appearance, color, smoothness, cling to spoon, or other selected characteristics. An alternative to a chosen standard might be a training session for a few people to familiarize themselves with the key characteristics of the product, and have the small group evaluate against their "mental standard."

The data for these evaluations could be plotted daily to spot gradual changes in characteristics and could also be used to support ship/not ship decisions. The data collected in this manner also takes the weight of a pass/fail decision out of the hands of the QC technician. The disadvantage is the added time required to standardize the preparation and evaluation. The cost/benefit must be considered by the individual company in deciding how far to support these tests.

An example of plotting data from a test like this is shown in Fig. 7.1. The data are from evaluations of smoothness on a 10-point scale, with 1 being not smooth, 5 being the target for smoothness, and 10 being extremely smooth. The graph, using hypothetical data from two panelists' mean values, shows that Batches 3, 4, and 5 were problem batches, and the problem was corrected with Batch 6. With limited sampling and testing like this, the limits around the target are usually set by some group such as, QA or marketing, within the company.

One important point to make for all end-product testing is to be sure to test the product as it is likely to be used. For example, to save time, an end product might be evaluated

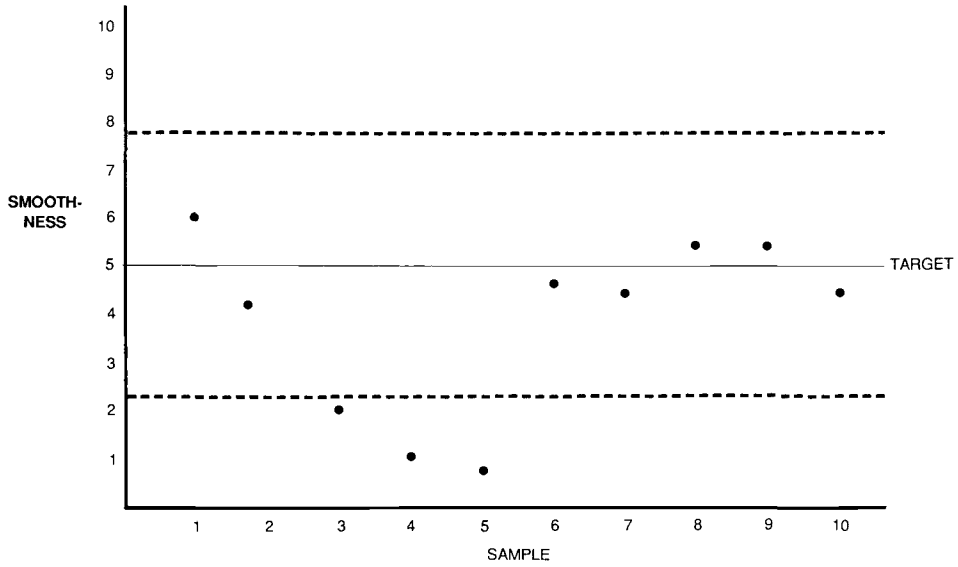


FIG. 7.1—*Smoothness testing over ten batches.*

cold by the QC staff instead of heating it as the customer would be likely to do. Perception of volatile materials will change depending on the temperature, and the time saved by not preparing samples uniformly and correctly is not worth the possibility of obtaining inaccurate results.

Example 2: Difference Testing

Company B manufactures candy. Their caramel candy is made with a special cooking process that produces a flavor that sets it apart from the competition. To maintain their distinctive flavor, the company depends on the line supervisors to evaluate the product on an ongoing basis to control the cooking process and maintain quality. If the line supervisor notes a product problem, the candy is put on hold and the QC lab arranges to conduct a difference test to decide if the product should be shipped, reworked, or discarded.

A sample of a recently produced batch within chemical and sensory specification is used for a triangle or duo-trio test, along with the product in question. This test can be done in two stages. The first step would be to have a group of 5 to 6 people who are familiar with the product taste the samples to decide whether more judges (30 to 40) are needed to provide the statistical support for the decision to be made. The difference test judges are recruited from the office and plant, and the samples are presented blind coded in randomized order. If there are no statistically significant differences between the acceptable batch and the one in question, the product is recorded as OK to ship. If differences are found, the product is reworked in successive batches or scrapped. The data from the difference test would be recorded as no significant difference (NSD) or as different with a probability level presented with it.

In summary, the difference test accomplishes the following for Company B:

1. It provides statistical support for the accept/reject decision and helps in training the line people as to how much difference in product characteristics can be perceived by unbiased judges.

2. It avoids total dependence on people who may know the product well and who are able to perceive minute differences that might not be noticed by people less familiar with the product.
3. It provides rapid turnaround, since the correct responses can be counted as the test is conducted.

Difference tests can be useful if the objective for management is to maintain the product the same as a day or two before. They will not provide information on product drift over time and will not tell how different the product was from the reference, or in which attributes the difference was seen. Difference tests also do not provide a "center line" around which variance can be plotted for QC reports. With difference testing, careful consideration must be given to the age and condition of the standard product selected. Difference tests require a fairly large judge pool, along with space and people to present the samples. On the positive side, difference tests require only task training of the judges, and the results are supported by statistics.

