The International Center for Research on the Management of Technology

The Voice of the Customer

Abbie Griffin[1]
John R. Hauser[2]

October 1991 **WP # 56-91**

Sloan WP# 3449-92

[1] Assistant Professor of Marketing and Production, Graduate School of Business, University of Chicago
[2] Kirin Professor of Marketing, Sloan School of Management, MIT

© 1991 Massachusetts Institute of Technology

Sloan School of Management
Massachusetts Institute of Technology
38 Memorial Drive, E56-390
Cambridge, MA 02139-4307

ABSTRACT

In recent years, many US and Japanese firms have adopted Quality Function Deployment (QFD). QFD is a total-quality-management process in which the "voice of the customer" is deployed throughout the R&D, engineering, and manufacturing stages of product development. For example, in the first "house" of QFD, customer needs are linked to design attributes thus encouraging the joint consideration of marketing issues and engineering issues. After reviewing QFD, this paper focuses on the "Voice of the Customer," that is, the tasks of identifying customer needs, structuring customer needs, and providing priorities for customer needs.

In the *identification* stage, we address the questions of (1) how many customers need be interviewed, (2) how many analysts need to read the transcripts, (3) how many customer needs do we miss, and (4) are focus groups superior to one-on-one interviews? In the *structuring* stage the customer needs are arrayed into a hierarchy of primary, secondary, and tertiary needs. We compare group consensus (affinity) charts, a technique which accounts for most industry applications, with a technique based on customer-sort data. In the stage which *provides priorities* we review existing research on measuring and estimating "importances." We then present new data in which product concepts were created by product-development experts such that each concept stressed the fulfillment of one primary customer need. Customer interest in and preference for these concepts are compared to measured and estimated importances. We examine data to address the question of whether frequency of mention can be used as a surrogate for importance. We examine the stated goal of QFD, *customer satisfaction*. Our data demonstrate a self-selection bias in satisfaction measures that are used commonly for QFD and for corporate incentive programs.

We close with nine application vignettes to illustrate how product-development teams have used the voice of the customer to modify existing products and services or to create new products and services.

Many leading US firms are focusing on total quality management techniques. For example, 106 firms applied this year for the Baldrige Award (the national quality award) -- an application process that is tedious, costly, and time-consuming but carries tremendous prestige for the winner. There were 180,000 requests in 1990 for copies of the Baldrige criteria (NIST 1991, Reimann 1991) and another 190,000 in 1991 (NIST, personal communication). This interest is based on the belief that quality improvements lead to greater profitability. For example, based on a study of the Baldrige finalists, the General Accounting Office (GAO 1991) suggests that those firms which adopt and implement total quality management tend to experience improved market share and profitability, increased customer satisfaction, and improved employee relations[1].

One aspect of the focus on total quality management has been the widespread adoption of Quality Function Deployment (QFD)[2]. QFD is a product (service) development process based on interfunctional teams (marketing, manufacturing, engineering, and R&D) who use a series of matrices, which look like "houses," to deploy customer input throughout design, manufacturing, and service delivery. QFD was developed at Mitsubishi's Kobe shipyards in 1972 and adopted by Toyota in the late 1970s. In part, because of claims of 60% reductions in design costs and 40% reductions in design time (see Hauser and Clausing 1988), it was brought to the US in 1986 for initial applications at Ford and Xerox. By 1989 approximately two dozen US firms had adopted QFD for some or all of their product and service development. We estimate that in 1991 there are well over 100 firms using some form of QFD. (For those readers unfamiliar with QFD we provide a brief review in the next section of this paper.)

