

Chapter 15

Consumer Field Tests and Questionnaire Design

Abstract This chapter presents an introduction to consumer testing in various settings including central location and home use tests. The construction of a useful consumer questionnaire requires both skill and experience. General rules for questionnaire design and question construction are presented. Various question formats such as agree–disagree scales and open-ended questions are discussed.

Developing products is easy, developing products that appeal to consumers is less so, and developing products that appeal to a sufficient number of consumers and achieve commercial success based on specific business criteria is very difficult.

—Stone and Sidel (2007)

Contents

15.1	Sensory Testing Versus Concept Testing	349	15.6.5	Avoid Vague Questions	365
15.2	Testing Scenarios: Central Location, Home Use	351	15.6.6	Check for Overlap and Completeness	365
15.2.1	Purpose of the Tests	351	15.6.7	Do Not Lead the Respondent	365
15.2.2	Consumer Models	352	15.6.8	Avoid Ambiguity and Double Questions	366
15.2.3	Central Location Tests	353	15.6.9	Be Careful in Wording: Present Both Alternatives	366
15.2.4	Home Use Tests (HUT)	354	15.6.10	Beware of Halos and Horns	366
15.3	Practical Matters in Conducting Consumer Field Tests	355	15.6.11	Pre-test	366
15.3.1	Tasks and Test Design	355	15.7	Other Useful Questions: Satisfaction, Agreement, and Open-Ended Questions	367
15.3.2	Sample Size and Stratification	355	15.7.1	Satisfaction	367
15.3.3	Test Designs	356	15.7.2	Likert (Agree–Disagree) Scales	367
15.4	Interacting with Field Services	358	15.7.3	Open-Ended Questions	367
15.4.1	Choosing Agencies, Communication, and Test Specifications	358	15.8	Conclusions	368
15.4.2	Incidence, Cost, and Recruitment	359	Appendix 1: Sample Test Specification Sheet	370	
15.4.3	Some Tips: Do’s and Don’ts	360	Appendix 2: Sample Screening Questionnaire	371	
15.4.4	Steps in Testing with Research Suppliers	360	Appendix 3: Sample Product Questionnaire	374	
15.5	Questionnaire Design	362	References	378	
15.5.1	Types of Interviews	362			
15.5.2	Questionnaire Flow: Order of Questions	362			
15.5.3	Interviewing	363			
15.6	Rules of Thumb for Constructing Questions	364			
15.6.1	General Principles	364			
15.6.2	Brevity	364			
15.6.3	Use Plain Language	364			
15.6.4	Accessibility of the Information	365			

15.1 Sensory Testing Versus Concept Testing

Why do so many products fail in the competitive marketplace? To minimize new product failures, one strategy is to insure that the consumer perceives, through the senses and repeated experiences, the characteristics

Table 15.1 Sensory tests versus product concept tests

Test characteristic	Sensory test	Product concept test
Conducted by	Sensory Evaluation Department	Marketing research Department
Primary end user of information	Research and Development	Marketing
Product labeling	Blind-minimal concept	Full-conceptual presentation
Participant selection	Users of product category	Positive response to concept

that make the company's product superior to competitors' products and thus more desirable. Furthermore, this perception must be maintained to build brand loyalty and to insure repurchase. The sustainability of the perception of quality is important long after the initial rush of interest from advertising claims and promotions that surround a new product introduction. The purpose of this chapter is to discuss the techniques for consumer product testing in the field on a blind-labeled basis that insure the sustainable perception of positive product characteristics.

A consumer field test with new product prototypes or market candidates can provide several pieces of useful information to product developers. The blind-labeled sensory test can be an important step before a multi-city marketing research field test and product "launch." It provides an opportunity to determine consumer acceptability on a sensory basis, without the concepts and claims that will normally appear in advertising or on packaging. The sensory consumer test can facilitate diagnosis of problems before more expensive marketing tests. Costly mistakes can be avoided and problems uncovered that may not have been caught in laboratory tests or more tightly controlled central location tests. It can provide direction for re-formulation if needed. Multiple formulas or candidates can be compared on a blind-labeled, pure performance basis. Poorly performing products can be dropped from further consideration. Finally, since the tests are done with target consumers, the company can obtain data that may be used for claim substantiation. This can be valuable in defending challenges from competitors and responding to the requirements of advertising regulators.

At first glance, a consumer sensory evaluation field test looks a lot like the consumer tests done in marketing research. It is worth understanding some of the important differences. The research arm of a consumer product manufacturer may rely on the important technical support provided by the sensory consumer test. It provides validating data about reaching goals in terms of sensory factors and the perception of product

performance by consumers. Important differences exist between sensory consumer tests and the typical marketing research consumer tests. Some of these are shown in Table 15.1. In both tests products will be placed with consumers and their opinions surveyed after a trial period. However, they differ in the amount of information given the consumer about the product and its conceptual features.

The marketing research "product concept" test usually proceeds as follows: participants are shown the product concept, via a storyboard or videotape mockup that often resembles a rudimentary advertisement for the idea. They are then questioned about their response and expectations of the product based on the presentation. Note that this is important strategic information for marketers. Next, *those that respond positively* are asked to take the product home and use it, and later to evaluate its sensory properties, appeal and performance relative to expectations. This selection may appear biased, but it is based on the idea that people who do not like your idea to begin with are probably not part of the target market. In the sensory-oriented consumer tests the conceptual information is kept to a minimum. The rule of thumb is to give only enough information to ensure proper use of the product and evaluation relative to the appropriate product category. For example, the product may be labeled simply "pizza" or "frozen pizza." In a market research test, it might be being evaluated after a conceptual presentation that communicates a number of features such as "new—improved—low-fat—high fiber—whole-wheat—stuffed-crust—convenient—microwavable—pizza." Ratings of attributes and product acceptance are contaminated by any additional information or by the expectations that are built up as a function of showing the detailed product concept. The sensory product test attempts to ascertain their perceptions about the sensory properties in isolation from other influences. These other influences can be quite profound. For example, the introduction of brand identity or other information can produce differences in the apparent acceptability of products where there

is no differentiation on a blind-tested basis (Allison and Uhl, 1964; El Gharby, 1995; Gacula et al., 1986).