One improvement that Company B might consider is to provide the line supervisor with a "reminder sheet" with key characteristics of the product at the point of evaluation on the line. A reminder sheet gives the supervisor a list of what to look for as well as a means of keeping a record of having looked at the product.

Example 3: Difference from Control Testing

The degree of difference test was first described by Gacula et al. [*10*]. An adaptation of the test, difference from control, is described by Meilgaard et al. [*11*]. The objective of the degree of difference test as it was first described was to deal with batch to batch variations of formulated processed products. The difference from control test tells the user how different one or more samples are from a control or reference sample. Examples of products where this type of test might apply are variations of natural products, such as fruits, or formulated products, like sauces or cakes.

The examples presented here represent some of the ways in which the difference from control test can be applied. The difference from control (or degree of difference) test recognizes that most foods, being natural products or formulated and processed, have some variance, and with some products the band of variance may be so great that the high and low ends would not pass a triangle or duo-trio test. It is with products with some expected and acceptable variation that the degree of difference test is most useful. Also, as indicated in the discussion of the difference test in Example 2, the results of ordinary difference testing cannot be reported in graphic form. The results of difference from control/degree of difference testing can be graphically presented.

These tests are all based on a reference standard. The degree of difference test differs from difference from control tests in that variance as a result of judge (as measured by reference versus blind-coded reference) is averaged with variance as a result of batch or lot of product. Variation as a result of batch is measured by testing the reference versus a blind-coded second sample from a different batch or lot. The variance as a result of product and judge is then used in the F test of reference versus test product. The results are analyzed by a variation of the analysis of variance, testing whether the test sample score is different from the average of the two reference samples.

The difference from control test uses a blind-coded replicated sample to measure only panel variance.

Degree of Difference

Company C produces a potted meat product. Variations in raw materials and processing can cause changes in the finished product. A product in question was presented to a panel of 24 judges as follows:

1. The reference.
2. The reference, blind coded labeled as 894.
3. Product from another batch, blind coded as 673.
4. Test product, blind coded as 258.

The panel was asked to rate the samples as to the degree of difference from the reference on a 0–10 scale where 0 was no difference, 1 was a very small difference, and 10 was extremely different. The test data are presented in Table 7.2.

The data from this test can be analyzed by analysis of variance or some other way of comparing values if a computer is not available. The ANOVA table is shown in Table 7.3.

The sums of squares in the ANOVA table were obtained by standard procedures as follows:

Total sum of squares

$$(2)^2, (0)^2, \ldots, (6)^2 = 1166 - (266)^2 = 456.61/72$$

TABLE 7.2—*Data from the degree of difference test.*

	Sample		
Judge	Reference Versus 894	Reference Versus 673	Reference Versus 258
1	2	1	6
2	0	3	7
3	1	2	5
4	1	3	7
5	0	3	6
6	2	2	6
7	3	1	6
8	2	3	6
9	2	2	6
10	3	4	6
11	1	2	7
12	0	1	7
13	3	1	4
14	0	2	8
15	0	0	6
16	0	1	8
17	1	1	7
18	3	4	6
19	1	1	9
20	0	3	6
21	0	1	7
22	1	2	6
23	2	1	4
24	1	1	6
Sum	29	45	152

TABLE 7.3—*ANOVA table for degree of difference test data.*

Source	*DF*	*SS*	*MS*	*F*
Total	71	456.61	. . .	. . .
Test versus both references	1	367.36	367.36	326.5
Pure error	24	27	1.13	. . .
Residual	46	62.25	. . .	. . .

Correction factor

$$(299)^2/72 = 709.4$$

The test versus reference is the sum of the average of the two "controls" and the test, squared, minus the correction factor:

$$(29 + 45)^2/2 = (74)^2/2 = 5476/2 = 2738$$

$$2738 + (152)^2/24 = 25\,840/24 = 1076.76 - 709.4 = 367.36$$

The pure error is determined by the differences between the two "control" samples, squared and summed

$$(1)^2 = (3)^2 = (1)^2 \ldots (0)^2 = 54/2 = 27 \text{ with 24 degrees of freedom}$$

The residual is then 62.25, with 46 degrees of freedom.

Acceptance testing can be conducted to determine the F level for acceptance of rejection of a particular product.

Without a readily available computer program, the degree of difference test can still be used. For example, if the average for the two blind-coded references is 1.5, while the test product mean is 6.3, one would assume that the test product is indeed different. Experience or acceptance testing must determine the extent to which a product can differ from the mean of blind-coded controls and still be acceptable. The importance of the degree of difference test is the recognition of normal variation of product that is still within chemical and physical specifications, so that acceptance or rejection of a batch is done on a realistic basis. The degree of difference test is similar to looking at products on X charts, in which test products are plotted among other products above and below the mean line, and the relationship of a questionable product to normal variation can be seen visually.