From the perspective of marketing science, QFD is interesting because it encourages other functions, besides marketing, to use, and in some cases perform, market research. Each of these functions brings their own uses and their own demands for data on the customer's "voice." For example, engineers require greater detail on customer needs than is provided by the typical marketing study. This detail is necessary to make specific tradeoffs in engineering design. For example, the auto engineer might want data on customer needs to help him (her) place radio, heater, light, and air-conditioning controls on the dashboard, steering column, and/or console. However, too much detail can obscure strategic design decisions such as whether the new automobile should be designed for customers interested in sporty performance or for customers interested in a smooth, comfortable ride. Because QFD is an interfunctional process it requires market research that is useful for both strategic decisions (performance vs. comfort) and for operational decisions (placement of the cruise control).

To address both strategic and operational decisions, industry practice has evolved a form of customer input that has become known as the "Voice of the Customer." The voice of the customer is a hierarchical set of "customer needs" where each need (or set of needs) has

---

[1] If only those firms that do well on these criteria can be expected to apply, then this data may contain some self-selection bias. However, the report and the congressional reaction to it (Stratton 1991) are indicative of the national interest in quality.

[2] Among the US and Japanese firms reporting QFD applications in 1989 were General Motors, Ford, Navistar, Toyota, Mazda, Mitsubishi, Procter & Gamble, Colgate, Campbell's Soup, Gillette, IBM, Xerox, Digital Equipment Corp., Hewlett-Packard, Kodak, Texas Instruments, Hancock Insurance, Fidelity Trust, Cummins Engine, Budd Co., Cirtek, Yasakawa Electric Industries, Matsushita Densko, Komatsu Cast Engineering, Fubota Electronics, Shin-Nippon Steel, Nippon Zeon, and Shimizu Construction.

assigned to it a priority which indicates its importance to the customer. The voice of the customer becomes a key criterion in total quality management. For example, the National Institute of Standards and Technology (NIST) states as the first key concept in the Baldrige Award criteria that "quality is based on the customer (NIST 1991, p. 2)." See also Juran (1989).

This paper focuses on the customer input used for new-product development. We adopt industry terminology for the customer input and we work within the QFD framework. Marketing readers will notice a similarity between many of the QFD constructs and those that have long been used in marketing. One goal of our paper is to introduce the problems and challenges of QFD to the marketing audience. Another goal is to present new data on some of the techniques that are commonly used by industry.

Following the philosophy of total quality management, we focus on incremental improvement of the techniques for QFD's customer input. In most cases we draw from the rich history of research in marketing and focus on the changes and modifications that are necessary for QFD. We cite new data on comparisons that we have made. Naturally, we can not compare all the possible techniques for any given step in the customer input. Instead, based on our experience over the past four years with over twenty-five US corporations[3] and based on our discussions with market research suppliers, we focus on those techniques that are applied most often in the QFD framework. Because our comparative research provides incremental improvement, it is never completed. Based on the data presented in this paper we fully expect that other researchers will experiment with other techniques and provide incremental improvements relative to the techniques we report.

The structure of this paper is as follows. We begin with a review of QFD and the voice of the customer. We define customer needs and we indicate briefly how they are tied to design goals and design actions. We then focus on each of the three steps in the measurement and analysis of QFD's customer input: (1) identifying customer needs, (2) structuring customer needs, and (3) setting priorities for customer needs. Because QFD's voice of the customer should help the product development team understand how to satisfy the customer, we close with some data on QFD's stated goal of customer satisfaction. We format our presentation within each section around those research questions that we have heard most often in applications (and for which we have data to address).

QUALITY FUNCTION DEPLOYMENT – A BRIEF REVIEW

There is a well-established tradition of research in the management of technology that suggests that cooperation and communication among marketing, manufacturing, engineering, and

---

[3] These applications include computers (main-frame, mid-range, work stations, and personal), software, printers, cameras, airline service, paints, surgical instruments, diagnostic instruments, office equipment, consumer products, tools, retirement plans, movie theaters, health insurance, distribution networks, automobiles and automobile subsystems and components.