Sensory consumer tests and marketing product concept tests also differ in the participant selection. Only those who show an interest and react positively to the product concept are usually included in the actual product test in the marketing research scenario. As these participants have shown an initial positive bias, it is perhaps not surprising when the product receives good scores in the test. On the other hand, the sensory consumer test merely screens participants for being users of the product category. Given these differences, the two types of tests may give different evaluations of the consumer appeal of the product. The tests provide different types of information, viewed from different frames of reference by the consumers and they use different pools of respondents. In the concept test, product perception may be biased in the direction of assimilation toward their expectations (Cardello and Sawyer, 1992). Results from both types of tests may be equally “correct;” they are simply different techniques seeking different types of information. Both types of information should be weighed in management decisions to go forward or to seek further modifications to optimize the product.

Critics of the sensory approach often remark that the product will never be seen on the store shelf with a generic description and a three-digit code, so why bother evaluating it in the blind-labeled, concept-free form? The answer is simple. Suppose the product fails in the marketplace? How does one know what went wrong if only the product concept test was performed? Perhaps it did not have good sensory properties or perhaps the marketplace did not respond to the concept as predicted by the marketing tests. Without the sensory test, the situation is ambiguous and the direction for fixing the product is unclear. The research team may have designed a poor product that was only carried along by a catchy concept. However, after extended use, consumers may have figured out that the product does not deliver benefits in keeping with their expectations and they stopped buying it. On the other hand, the marketing team may have simply designed a poor concept that somehow moved forward in the rush of initial enthusiasm for a new product idea (Oliver, 1986). The research and development team has a need to know if their efforts at meeting sensory and performance targets were successful in an unambiguous blind test.

The following sections of this chapter are devoted to how consumer tests are conducted, emphasizing field testing and questionnaire construction. Although there is a substantial literature on survey techniques and questionnaire design for marketing research and for opinion polling, there is little published research on sensory field testing of consumer products. Product placement and interviewing in the field is a complicated, expensive, and time-consuming enterprise. Training is often obtained by “shadowing” an experienced researcher in industry. However, there are some general guides to consumer sensory tests (Schaefer, 1979; Sorensen, 1984). The book, *Consumer Sensory Tests for Product Development*, by Resurreccion (1998) contains guidelines and much practical advice. It provides detailed information on topics such as checklists for conducting various types of consumer tests, sample questionnaires, project management guidelines, and maintaining consumer databases of test participants for recruitment. Another resource dealing specifically with the issues and methods for claim substantiation is the extensive ASTM standard E-1958 (ASTM, 2008).

15.2 Testing Scenarios: Central Location, Home Use

15.2.1 Purpose of the Tests

The primary goal of a consumer field test is to assess the acceptability of a product or group of products or to determine whether a product is preferred over other products. Some typical situations that justify a consumer field test are (1) a new product entering the marketplace, (2) a reformulated product, that is, ingredient, process or packaging changes of a major nature, (3) entering a competitor’s product category for the first time, or (4) competitive surveillance, as in a category appraisal (see Chapter 19). It is also an opportunity to collect some diagnostic information on the reasons behind consumer likes and dislikes. Reasons for liking are usually probed with a variety of techniques such as open-ended questions, agree–disagree scales, and just-right scales. Agreement with label claims of a perceptual nature (e.g., “crispier”), assessing consumer expectations and consumer satisfaction with the

product can all be surveyed through questionnaires and interviews.

Four general categories of consumer sensory tests are often distinguished. The first is the use of internal panels checking acceptance using on-site testing, usually with employees. The second is the use of a local standing consumer panel. These are people, sometimes from social groups that are recruited for multiple tests for a period of time. We will refer to these two types of panels as “consumer models.” The third is the central location test or CLT and the fourth is the home use test or HUT.

15.2.2 Consumer Models

A variety of consumer testing situations are used to assess the appeal and overall acceptability of products. Due to resource constraints in time, money, or security concerns, there are several types of acceptance testing that use what can be termed “consumer models.” Such “consumer” groups may consist of employees or local residents, but often there is little or no attempt to insure that the group is representative of consumers at large. It is of course essential that the group be users of the product category. It would make no sense to ask people about the appeal of several extruded puffed breakfast cereals if they never consume such a product.

Internal consumer tests are tests conducted in the sensory test facility of a company or research department using employees. A major liability of employee panels is that they are not necessarily blind to the brand of the product and they may have potentially biasing information and assumptions about what is being tested (Resurreccion, 1998). Technical personnel may view the product quite differently from consumers, focusing on entirely different attributes. Such an internal consumer panel should be routinely compared to an outside sample of non-employee consumers by testing the same products with both groups. Stone and Sidel (2004) describe the use of a split-plot ANOVA with internal and external panelists evaluating the same products to assess the degree of agreement. Unfortunately this is rarely done during product development, because the diagnostic information from the early tests with the internal panel is used to make adjustments or optimizations before subsequent expensive field tests. So the product changes as it moves through the development process.

Another cost-efficient approach using consumer models is to use local standing consumer panels. The oldest record of a local standing panel that we could find was the use of 300 families in the Columbus, Ohio area by the Ohio State University and Ohio Agricultural Experiment Station in the 1950s (Gould et al., 1957). The families were stratified by socioeconomic class via census records of rental costs and they participated in simple preference tests. This panel was similar to the Kroger company’s mail panel of the 1940s, described by Garnatz (1952), which had a more geographically diverse makeup. Test products were delivered to their homes and questionnaires returned by mail. Another way to recruit and set up standing panels is through community groups. These groups may be affiliated with schools, churches, fraternal, or hobby-oriented clubs (e.g., singing groups) or virtually any other organization that meets in a nearby location on a regular basis. Social groups can be used for central location tests, sometimes in their own facilities (Schaefer, 1979) or to facilitate distributions for home placement. They can be contacted through a panel leader or coordinator for product and questionnaire distribution, offering some time savings. Such panels are re-used for a period of time, so like the employee consumer models, they offer convenience and time savings in recruiting respondents and testing products on a routine basis. Incentives can be directed to the organization itself, so there can be social pressure, a powerful motivator, to participate.