Difference from Reference

The difference from reference test uses a reference sample, a set of test samples, and a blind-coded reference sample which is used as the estimate of panel variance. Occasionally, a reject sample might be placed in the array to check the panel and to relate results to previous tests. The reference sample in most cases should be the "target" sample, a product representing the middle of the quality or acceptance range.

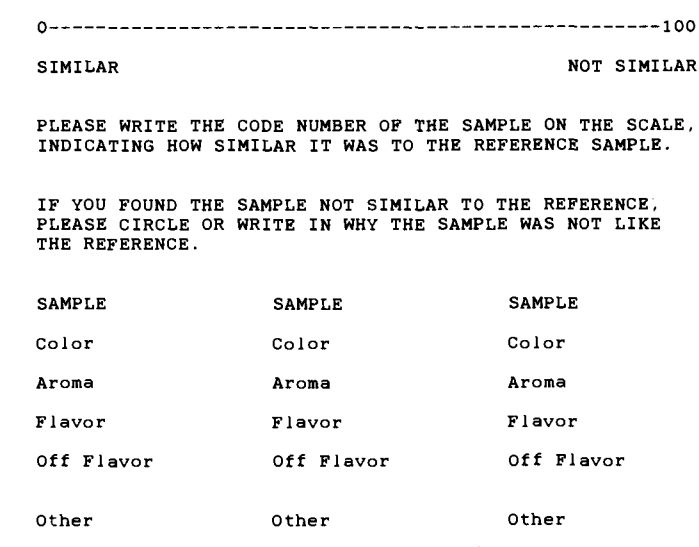
Company D manufactures several products in its main plant. The plant sensory QC person has handled the challenge of maintaining needed sensory quality of ingredients, intermediates, and finished products by using variations of the difference from reference test. For example, Company D uses honey as an ingredient in a product. The specification for honey includes a range of color and flavor of the honey to be purchased. Attributes of color,

aroma, flavor, and "off" flavor all contribute to the sensory quality of the honey. The QC group at the plant has three options with respect to the incoming ingredient: accept, reject, or blend.

A group of six to ten people who are familiar with the sensory characteristics of the honey used at the plant are presented with a reference which represents the "target," and an array of four to five blind-coded samples, including a blind-coded reference. If possible, a blind-coded "reject" sample is also included. Their primary task is to score on a line scale how different the products are from the reference. The scale can be open or divided, and can be assigned any number range. This example uses a 0–100 scale. While all samples are scaled based on how similar they are to the labeled reference, the degree of difference for each sample is reported based on how the incoming samples relate to the blind-coded reference. Note that in this case, only the blind-coded reference is used for comparison of the test samples. The reason for not including an extra blind-coded acceptable sample from a previous batch, as was done with the degree of difference test, is to enable the panel leader to test a larger quantity of incoming samples. If too many duplicates are presented, the number of incoming samples must be limited to prevent panel overload. Testing an array in this fashion can be used as a "rough cut." If there is a question about a given sample, it can be retasted with other replicated samples to provide more confidence in the accept/reject decision.

Commonly on a 100-point scale, a duplicate reference sample will be scored around 15, even with experienced judges. If the incoming sample is within the range of difference seen between the reference and the blind-coded reference, the sample is accepted for use. If it is outside the difference range, the next step is to look at a list of attributes that are key to whether the product can be blended and used. These attributes are listed on the questionnaire, following the degree of difference scale. The attribute that caused the sample to be "different" is circled or written in by each judge. The characteristic contributing to the difference will aid in deciding whether to reject or blend and how to blend.

A sample questionnaire, showing the mean panel results for two test samples and the blind reference, is shown in Fig. 7.2. The results from eight judges tasting the honey are shown in Table 7.4.



0----------100

SIMILAR NOT SIMILAR

PLEASE WRITE THE CODE NUMBER OF THE SAMPLE ON THE SCALE, INDICATING HOW SIMILAR IT WAS TO THE REFERENCE SAMPLE.

IF YOU FOUND THE SAMPLE NOT SIMILAR TO THE REFERENCE, PLEASE CIRCLE OR WRITE IN WHY THE SAMPLE WAS NOT LIKE THE REFERENCE.

SAMPLE	SAMPLE	SAMPLE
Color	Color	Color
Aroma	Aroma	Aroma
Flavor	Flavor	Flavor
Off Flavor	Off Flavor	Off Flavor
Other	Other	Other

FIG. 7.2—*Sample questionnaire for difference from reference test for ingredient testing.*

TABLE 7.4—*Results of honey difference from reference test.*

Judge (Blind Reference)	Sample		
	589	623	754
1	15	22	18
2	10	30	24
3	18	24	27
4	40	14	16
5	16	33	29
6	21	36	31
7	17	30	19
8	13	29	23
Mean	15.5	30.5	23.3

The degree of difference from the reference has been set by QA as 15–25 = accept, 26–40 = blend, over 41 = reject. A record of samples can be maintained on a simple graph, such as the one shown in Fig. 7.3.

If the incoming sample is within the set range of difference between the reference and the blind-coded reference, the sample is accepted for use. If it is outside the difference range, the next step is to look at the list of attributes circled as reasons for the product having been seen by the judges as different. The characteristics contributing to the difference will aid in deciding whether to reject or blend and how much to blend.