R&D leads to greater new-product success and more profitable products. Some of this evidence is based on large questionnaire studies with hundreds of firms as data points, e.g. Cooper (1984a), some of this evidence is based on multi-year projects covering more than fifty firms, e.g., Souder (1987, 1988), and some is based on in-depth ethnographic analysis of a relatively few firms, e.g., Dougherty (1987). The evidence is consistent and persuasive. If engineering, manufacturing, and R&D understand customer needs and if marketing understands how customer needs can be linked to product or service changes, then a product or service is likely to be profitable. See Cooper (1983, 1984a, 1984b), Cooper and de Brentani (1991), Cooper and Kleinschmidt (1987), Dougherty (1987), de Brentani (1989), Griffin and Hauser (1991b), Gupta, Raj and Wilemon (1985), Hise, O'Neal, Parasuraman and McNeal (1990), Moenaert and Souder (1990), Pelz and Andrews (1966), Pinto and Pinto (1990), Souder (1978, 1987, 1988), and others.

QFD improves communication among the functions by linking the voice of the customer to engineering, manufacturing, and R&D decisions. It is similar in many ways to the new product development process in marketing (Pessemier 1982, Shocker and Srinivasan 1979, Urban and Hauser 1980, Wind 1982), the Lens model (Brunswick 1952, Tybout and Hauser 1981), and benefit structure analysis (Myers 1976). For example, like these marketing processes QFD uses perceptions of customer needs as a lens by which to understand how product characteristics and service policies affect customer preference, satisfaction, and, ultimately, sales. One advantage of QFD is that is uses a visual data-presentation format that both engineers and marketers find easy to use. This format provides a natural link among functions in the firm. Since its development in 1972, QFD has evolved continuously to meet the usage requirements of the product-development teams.

QFD uses four "houses" to present data. As shown in figure 1 the first house, the "House of Quality," links **customer needs** to **design attributes**. Design attributes are engineering measures of product performance. For example, a computer customer might state that he (she) needs something which makes it "easy to read what I'm working on." One solution to this need is to provide computer customers with monitors for viewing their work. Design attributes for the monitor might be physical measurement for the illumination of alphanumeric characters, for the focus of the characters, for the judged readability at 50 centimeters (on an eye-chart-like scale), etc.

The second house of QFD links these design attributes to **actions** the firm can take. For example, a product-development team might act to change the product features of the monitor. The product-development team can affect the design attribute of readability at 50 centimeters (as measured by an eye-chart scale) by changing the number of pixels, the size of the screen, the intensity of the pixels, the refresh rate, or whether the monitor is interlaced or not[4]. One action which might affect the design attributes is to change the material in the monitor's screen. A

---

[4] A pixel is a dot on a screen, for example a standard VGA monitor has 640 by 480 pixels while an XGA monitor has 1024 by 768 pixels. A monitor is interlaced if all of the odd rows of pixels are activated and then all of the even rows of pixels are activated. Refresh rate is the number of times per second the pixels are reactivated.