However, there are several liabilities with a locally recruited and ongoing consumer panel. First, the sample is not necessarily representative of opinions beyond the delimited geographic area of the club or group. Second, the participants may know each other and talk to each other on a regular basis, so there is no guarantee that the opinions are completely independent. Using a variety of random codes for products may reduce this liability, but there is no airtight guarantee that their judgments are not influenced by others. Finally, unless an outside agency or a disguised testing lab is used for the contact and distribution, the participants may become aware of what company is conducting the test. Opinions or pre-existing attitudes about products from the originating company may bias results. If they view the company favorably, they may evaluate its products more favorably. As in any consumer acceptance test, the participants should be carefully screened for regular usage of the product category. That is, you must

eliminate those members of the group who are not regular users even if it engenders some disappointment among people who were looking forward to being in the test. This possibility must be spelled out during orientation sessions after the group is recruited.

In spite of the apparent savings in recruitment and completion time for tests, the setup and maintenance of a local ongoing consumer panels can require considerable effort on the part of the sensory staff. If the program is big enough and has contact with ten or more consumer groups, a full-time staff member may be required to supervise such a program. Pickup and recall of products and questionnaires must be arranged, and returned questionnaires should be carefully examined for any evidence of cheating. People may fill out questionnaires and return them without actually trying the product. Containers returned full or nearly full, questionnaires with illogical answers or use of only one scale point on every question are some hints (Resurreccion, 1998). These respondents should be deleted from the data set and noted for future exclusion or monitoring. Maintaining good relations and close communication with a member of the group acting as the local contact person is key. These group coordinators or contact persons have many of the responsibilities for oversight that a field agency supervisor has, but bear in mind that this is not their profession. Cultivating this important contact person, orienting them to procedures and providing motivation may also take time and considerable social influence on the part of the sensory specialist. Such a program of local panels should also involve periodic furlough of each group and rotation of active participation cycles since some boredom and disinterest in the tests can set in over time. A 6-month active testing cycle within every 18 months to 2 years is reasonable.

In spite of their obvious problems in being non-representative of outside consumers, employee panels and local consumer panels can provide valuable information on a cost-efficient basis. For decades, the US Army has used employee panels at the Natick, Massachusetts laboratories to evaluate military rations and other foods. These panels have been reasonably well predictive of soldiers' opinions of the same foods (Meiselman and Schutz, 2003; Peryam and Haynes, 1957). However, more recent work showed that the correlations were higher for snack foods than main dishes and meal components and were more predictive when the laboratory test involved an element of choice

(de Graaf et al., 2005). In food companies, the risks from using an internal panel can be a bit higher. For a major roll-out of a new product with millions of dollars of advertising to be spent, it is much safer if not imperative to continue testing with a true consumer field test, in a home use scenario in multiple locations.

15.2.3 Central Location Tests

Probably the most popular type of consumer test with foods involves product trials at a central location. The central location test (CLT) is often conducted in the facilities of a field testing agency (a service provider), for example, in a shopping mall. However, there are just about as many variations on this theme as one could imagine including testing in retail outlets, recreational facilities, and schools (Resurreccion, 1998; Sorensen, 1984). A picture of a CLT setup classroom style is shown in Fig. 15.1. Consumers can come to a corporate sensory test lab, although that defeats the non-branded nature of the blind sensory test since the company identity is obvious. Having consumers on the company property may entail some security risks. If the testing program is extensive, it may be economically justifiable to set up a disguised test facility actually administered by the company's own sensory program, rather than subcontracting to outside testing services. A sensory group can use a mobile testing laboratory to change locations. The University of Georgia



Fig. 15.1 A consumer central location test being conducted in a classroom style arrangement. Photo courtesy of Peryam and Kroll Research.

maintained such a mobile testing laboratory moved by a truck (Resurreccion, 1998). This offers enormous flexibility in the opportunities for consumer contact. For example, foods targeted at summer picnics or outdoor cooking could be tested at or near campgrounds or parks. Such on-site testing can introduce a realistic element in the testers' frame of reference. Products aimed at children could be brought to school locations and the mobile test laboratory can provide a site for proper product preparation and controlled presentation. For some foods, special considerations for preparation are minimal and tests can be conducted at any site where people are gathered in large numbers and have flexible time. For example, sensory testing can be done at state fairs or other recreational events.

The central location test offers conditions of reasonably good control of product preparation as the staff can be trained in product preparation and handling. Compliance with instructions, manner of examining samples, and ways of responding may be monitored and controlled (Resurreccion, 1998; Schaefer, 1979). It is easier to isolate respondents in test booths or separate areas to minimize outside influence. Security can be maintained more easily than in a home placement. The tradeoff occurs against the necessarily limited product usage, i.e., participants' exposure to the product is much shorter than in the home placement and usually only limited amounts of the product are tasted or consumed (Schaefer, 1979). It is of course possible that the limited product interaction in a CLT can give erroneous results or lead to faulty conclusions (Oliver, 1986).

15.2.4 Home Use Tests (HUT)

The most expensive but most realistic situation is when consumers take the product home and try it under normal circumstances on several occasions. Home use tests are time consuming to set up and administer. They can be costly, especially if external field testing services are hired to do most of the work. However, HUTs offer tremendous advantages in terms of face validity of the data. This can be important in advertising claim support. Also, the opinions of other family members can enter the picture as they do in everyday use of purchased products. The primary advantage is that the consumer uses the product over a period of

time and can examine its performance on several occasions before forming an overall opinion. For foods, this becomes less of an issue where flavor, appearance, and texture are rapidly appreciated and the hedonic reaction of a person is virtually immediate upon tasting the item. For consumer products such as a shampoo or a floor wax, it may be critically important to have some extended use in the actual home and to see how the treated substrate (in this case the hair or the floor) holds up over time. Home placement provides a chance to look at the product in a variety of settings (Anon., 1974). Another important opportunity is the chance to test product and packaging interactions. Some products may be well or poorly suited to their package design (Gacula et al., 1986) and the home use test provides an excellent chance to probe this. Finally, the home use test can facilitate a more critical assessment of the product relative to the consumer's expectations.