Maintaining a graph of evaluated samples can be valuable in reviewing supplier reliability. For example, in Fig. 7.3 the results show Suppliers *Y* and *X* to be reliable, but Supplier *X* seems to be having some difficulties, with one sample rejected and the two succeeding

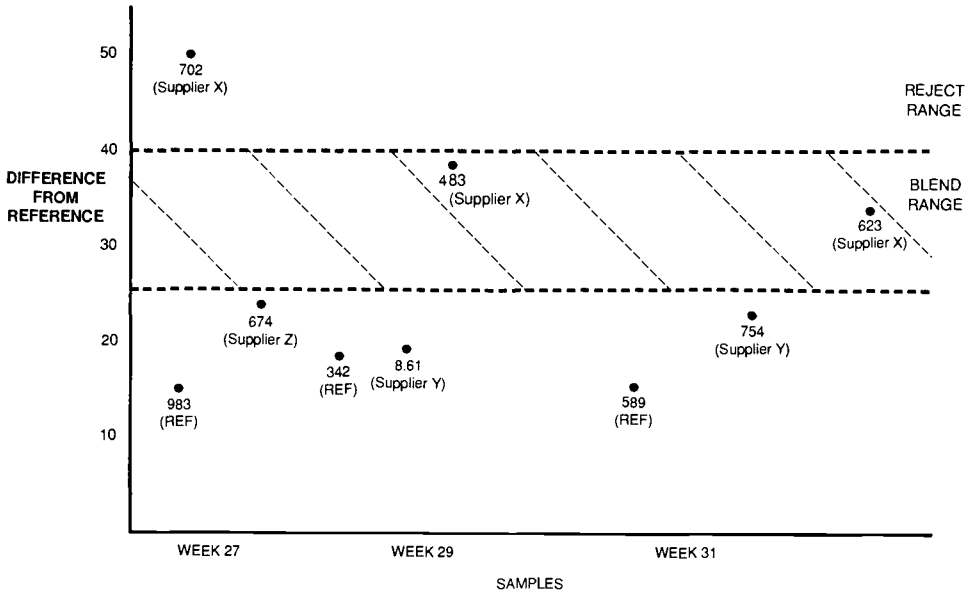


FIG. 7.3—*Ongoing record of sensory testing of incoming honey.*

samples in the blend range. Showing suppliers graphic data such as this can be useful in establishing and maintaining quality ingredients from them.

Variations of the difference from control test can be conducted with a larger number of judges or with replicated tastings to provide statistical support if desired.

An alternate approach to the difference from reference test questionnaire would be to rate the reference versus the same test array for intensity of selected characteristics instead of overall similarity. The advantage of using attributes is in having quantitative measurements for discussing quality problems with suppliers. The disadvantage is in having more data to handle.

A questionnaire for difference from reference testing of selected attributes is shown in Fig. 7.4. Hypothetical data from a test of weak to strong honey flavor are presented in Table 7.5, and a plot similar to that from the difference from reference test is shown in Fig. 7.5. For attribute testing, mean values of attribute intensities for reference or "target" samples are sometimes established to anchor the values for the test samples. These mean values can be marked on the questionnaire, and the test samples are then judged as weaker or stronger than the reference.

Figure 7.5 shows the strength of honey flavor of the samples in the test array relative to the mean honey flavor intensity given to reference samples in past testing. The graph shows that the blind reference and Sample 754 were perceived as slightly weaker in honey flavor than the reference, while Sample 623 was stronger. Whether the strength of flavor alone caused it to be seen as different enough from the reference to be in the blend range shown in Fig. 7.3 can be judged by looking at the other attributes tested. This type of attribute

```
Please rate the samples for attribute intensities relative
to the reference.  Values for the reference are marked on
the scale.

Light Color---------X-------------------------Dark Color


Weak Aroma ------------X----------------------Strong Aroma


Weak Flavor----------X-------------------------Strong Flavor


No Off                                          Strong Off
Flavor      -X----------------------------------Flavor
```

FIG. 7.4—*Questionnaire for difference from reference attribute testing.*

TABLE 7.5—*Data for weak to strong honey flavor, relative to reference.*

	Sample		
Judge	589	623	754
1	15	40	19
2	18	33	17
3	12	28	10
4	15	24	10
5	14	26	12
6	20	31	11
7	12	35	10
8	15	27	8
Mean	15.1	30.5	12.1