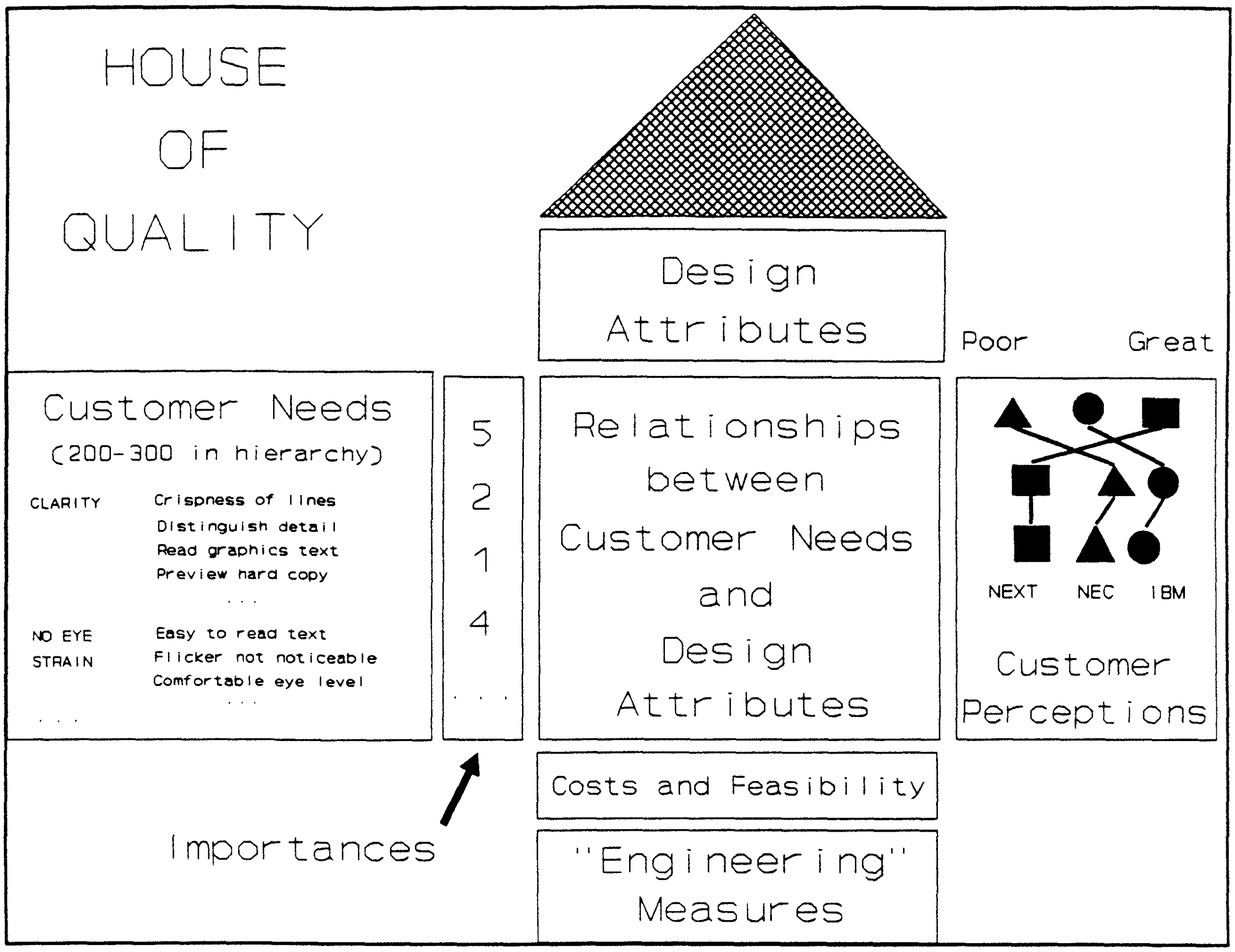

Figure 1. The House of Quality from Quality Function Deployment

more radical action might eliminate the monitor and provide a system which projects the work on a wall or on very small stereoscopic screens which the user wears as goggles[5].

The third house of QFD links actions to **implementation** decisions such as manufacturing process operations. For example, the third house might be used to identify and select the manufacturing procedures that produce a monitor with a target refresh rate or the manufacturing procedures that produce the material that was selected for the monitor's screen. The final house of QFD links the implementation (manufacturing process operations) to **production planning**.

We begin by describing the customer input to the House of Quality. We then review briefly the other aspects of the House of Quality. For greater detail see Clausing (1986), Eureka (1987), Griffin (1989), Hauser and Clausing (1986), King (1987), Kogure and Akao (1983), McElroy (1987), and Sullivan (1986, 1987), as well as collections of articles in Akao (1987) and

[5] MIT's Media Laboratory is working on such "virtual reality" solutions.

the American Supplier Institute (1987).