In the case of product fragrance testing, the short exposure in a central location may overestimate the appeal of very sweet or perfume-like fragrances. When used in the home for extended periods, such fragrances may become cloying and a type of hedonic fatigue can set in, even though they score well in briefer laboratory sniffing tests. In general, it may be dangerous to screen fragrance candidates in lab-based sniffing tests, especially for functional products. A very appealing fragrance sniffed from a bottle may not communicate a message of efficacy with industrial cleaning products or insecticides. For functional products, the fragrance has to be chosen to support the perception of product efficacy. Mismatches can occur in flavors as well, for example, there is some resistance to candy-like flavors in toothpastes (Jellinek, 1975). Similarly, very sweet food products may score well in a central location, but do less well when used over an extended period.

In summary, four major categories of consumer tests are commonly used—employee consumer models, local standing consumer models, central location tests, and home placements. The remainder of the chapter will focus on field tests and on questionnaire design. Table 15.2 shows the characteristics of these levels of consumer testing and how they vary. The employee tests are the quickest, least expensive, and the most secure but have the greatest liabilities in terms of potential bias, lack of a representative sample, and lack of realism in the testing situation. The choice of test in any specific situation usually represents a compromise between time and expense on the one hand and

Table 15.2 Kinds of consumer tests

Type	Advantages	Disadvantages
Internal employee panel	Secure Low cost Rapid results	Not representative
Local standing panels	Reasonably secure Lower (?) cost Easy distribution	Not a random sample Panelists may discuss products
Central location test	Representative sample Control over product preparation	Requires test agencies Cost, slower results
Home use test	Representative sample Realistic testing Whole family input Can test use directions	Requires test agencies Slowest to conduct, costliest Lack of product control Security risk

the need for the most valid information on the other. The business risk of basing major decisions on less valid testing situations should be weighed against the cost of more extensive testing.

the added levels of complexity compared to simple in-house acceptance/preference testing. Field tests are costly and require a high level of attention to detail. Simple mistakes can render the results of a test invalid (Schutz, 1971), at potentially great expense if the test must be repeated.

15.3 Practical Matters in Conducting Consumer Field Tests

15.3.1 Tasks and Test Design

A number of considerations will enter into the design of a home use test, many of which need to be negotiated with the client or end-users of the information and also with any field testing services that collect data. Some of the primary decisions of the sensory professional will include sample size, experimental design, qualification of participants, choice of locations and agencies, and structure of the interview or questionnaire. There are several dozen activities and decision points in setting up and conducting a field test, rendering this type of test as one of the most complicated projects that a sensory specialist may perform. The most important decisions that affect the experimental design include the number of consumer required, the number of products, and how the products will be compared. Statistical consultants, if needed, should be brought in at this stage. The specific tasks involved in conducting a consumer field test are discussed below. Resurreccion (1998) provides checklists for the various types of consumer tests, home use, central location, etc. These can be very useful for the sensory specialist who is new to consumer field tests and may not realize

15.3.2 Sample Size and Stratification

In this instance, sample size refers to the numbers of consumer participants, not the size of the portion or amount of product served. How many people should be in the test? More powerful tests are less likely to miss a real difference or an important effect and having a sufficient sample size is the first concern in test design. A statistical consultant can help with estimates of test power, but there are ultimately some subjective decisions to be made about the size of a difference one is willing to miss, or conversely, that one must be sure to detect. This decision is akin to determining how big of a difference is practically meaningful or what small differences can safely be ignored. Once the effect size is specified, the probability of detection (one minus beta-risk) also must be chosen. This is called the power of the test (see Appendix E). These may be difficult concepts for management to understand, unless they have had extensive statistical training and a good deal of practical experience. The level of variability also affects the test power but it can be used as a kind of yardstick. The size of the difference you wish to detect can be stated in terms of standard deviations. A reasonable rule of the thumb for the level of error in consumer tests is that standard deviations will

be in the range of 20–30% of the scale (or about two points on a 9-point scale, which has only eight intervals). The range of variability can be slightly lower for intensity scales than for hedonic scales. Variability will also be lower for “easy” attributes, like those having to do with appearance or some simple texture attributes, as opposed to taste characteristics or olfactory or aromatic attributes, which are the most difficult of all. Given this rule of thumb, it makes sense to have a sample size in the range of 75–150 persons (per product) for most tolerable levels of risk. A generally useful equation for evaluating the required sample size based on scaled data (like the 9-point hedonic scale) is shown here:

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2 S^2}{(\mu_1 - \mu_2)^2} \quad (15.1)$$

where N is the number of consumers needed in the test, Z_{α} and Z_{β} are the Z -scores associated with your chosen levels of alpha- and beta-risk, S is the anticipated standard deviation of the scores (or a pooled estimate), and $\mu_1 - \mu_2$ is the difference between means or the size of the difference you want to be sure to detect.

Can a test be too big? Although some marketing groups are prone to do tests with hundreds or even thousands of respondents, this derives from a false sense of security in numbers (Stone and Sidel, 2004). There is a law of diminishing returns with sample size and statistical power, just as there is in interviewing in general. The largest amount of information is obtained from the first few interviews and additional testers yield less and less new information (Sorensen, 1984). It is also possible to have a test that is too sensitive, i.e., to show statistically significant results in an area that is of little practical consequence to consumers (Hays, 1973; Schutz, 1971). Stone and Sidel (2004) discuss “the curse of N ” in testing in the sense that people put too much faith in large numbers. Statistical significance must be weighed against practical significance. It is unfortunate that the technical meaning of statistical significance refers to issues of confidence and likelihood, while the common everyday meaning is synonymous with “important” (Sorensen, 1984). Management must be reminded of the difference in these usages in order to keep from over-interpreting statistical significance, especially in large test populations. Finally, it is better to have a small test of high quality that is well designed, with careful

attention to detail and close monitoring of field agencies than it is to have a sloppy test that uses large numbers of consumers to compensate for the added variability.