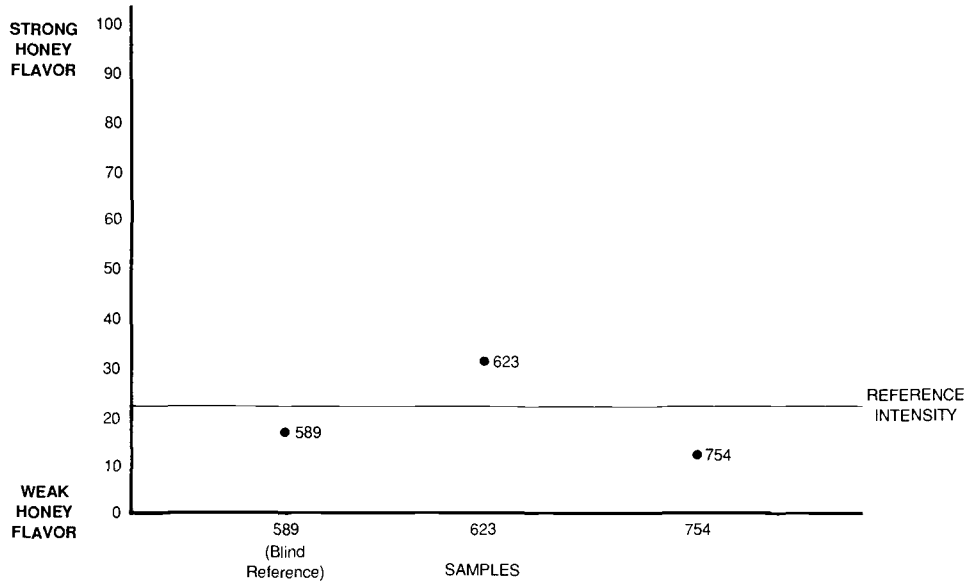


FIG. 7.5—*Attribute intensity difference from reference.*

testing might be used in combination with the overall difference from reference test when more information is desired, such as when and how much to blend. If, for example, the difference in Sample 623 was more related to color then a decision might be made to accept it for use rather than blending, or to use more of it in a blend.

In summary, the use of the difference from reference test has provided Company D some flexibility in purchasing a natural variable product. It can provide a method of tracking quality and variability of different suppliers. The use of attributes provides opportunity for troubleshooting. The use of the scale and attributes also facilitates blending of honey shipments. The test requires some advance work in identifying unacceptable ranges and setting tolerances by QA and other groups, such as R&D or marketing, and in establishing mean attribute intensities of the target product if attributes are measured. The overall similarity and scales of attributes can be mixed to suit the information desired and the ability to handle quantities of data, making this test very versatile for a plant QC application.

Another application of the difference from reference test is in control of mixing and processing where on the spot decisions must be made.

Company D manufactures a product in two stages. It manufactures an intermediate product, which is then made into a finished product by another department in the plant. It has a system for evaluating incoming ingredients as previously discussed, but controlling only the ingredients does not predict the effects of mixing and processing or possible ingredient interactions. Company D added a sensory analysis to the physical and chemical measures used for controlling the intermediate product.

A sample of an intermediate mix is prepared according to the product specification. Observations on the mix appearance are recorded by the technicians preparing the samples. The intermediate product is then evaluated by a group of three to six panelists from the QC staff. They are presented with a reference, a blind-coded reference, the sample, and one or two other samples that have been previously accepted. Experience must dictate how much difference from the blind-coded reference can be tolerated before the intermediate

product is put on hold. The use of the difference from reference in this case enables the company to know how much variation there is in the intermediate product, and with the small panel a decision can be made quickly as to whether to use the intermediate product or not. The difference from reference test conducted in this fashion provides no statistical support of the decision, but the larger number of judges does provide more reliability than a single judgment by a line supervisor. Also, a record of the mean values can be maintained, showing possible drifts over time. Showing the mean values on a product quality graph can also provide motivation for the panel.

Company D also makes salad dressings. For dressings, they use another variation of the difference from reference test. This test involves the use of a "mental scale" by a group of trained panelists. This mental reference is especially important with products that have a limited shelf life or a flavor that changes with time so that a physical reference is a "moving target." For this type of test, the critical characteristics are scaled and compared to values of the attributes for the "mental standard." Panel training and experience in working together are very important in this application. The questionnaire is similar to that of the attribute difference from reference test. The value for the mental standard is pre-marked on the questionnaire, and the sample numbers are written on each attribute line by the judges relative to their mental reference as shown in Fig. 7.6. It can be seen from the marks on the scales that Sample 2, the blind-coded reference, is similar to the mental reference. However, Sample 1 is stronger in acetic acid and rancid oil, and weaker in black pepper than the mental reference.

Use of mental references requires training and experience. An ongoing question is how one keeps either mental or physical references from changing or drifting over the years. There are no easy answers to the question.

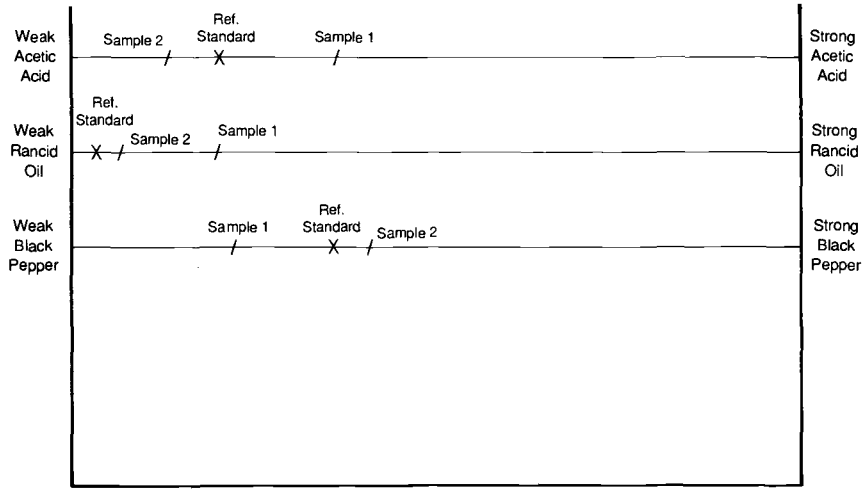


FIG. 7.6—*Using a difference from reference test with a "mental" reference.*

ATTRIBUTE SCALING

Flavor ingredients are often major determinants of the character of the finished product. While the base product can be controlled by physical and chemical analyses, many flavor

compounds are difficult to define chemically. Chromatography can identify major components, but minor components that differentiate delicate parts of a flavor are usually measured by human senses.

Company E makes strawberry flavored syrup and decided to reduce the risk of variation in the product by working with the vendor to improve the shipment to shipment consistency of the product. The flavor ingredient used in the product manufactured by Company E changes with time, thus precluding the possibility of using a physical reference in a QC test.

The method selected by Company E was to have preshipment samples of each lot of flavor sent for QC approval, based on the following procedure:

1. The flavors are diluted for tasting according to the written specification for the flavor. Procedures written into the specification provide a means of communication between the buyer and vendor.
2. Ten panelists familiar with the flavor are used. Their task is to provide quantitative responses to attribute questions using terms and scales defined by the specification.
3. At least two other samples are included in the array presented to the judges: an approved sample from a recent shipment and a rejected sample if available. The samples are presented in random order. The tasting is replicated in four sessions, thus providing 40 data points for a one-way analysis of variance. The panel data are handled as individual points, not as means.

Decisions on acceptance or rejection of the preshipment sample are made based on a statistically significant difference in one or more flavor attributes. An example of the questionnaire for an attribute scaling test like this is given in Fig. 7.7.

In summary, Company E has achieved the following with this QC approach:

1. Establishment of working lines of communication between the vendor and the company.
2. Active involvement between purchasing and QC through writing and understanding of the ingredient specification.
3. Establishment of a semitrained panel but without dependence on one or two "top tongues."
4. Statistically supported decisions.

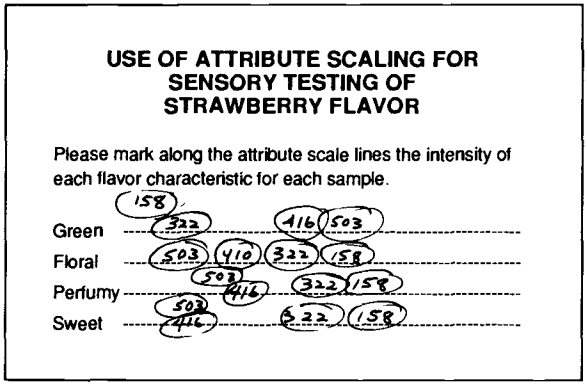


FIG. 7.7—*Use of attribute scaling for sensory testing of strawberry flavor.*

Training the vendor's panel in the use of the same test would assure the use of a common language and understanding between manufacturer and supplier.

DESCRIPTIVE ANALYSIS: QUALITATIVE AND QUANTITATIVE

Descriptive analysis can take many forms in sensory testing for quality control. It can range from a flavor profile of the desired finished product or ingredient developed by a trained group of judges to statistically supported replicated tasting by trained groups. Rather than providing detailed procedures and applications in this manual, the reader is referred to Refs *12* and *13*, which contain a thorough presentation of descriptive techniques and applications.

Descriptive testing can be useful to vendors and manufacturers in checking the quality of ingredients against a specification. The "spider web" graphs used to report quantitative descriptive results on product differences can be effective tools for QC work.

QUALITY SCALES

Quality scales, along with pass/fail decisions, are among the most widely used and recognized of all QC sensory procedures. Quality scaling assigns a number or letter grading to a product based on the level of one or more attributes. The grades are also sometimes represented by words such as "typical/good," "acceptable/fairly good," "marginal/fair," "poor/inferior," or "unacceptable/reject." Quality scales are often used by quality control groups who have people assigned to be specialists for a product. The specialists are taught or develop a level of expertise on the product that enables them to give an integrated grade for the sample. Commodity items, such as fruit juices and vegetables, are often graded with these scales.

The important characteristics of any product can be weighted on a score sheet like the one used by Company F, a bottled beverage manufacturer, presented in Fig. 7.8. The use of a scoresheet like this implies a knowledge of what the typical product should be so that it can be defined with weighted attributes as well as the training of personnel in understanding the point system.

In summary, Company F takes advantage of its experts on staff to maintain beverage quality and uses the quality scale to monitor product quality and drifts. Quality scales are dependent on well-written and understood definitions for each level of the scale and on judges who are familiar with the product.