Of course, the sampling strategy may not be from a single group. It is sometimes desirable to look at different geographic locations, different demographic strata (e.g., age, gender, income) or groups with different product usage habits (Schaefer, 1979). There are two reasons for stratification of the sample group. The first is to insure a certain amount of diversity in the group so that it mirrors the target population. Thus there may be quotas for men, women, different age brackets, etc. This kind of quota sampling is very important in central location tests where the participants may be recruited through a mall intercept (recruitment on site). The second reason for stratification is that examining differences among these groups may be part of the research plan. If the test groups are stratified in this way, it becomes necessary to increase the total pool to maintain the minimum subgroup size in the range of 50–100 respondents. Obviously, such variables should be chosen very carefully and with solid justification, for they can dramatically increase the size and cost of the test.

15.3.3 Test Designs

There are three primary designs used in consumer testing. Side-by-side tests are sometimes done in which both products are placed simultaneously. These are more often performed in central location tests than in home placements. Under controlled circumstances, the side-by-side test will have great sensitivity, since the same people view both products. Difference scores (as in the dependent or paired t -test) or complete block (repeated measures) analysis of variance can be used to analyze the data. Comparisons are both statistically and perceptually direct. However, putting more than one product out simultaneously in a home placement test can lead to confusion for the participants. There are many chances for errors in following directions for product use, order of evaluation, and questionnaire usage if it is self-administered. The side-by-side evaluation is better suited to situations where the product–person interaction can be controlled and monitored.

More common designs in field tests are the monadic and monadic sequential placements. In the monadic test, only one product is placed with an individual. This is usually a faster test scenario, and can be completed more quickly with fewer drop-outs. However, it requires larger numbers of participants, one group for each product. It may not be practical if the incidence of product use is fairly low or participants are difficult to find and recruit. The statistical comparison between products is necessarily a “between-group” comparison. The opportunity to use consumers as their own baseline of comparison cannot be taken advantage of when different groups are compared. So there is a potential loss of sensitivity in this design due to high inter-individual variability. Conversely, the monadic sequential design permits the use of individuals as their own baseline. Scale usage habits or other individual peculiarities are the same for both products and can be statistically factored out of the analysis. This generally leads to a more sensitive test (discussed in the section on the paired t-test in Appendix A).

In the monadic sequential design, the products are placed one at a time in sequence. A questionnaire is normally administered at the end of each product’s usage period, while the sensory characteristics and performance are fresh in the person’s memory. Of course, careful counterbalancing of orders across the subgroups is necessary. Bear in mind that the first product used in a monadic sequential test will have the same frame of reference (or lack thereof) as a simple monadic design (Sorensen, 1984). So analysis of the first product used by each person can be informative if there is concern about sequential bias or any order effects on the second or third product. The monadic test leads to higher rates of attrition (non-completion). It does permit a preference question after completion of the second product placement.

Some situations arise in which monadic sequential tests are inapplicable. When the substrate or evaluation process is irreparably or severely altered by the initial product usage, the second placement becomes unworkable. For example, with a pharmaceutical, a personal care product like a hair conditioner or a home insecticide, use of the product may create such changes in the substrate that it is not practical to get a clear picture of product performance for the second product in a sequence. Of course, multiple products can be tested after a “wash-out” or recovery period, as is sometimes done in pharmaceutical tests. This may not be practical

for market-driven new product tests where time is of the essence.

The number of products to include is also a consideration in test design. It is possible to test more than two products in sequence or to use incomplete sampling designs like balanced incomplete block designs (Cochran and Cox, 1957; Gacula and Singh, 1984; Gacula et al., 2009) to test a number of possible alternatives. Due to the expense and effort in conducting home placement or central location field tests, the number of alternative formulas should have been reduced to only a few highly promising candidates through earlier phases of testing. One design to avoid is the one-product monadic test, which is really not a test at all, but an exercise in confirmation of the intuition of the project supervisors. A one-product “test” puts far too much faith in the raw value of the scores received. As humans are very poor absolute measuring instruments (see Chapter 9) and prone to context effects, the absolute value of the scores is nearly meaningless, even in comparison to historical data with similar products due to a potential change of context. It is far safer and much more scientifically valid to include a baseline product for comparison. Examples of useful baselines for comparison are an alternative formula, the current product, or a repackaged sample of a competitor’s most successful formula.

A final consideration in the test design is whether to include a question of paired preference. In a monadic sequential test, there may be considerable challenge to the participants’ memory if a preference question is asked following the use of the last product in the sequence. Due to the possibility of sequence effects, it is also wise to look at preference ratios for each separate order of presentation and not just for the overall test population. Paired preference may still be confirmatory to comparison of (scaled) acceptability scores, so it can be used as an additional source of information in developing the “story” that is told by the test and the scorecard. It is also possible to get conflicting results from acceptance and preference questions from some individuals. This can occur if they change their basis for decision-making between the two questions. For example, the acceptability question may be answered with sensory properties like taste or texture in mind, while the preference question may consider preparation time or some convenience factor. It is also possible that a product may win in a preference comparison, but because there is a highly

dissatisfied minority, the acceptance scores for the winning product are lower (an example is given in Stone and Sidel, 2004). There is a widespread belief that the paired preference question is somehow more sensitive than acceptability ratings, but this notion lacks empirical support. Nonetheless, considerable pressure may arise from clients to include a preference question. Practical considerations such as the length of the product usage period, results from pre-testing and needs for information for claim substantiation (claims such as superior to, unsurpassed, equally preferred) may determine whether or not to include a paired preference question.

15.4 Interacting with Field Services

15.4.1 *Choosing Agencies, Communication, and Test Specifications*

Field services are variously called agencies, vendors, suppliers, and field services. Some workers in this area prefer the term “suppliers” because they are supplying research information. Choosing a good field service or a test agency is largely a matter of experience. In a company with an ongoing product testing program, it makes sense to keep a record of those agencies that deliver timely and cooperative service, demonstrate quality in their interviewing, and show attention to detail in handling of products and questionnaires. The quality of service is not necessarily proportional to cost—high bidders may not always provide the best service (Schaefer, 1979). Costs of field services will depend upon their level of involvement. In some cases there may be two levels of contracting, a primary agency that administers the test, and subcontracting field test agencies in different cities that actually conduct the test under the direction of the primary contractor. It is important to distinguish between full-service suppliers and basic field test sites. Full-service suppliers can provide valuable input on the screening and product questionnaires, the design, execution, analysis, and reporting of results. They act as an extension of your professional team. In other cases, the subcontractors can merely provide a testing service, i.e., product placement and interviewing, and act according to your specific directions.

In each agency, it is important to identify a single person, sometimes called the field service supervisor or project manager, who is ultimately responsible for the conduct and quality of the test. Reporting to this person there are often many part-time employees who may have different degrees of training in interviewing techniques. The nature of interviewing is that it attracts a lot of freelance or part-time workers. They should have excellent interpersonal skills, the ability to follow directions, and a sense of caring and integrity about the quality of the job (Schaefer, 1979). The field service supervisor should visit any subcontracting sites and participate in or view the testing process if at all possible. There has to be good communication, a written test specification sheet, and a briefing of the field sites (well before the actual test) to answer any questions. Good agencies will provide training for the interviewers and a briefing for each test to review respondent qualifications, instructions for sampling, and placement and questionnaire structure. Supervision of the interviewers is important in quality control. Practice interviews have been found to predict field behavior and can be used as a screening device (Blair, 1980). Problem areas include cheating on screening and qualification of improper respondents and faking of part or all of the interview (Boyd et al., 1981). Checkups or validation of a given percentage of the completed questionnaires (usually by phone) should be requested and monitored by the field supervisor.

In using a field service for central location tests, facilities are very important. They must allow for proper experimental control, product serving and preparation, and provide an environment free from distractions and conducive to sensory testing (Schaefer, 1979). If there is a central location test involved, it may be necessary to hire an agency with facilities for product preparation. If agencies derive most of their business from servicing marketing research tests and/or focus groups, they may not be set up for food preparation and serving. So confirming the facilities, preferably by an in-person visit, can be an important detail. Considerations include the ability to isolate respondents to minimize interaction and the resulting loss of independence in their judgments. In a home placement, the agency test facilities are less important. Having a focus group room for follow-up discussions is one consideration if follow-up groups are a part of the project plan. As most testing companies maintain

websites, some idea of the nature of the facilities can be assessed from the pictorial material on the website.

After the agency is hired, it is important to make sure all the test details and instructions are communicated in writing. Most of the general facts will have already been communicated in order for the agency to make a cost estimate. Further details are critical to a successful test. Test specification sheets should normally be sent to an agency, giving as many details as possible about the test design, qualification of respondents, quotas, deadlines and services to be delivered, including data tabulation and analysis, if any. Expectations about security, confidentiality, and the professional conduct of interviewers can also be spelled out. Arrangements for product retrieval and disposal can be specified, as well as shipping of completed questionnaires. A sample test specification sheet is shown in Appendix 1.

15.4.2 Incidence, Cost, and Recruitment

It is up to the sensory professional, in consultation with the clients and perhaps with input from the field supervisor, to determine the screening qualifications of the participants. Certainly participants should be users of the product category and usually people who also actually like the product—the two are often overlapping but not completely synonymous. In addition, one needs to determine what level of product usage is sufficient to qualify an individual. A screening questionnaire will normally include several usage frequency categories, in order to eliminate those consumers that only use the product so rarely that they are really not in the target market. A sample screening questionnaire is shown in Appendix 2. A major consideration in determining the cost of the test is the incidence of users of the product category (Sorensen, 1984). What percentage of consumers in the general population use this product or category of products? When hiring an agency or requesting a cost estimate, incidence figures will be required to estimate the time required and thus the cost of recruitment. Marketing data can be helpful in this regard.

Another consideration is whether to recruit participants by phone, by intercept or from existing subject pools using a database of product use information. Some recruiting may be done over the Internet. It

has become increasingly common to recruit from a national database and then ship the product to the consumer's home. E-mail or Internet recruiting may work well with younger consumers. Recruiting by telephone can be time consuming, but may provide the closest approximation to a random sample of the area. Unfortunately it may miss people with unlisted numbers or people with no land lines (only cell/mobile phones), who represent a demographically different population segment (Brunner and Brunner, 1971). Intercepting individuals at a site like a shopping mall has been popular when the field agency has a testing facility in a mall. However, the nature of the sample must be carefully scrutinized due to the biases inherent in sampling shoppers. If there are stratification quotas for age, gender, etc., these are very important. Some larger testing agencies may maintain databases on pools of local consumers who have been recruited for general service in multiple tests. They may have answered questions on product use and therefore can save a lot of time in locating regular users of the products. However, since habits and situations change (e.g., health, dietary restrictions, family members changing residence), it is necessary to confirm their current suitability through the normal screening questionnaire. Furthermore, it is important to guard against overuse of people from standing databases or retested subject pools. They can become jaded or take on characteristics of professional testers. The participants should be screened for testing frequency or having not participated within a given time frame, usually several months. Three months between tests is a common requirement although Resurreccion (1998) recommends 6 months between tests.

Some other requirements and choices will affect the activities of the field agencies. Consider the stability of the product relative to the holding time that may be encountered during recruitment, for example. This is especially pertinent if there is a low incidence of users. Recruiting time may be very long, but should not exceed the freshness limitations on the product. Distribution or shipping may be a factor as well. One of the authors supervised a test in which frozen pizzas were shipped to the United States from Europe, but due to unexpected delays in U.S. Customs clearance were ruined and useless by the time they reached the test site. The method of product distribution can also be an issue. The two primary choices are personal pickup or mailing. Delivery to the home is also possible but can

be costly. If the individuals pick up the product at the agency facility it may be necessary to build in an extra incentive payment to cover their time and travel costs into the overall incentive for participation. Mailing a product risks mishandling, misdirection or delay, and the possibility of unknown temperature history, but it is a low-cost alternative if product stability is good.