Quality scales can be presented in graph form over time, similar to the data presented for the difference from reference test. Caution must be used with quality scales. The weights for each characteristic must be determined carefully and care must be taken that one very bad flaw is not overlooked because the sum of the remainder compensated for it. Because the quality scale number is a sum of a number of attributes, it is difficult to track individual attribute drifts, unless each attribute score is graphed or tracked mentally by the person grading products.

HEDONIC TESTING

Hedonic testing has been mentioned previously, relative to establishing limits as to how far a product can be from the target product before acceptance is affected. Hedonic testing conducted at the plant level is generally not recommended because of the inherent bias

Appearance	Points
Typical color, cloud, Bright color, no sediment	+3
Clean label	+1
Dull color,sediment, cloud ring, other	-1 per fault
Total appearance points	4 maximum
Total minimum appearance points acceptable	2
Aroma	
Typical aroma	+2
Clean,characterizing aroma, good level	+2
Off odor or other	-1 to -2
Total aroma points	4 maximum
Total minimum points acceptable	1
Taste/Flavor	
Typical sweetness and flavor, good sweet/tart balance	+7
Good characterizing flavor	+1 to +5
Off flavor, low sweetness too tart	-1 to -7
Total flavor points	12
Total minimum points acceptable	5

FIG. 7.8—*Example of a quality scale for a bottled beverage.*

toward the company product on the part of the employees. This bias relates both to the desire to want to like what the company produces out of company loyalty and to familiarity with the product, which creates a comfortable relationship with it.

Some companies use hedonic testing for quality control, sending regular production samples to an established panel of consumers.

Household products' Company G produces a cleaning agent. The product specifications state that the product will have a cleaning efficacy comparable to or better than the other leading brand. Data from chemical and physical analyses are used to control daily productions. Sensory testing is done on 200 boxes sampled from 1 week's production. These samples, along with 200 boxes of competitive product, are purchased by a local pickup service. A large pool of panelists is maintained such that each week 200 placements of 2 products can be made, resulting in each panelist family receiving a placement every 6 weeks. The panelists

represent the demographic/geographic target consumer population. They have been instructed on the use of an efficacy scale, and they have participated in at least 2 "trial" placements before their data are used. Their performance is monitored over time. Erratic inconsistent panelists are discovered not only by their performance on the trials but also by "ringer" tests, which include duplicate samples or samples with known defects that have been shown in other testing to be significantly different from target product. The data are collected by telephone two weeks after placement. The data are fitted to control charts and appropriate changes are made to raw material evaluations or processing if the product is outside of control limits.

This type of testing is more extensive and expensive than most sensory testing for plant quality control. It does directly measure consumer response to the product in a realistic fashion. It is subject to competitive pressures on the product, which might not be directly related to maintaining defined product attribute and quality targets.

Company G uses the consumer hedonic data as a means of quality control and ongoing product development or improvement, based on how the product fares against competition.

8

Sampling

HOW MANY SAMPLES?

Often the number of samples evaluated is dictated more by limitations of testing than by an ideal number from a QC textbook. Some questions to be asked by the person setting up the program might be as follows:

1. How many samples can be prepared and served in the time available?
2. How variable are the components, intermediates, or finished products to be tested?
3. How important is the measured variability to perceived product quality, that is, does the observed variability affect acceptability?

Allowing for limitations in resources, it is best to look at several samples encompassing a wide range until some understanding of the uniformity is established.

WHAT SAMPLES?

The quality of incoming raw materials (ingredients and packaging) assures the quality of the finished products, assuming the process is in control. Often water is overlooked in sampling of raw materials. Water used as an ingredient is a key raw material and should be included in all sampling plans. All raw materials that have the potential of contributing off flavors to the finished product should be tested. They should be tested as an incoming ingredient, and they should also be tested during warehouse storage. If there is a possibility of line contamination or heat damage between two points in the plant, the raw materials should be tested at the point of use. Selecting samples requires knowledge of the characteristics. The benefits of sensory analysis of raw materials are as follows:

1. Prevention of failure caused by unspecified materials—In-process or finished products have incurred production costs. It is desirable to prevent the failure before manufacturing costs are added.
2. Providing data for supplier quality management—Vendors should be informed of prescribed sensory specifications, and an inbound report of each raw material should be supplied by the vendor guaranteeing the quality of their product. Testing for pertinent specifications assures vendor reliability and assures no contamination or degradation of material resulted during shipment.

Sensory testing of all raw materials may require more resources than the product requires or cost can justify. Every lot of every ingredient may not need testing. Suggested alternatives are as follows:

1. Establishing a priority according to how critical the ingredient is (for example, flavors need more sensory testing than malto-dextrin).
2. Skip-lot sampling with provisions that new suppliers or new ingredients may require more intense sampling initially.
3. Suppliers generally known to be reliable may be asked to submit preshipment samples from a large production lot. This lot can then be delivered in several smaller and timely shipments. This allows careful, more complete analysis of the preshipment sample and reduces the pressure and need for hasty decisions on receipt of shipments.
4. Issue sensory specifications for raw materials to vendors, placing responsibility on them for meeting specifications.

CONSIDERATIONS IN SENSORY TESTING OF PROCESS CONTROL SAMPLES

Thorough knowledge of the process and how process variation relates to sensory product changes can help in developing a test program of critical control points. For example, an error in weighing cinnamon is more critical to sensory characteristics than an error in weighing flour. Time at a higher temperature (pasteurization, for example) might be more critical than time at a cooling process step.

CONSIDERATIONS RELATIVE TO PLANT SENSORY TESTING OF FINISHED PRODUCTS

Most production facilities assume "hands on" evaluation of products as they are produced. Management is often included in these evaluations, either formally or informally. As previously stated, finished product evaluation is the least effective place for plant sensory to work because evaluation is done "after the fact." It is more effective to evaluate ingredients and in-process products because steps can be taken to correct defects. It is costly to reject finished, packaged product. Finished products have all ingredient, process, and package costs in them, and rejection requires thorough documentation of the sensory information.

Another factor to consider in doing finished product sensory analysis is that fresh product may have to age for a day or two or more before the characteristics seen by the consumer develop. For example, peanut butter crystal structure requires 24 h to stabilize. Salad dressings require some time to blend. Staleness and off flavors are often more difficult to perceive in very fresh products compared to those aged for a few days. Because of the "aging" needed for some products, it is also difficult to maintain reference samples against which to test some ongoing production products.

Because of the pressure not to reject something that is already made and packaged, it is easy to see that placing the product on "hold" or rejecting it is more difficult if it is already in the warehouse ready for shipping.

SENSORY RETAIN SAMPLES

Provision should be made for appropriate storage of retained samples of critical ingredients or finished products. This provision includes storage in sealed containers that may or may not be airtight and storage at proper temperature and humidity to minimize degradation. These samples should be retained for a sufficient period of time to answer any trouble-shooting questions before sale. These samples are not the same as reference standards which are used for sensory analysis of incoming ingredients or of finished products. Reference standards should also be maintained in proper packaging under temperature and humidity conditions that will prolong their useful life as a reference standard.

Summary

This manual was written to bring sensory testing for quality control in a manufacturing plant out of the "touchy/feely" and value judgment realm into the world of quality control data reporting. It was also intended to help a person or group who is beginning a sensory test program at a manufacturing facility. There are many things to consider in starting such a program, and many of the points raised in the book were the result of successes and failures shared by ASTM members working in the field.

The terms "attributes" and "sensory characteristics" were stressed throughout, because these are appropriate to QC sensory testing. Like/dislike testing is used in developing and determining the kind of product that the customer will be likely to purchase. Once the product is developed, the objective of quality sensory testing is to characterize the target product and maintain the manufactured product similar to the target.

Measuring how close the product is to the target is the basis for obtaining quantitative sensory data. The overall aim of this book has been to orient the planning of sensory testing so that objective, quantitative results can be provided to those responsible for overall quality control.

References

[1] Eggert, J. and Zook, K., Eds., *Physical Requirements Guidelines for Sensory Evaluation Laboratories, STP 913*, American Society for Testing and Materials, Philadelphia, 1986.
[2] ASTM Committee E-11 on Quality and Statistics, *Manual on Presentation of Data and Control Chart Analysis, MNL 7*, American Society for Testing and Materials, Philadelphia, 1990.
[3] Kramer, A. and Twigg, B., *Fundamentals of Quality Control for the Food Industry*, Avi Publishing Company, Westport, CT, 1966.
[4] Mendenhall, W. and Reinmuth, J. E., *Statistics for Management and Economics*, 2nd ed., Duxbury Press, North Scituate, MA, 1971.
[5] Jellinek, G., *Sensory Evaluation of Foods*, VCH-Ellis Harweed, Publisher, Chichester, England, 1985.
[6] Poste, L., et al., *Laboratory Methods for Sensory Analysis of Food*, Agriculture Canada, Publication 1864/E, Ottawa, Canada, ISBN 0-660-13807-7.
[7] Moskowitz, H., *Applied Sensory Analysis of Foods*, Vol. 1, CRC Press, Boca Raton, FL, 1988.
[8] Amerine, M., Pangborn, R., and Roessler, E., *Principles of Sensory Evaluation of Food*, Academic Press, New York, 1965.
[9] Gacula, M. and Singh, J., *Statistical Methods in Food and Consumer Research*, Academic Press, London, 1984.
[10] Aust, L. B., Gacula, M., Beard, S. A., and Washam, R. W., II, "Degree of Difference Test," *Journal of Food Science*, Vol. 50, 1985, p. 511.
[11] Meilgaard, M., Civille, G., and Carr, B. T., *Sensory Evaluation Techniques*, Vols. 1&2, CRC Press, Boca Raton, FL, 1987.
[12] Stone, H. and Sidel, J., *Sensory Evaluation Practices*, Academic Press, New York, 1985.

Index

T–X

ISBN 0-8031-1486-9