15.4.3 Some Tips: Do's and Don'ts

A successful and useful relationship between a sensory researcher and a test agency requires good communication and a good working relationship. Here are some suggestions for situations to avoid when dealing with field services: First, resist the temptation to change the study design at the last minute. Do not expect to change your design, questionnaire, number of products, or recruitment criteria the day before the test. The testing service has scheduled the facilities and set up the test based on your specifications. It may not be possible to make the changes and keep the same schedule. It may be a simple matter to add or change a question or two, but it may not. Second, when requesting shelf life testing do not expect testing agencies to have a time machine. If you need shelf life data you must be willing to wait for it. Do not wait until 3 weeks before the product launch if you need 6 month's data. Even accelerated testing and the Arrhenius equation will not save you. Third, do not take any assumptions that the field service will fill in details the way you expect them to. Spell out all the test details in the test request or specification sheet (in writing) and re-visit them during the agency briefing (verbally) to field questions and resolve ambiguities. It is also a good idea to visit some of the test sites during actual testing. Finally, if the results do not turn out the way you would like, do not blame the field service.

15.4.4 Steps in Testing with Research Suppliers

Table 15.3 lists steps in conducting a test using subcontracting field services. Most of the items are self-explanatory and some are part of the normal testing process for any sensory test (such as problem identification). A few comments are made here as

guides to the sensory professional. Bear in mind that the exact nature of testing differs from product to product, among companies as a function of policies including security concerns. Electronic means of data collection will continue to supplant paper questionnaires.

The setup of the questionnaire takes considerable effort and involves negotiation with the clients and circulation of drafts. The contact person in the agency may be involved to some extent as appropriate to their level of expertise and the degree to which tasks are delegated to the field service. They may also be able to assist in some pre-testing of the questionnaire. Interviewers should be given explicit instructions. These usually include the following: (1) read the questions exactly as worded, (2) do not comment on meaning, (3) do not suggest any acceptable answers, (4) answer every question, even if recorded as "do not know," and (5) do not deviate from the sequence or skip pattern. In spite of such instructions, monitoring shows that many interviewers do not follow these rules (Boyd et al., 1981).

At this time, a number of details need to be arranged concerning the physical products to be tested. If the test is large enough, it may not be practical to have the products made in the laboratory, and arrangements for pilot plant or manufacturing time may be necessary. Storage conditions can also be arranged to mimic the conditions encountered in the normal distribution system (Schaefer, 1979). If competitive products are to be evaluated, they may need to be disguised or repackaged for the test to insure the blind, unbranded nature of the test. It is important to obtain representative materials and to avoid samples that are abused or defective (Sorensen, 1984). Samples will have to be labeled with codes and the generic title of the product. The label usage directions should be carefully considered as part of the test design. Finally, shipping of products and questionnaires is arranged. For heat or cold-sensitive products, delivery, handling, packing, and unpacking are major considerations. Weekend delivery is sometimes difficult to coordinate. Delays in delivery may result in temperature abuse, as in our frozen pizza example.

Contact with the field agency is important before the testing begins. The questionnaires should be reviewed by their staff and any residual questions about the test, procedures, instructions, the qualifications of respondents, or the scorecard should be cleared up. A pre-test

Table 15.3 Steps in conducting a home use test using field services (“agencies”)

Stage 1: Before the test
Identify problem, goals, and negotiate test design with client
Write proposal, including budget
Confer with statistical consultant if necessary (sample size, etc.)
Obtain approvals
Obtain bids and hire agencies
Send test specification sheets concerning participants, products, timing, questionnaires, etc.
Prepare questionnaire
Confer with client, marketing, etc. to make sure all issues are included
Pre-test, revise if necessary
Obtain samples, place request with pilot plant or other supplier
Obtain competitive product if necessary
Design labels
Choose coding system
Confirm usage instructions with developers
Get labels printed and affix to sample product
Prepare shipping orders and send product to agencies
Print questionnaires (if paper questionnaires are used)
Ship questionnaires to agencies with instructions
Visit agencies before the test or hold telephone briefings to review test details and to field any questions
Stage 2: During and after the data collection
Visit agencies to observe testing and/or participate in “callbacks”
Arrange for keypunching and data analysis confer with statistical consultant if necessary
Develop coding sheets for open-ended questions
Receive questionnaires (if paper), unpack and check for completeness, cull mistakes, incompletes
Arrange for data entry if paper
Conduct follow-up discussion groups if desired
Perform statistical analysis
Write report
Schedule presentation
Prepare visual aids for presentation
Present results
Revise, print, and distribute reports
Process bills from agencies
Archive questionnaires and data
Dispose of unused or returned product

visit to handle these matters in person is called a briefing. It may be appropriate to physically visit the agency and observe the test in progress and to participate in some of the interviews or “callbacks” if the product has already been placed and used. Personal visits ensure a quality check on the conduct of the data collection and can provide valuable insight into the degree of attention to detail paid by the agency’s professional staff and interviewers.

Following the test, it is incumbent upon the sensory professional to guide and oversee the data entry and analysis, even if this has also been subcontracted. This does not mean that a physical presence is always necessary, but auditing of questionnaires, screening them

for potential fakes and eliminating botched interviews is part of the quality control of the test process. The person in charge of the test should also develop a coding scheme for answers to open-ended questions, to guide the data entry process. A second area of involvement comes in questionnaire validation. This is usually achieved by some telephone callbacks from the field supervisor to a proportion of the respondents, usually 10 or 15% to verify that their opinions were correctly recorded and the interview was not faked (Schaefer, 1979). The sensory project leader will set these quotas for validation and of course they enter into the cost analysis of the field agency. Common areas of difficulty occur from interviewers failing to follow the

sampling plan or screening criteria and allowing people to slip in who are actually unqualified. Validate the respondent's qualifications as well as participation and responses. Close supervision of the interviewers by the field supervisor may help curtail many of these problems. A large percentage of the errors may be concentrated in just a few interviewers (Case, 1971).

A final opportunity to interact with the field service comes after analysis of some or all of the data. Answers to some questions, particularly open-ended probed questions, may suggest additional issues that require follow-up. Most agencies have or can provide focus group interview facilities. It may be advantageous to recruit a number of participants for group interviews to probe additional issues. For example, you can call back consumers who were very positive or very negative (perhaps in different discussion groups) to further probe their reasons for liking or disliking the product. Review of the questionnaire results may suggest potential alternative formulations or line extension opportunities or features that need to be added to meet consumers' expectations.

15.5 Questionnaire Design

15.5.1 Types of Interviews

The exact form and nature of the research instrument will depend upon the test objectives, constraints of funding or time and other resources, and the suitability of the interview format. Interviews may take place in person, be self-administered by paper, conducted via a website, or by phone. Each method has advantages and disadvantages (Schaefer, 1979). Self-administration is obviously the least expensive, but does not lend itself to probing of open-ended questions, is open to respondent confusion and mistakes in following directions, and is not suitable for complex issues that may require an explanation. There is no insurance that the person will not read ahead or scan the entire questionnaire before answering any questions. They may not follow the order of questions as printed on the survey. Cooperation and completion rates are poorer with self-administration (Schaefer, 1979). In addition, illiteracy is a problem so a self-administered questionnaire may simply be unusable. Many people will try to hide their inability to read.

Telephone interviews are a reasonable compromise, but may not lend themselves to complex multi-point scales—questions are necessarily short and direct. Respondents may also feel an urge to limit the time they spend on the phone and may produce shorter answers to open-ended questions (Groves, 1978). Telephone interviews are somewhat prone to early termination by the respondent. The in-person interview is the most flexible and the questionnaire can be complex and include a variety of scales since the interviewer and questionnaire are both present for clarification (Boyd et al., 1981). Visual aids can be brought along to illustrate scales and scale alternatives if the interviewer is reading the questionnaire to the respondent. Advantages of this method may be offset by the higher cost (Boyd et al., 1981).

Consider carefully the length of the interview. A good rule of thumb for the length of consumer surveys is about 15–20 min, the attention span of most adults. The issue is not the number of questions, but the time commitment. Boredom and extraneous factors will begin to take a toll on the quality of responses if the time required is too long (Schaefer, 1979). Questionnaires that are too long will annoy respondents and may generate negative responses due to declining interest and change of attitude toward the interview. Length problems can arise when many individuals, perhaps both in research and marketing, have input into the survey issues and thus the number of questions gets too large. A good test for the necessity of each question is whether it is an issue that *you need to know* or whether it would simply be *nice to know*. The test procedure should not incur any costs to the participant like return postage for a mailed questionnaire. Prepaid incentives can reduce nonresponse rates (Armstrong, 1975; Furse et al., 1981).

15.5.2 Questionnaire Flow: Order of Questions

When designing a questionnaire, it is useful to make a flow chart of the topics to be covered. The flow chart can be very detailed and include all skip patterns or it may simply list the general issues in order. A flow chart can be very helpful to clients and other personnel who review the instrument before the actual test. It allows them to see the overall plan of the interview.

The primary rule for questionnaire flow is to go from the general to the specific. With food and consumer product testing, this requires asking about the person's overall opinion of the product first. An overall opinion question is recommended using the 9-point balanced hedonic scale. It is immediately followed by open-ended probing of the reasons for liking or disliking with an appropriate "skip pattern." The skip pattern drops to reasons for liking if the respondent was positive and then probes any dislikes. Conversely, the skip pattern next probes any reasons for disliking if the person was negative and then follows up to see if there were any positive characteristics. Open-ended questions (discussed further below) provide an important opportunity to get at some reasons for likes and dislikes in the respondents own words before other issues are brought to mind. Next, more specific attributes are investigated through the use of intensity, just-right, or liking scales (e.g., appeal of taste, appearance, texture). Finally, overall satisfaction or some other correlated index of liking can be checked again at the end and a preference question asked if more than one product has been tested.

The important principle here is to ask overall acceptability first, before specific issues are raised. These issues may not have been on the person's mind and may take on false importance if asked before the overall acceptance question. Respondents will try to figure out what the issues are and give the right answer or try to please the interviewer (Orne, 1962; Orne and Whitehouse, 2000). As the interview is a social interaction, respondents will adjust their answers to what they feel is appropriate for the situation (Boyd et al., 1981). They will begin to use the terminology that you introduce in the specific attribute questions (Sorensen, 1984). Also, consumers are naturally in an integrative frame of mind when thinking about a product (Lawless, 1994). Questions about individual attributes may cause respondents to become unrealistically analytical.

In addition to the screening questions that qualify a person for participation, other personal information can be gathered at the time of the final interview. Demographic information about personal characteristics such as age and income, number of family members, residence, occupation, and so on can be collected. Some of this material is of a sensitive nature. Participants may feel some reluctance about disclosing their income level, for example. Before the test

proceeds, it is important to assure them of the confidential nature of the data and its end use. It is also best to ask sensitive demographic questions last. At that time the participant should feel comfortable and familiar with the interview process and may feel some rapport with the interviewer. Since they have already committed to answering questions during the product-oriented phase of the interview, it should seem natural to simply continue along and answer a few questions about their personal situation.

In summary, the questioning should follow this flow under most circumstances (Sorensen, 1984): (1) screening questions to qualify the respondent if they have not been previously qualified; (2) general acceptability; (3) open-ended reasons for liking or disliking; (4) specific attribute questions; (5) claims, opinions, and issues; (6) preference if a multi-sample test and/or rechecking acceptance via satisfaction or other scale; (7) sensitive personal demographics. An example of a well-designed questionnaire for a consumer product test is shown in Appendix 3.

15.5.3 Interviewing

Participating in a few interviews can be a valuable exercise to get an impression of how the questionnaire flows in practice, as well as providing an opportunity to interact with actual respondents and get an appreciation of their opinions and concerns first hand rather than exclusively in a data summary. Of course, this is a time-consuming process and the potential value must be weighed against other uses of professional time. If the sensory professional actually takes part in the interview process, there are several guidelines to keep in mind. Remember to introduce yourself. Establishing some rapport with the respondent is useful in getting them to volunteer more ideas. A small to moderate degree of social distance may provide the most unbiased results (Dohrenwend et al., 1968). Second, be sensitive to the time demands of the interview. Try not to take more time than expected. If asked, inform the respondent about the approximate length of the interview (Singer and Frankel, 1982). This may hurt the overall agreement rate (Sobal, 1982), but will result in fewer terminated interviews. Third, if conducting an in-person interview, be sensitive to body language. Probe issues if there are signs of discomfort.