The Voice of the Customer

Customer needs. QFD lists customer needs on the left side of the house. A customer need is a description, in the customer's own words, of the benefit that he, she, or they want fulfilled by the product or service. For example, when describing lines on a computer monitor a customer might want them "to look like straight lines with no stair-step effect." Note that the customer need is not a solution, say a particular type of monitor (VGA, Super VGA, XGA, Megapixel, etc.), nor a physical measurement (number of noticeable breaks in the line), but rather a detailed description of how the customer wants images to appear on the monitor. The distinction has proven to be one of the keys to the success of QFD. If the product-development team focuses too early on solutions, they might miss creative opportunities. For example, the Hewlett-Packard (H-P) Laserjet III was a commercially successful product which enhanced graphics printing dramatically by rearranging the dots on the page. Suppose that H-P focused too quickly on a laser device that simply increased the number of dots per inch on the page (a design attribute). The clarity of the graphics would have been superior to the Laserjet II, but H-P would have missed an opportunity to develop a printer that was less costly and more effective. H-P would also have delayed the time to market of the Laserjet III. (The Laserjet III was introduced at a price that was lower than the Laserjet II.)

If the team focuses too quickly on physical measurements, they miss an understanding of all the influences on customer needs. For example, a computer-monitor team might be tempted to focus on the size of the monitor (12", 14", 16") to affect the size of the alphanumeric characters on the screen. However, the size of the alphanumeric characters is only one of the design attributes that affects the customer need of "easy to read text." The readability of a text string also depends on the ambient room light and reflections, the colors that the software designer chooses, the ratio of the height of small letters to that of capital letters, and even the style of the typeface (serif or sans-serif, proportional or fixed, etc.). All of these design attributes interact with the size of the monitor to affect the customer need of "easy to read text." Some may be less costly and more effective, some may be synergistic with changing the monitor's size, but all should be considered before a final design is chosen for the monitor.

Discussions with customers usually identify 200-400 customer needs. These customer needs include basic needs (what a customer assumes a monitor will do), articulated needs (what a customer will tell you that he, she, or they want a monitor to do), and exciting needs (those needs which, if they are fulfilled, would delight and surprise the customer). See Lillrand and Kano (1989).

Hierarchical structure. Not everyone on the product-development team works with the detail that is implied by a list of 200-400 customer needs. QFD structures the customer needs into a hierarchy of primary, secondary, and tertiary needs[6]. Primary needs, also known as strategic needs, are the five-to-ten top-level needs that are used by the team to set the strategic direction for the product or service. For example, the primary needs help the product-development team decide whether to develop a computer viewing system that emphasizes clarity and resolution, ease of viewing, viewing interactiveness, or visual impact.

Secondary needs, also known as tactical needs, are elaborations of the primary needs -- each primary need is elaborated into three-to-ten secondary needs. Secondary needs indicate more specifically what the team must do to satisfy the corresponding primary (strategic) need. For example, if clarity is the primary need, then the secondary needs tell the team how the customer judges clarity, say by the crispness of the lines, the ability to distinguish detail on all parts of the screen, the ability to read graphically generated text, and the ability of the user to see what he (she) will get on hard copy. These tactical needs help the team focus their efforts on those more-detailed benefits that fulfill the strategic direction implied by the primary need. Typically, the 20-30 secondary needs are quite similar to the 20-30 "customer attributes" that are common in marketing research and that often underlie perceptual maps. (See Green, Tull and Albaum 1988, Lehmann 1985, or Urban and Hauser 1980).

The tertiary needs, also known as operational needs, provide the detail so that engineering and R&D can develop engineering solutions that satisfy the secondary needs. For example, a person may judge the crispness of a line (a secondary need) by the following tertiary needs: the lack of a stair-step effect, the ability to distinguish lines from background images and text, and the ability to distinguish among individual lines in a complex drawing. These operational needs provide the detail which enables the engineer to make tradeoffs among alternative designs.

Importances. Some customer needs have higher priorities for customers than do other needs. The QFD team uses these priorities to make decisions which balance the cost of fulfilling a customer need with the desirability (to the customer) of fulfilling that need. For example, the strategic decision on whether to provide improved clarity, improved ease of viewing, or some combination will depend upon the cost and feasibility of fulfilling those strategic needs and the importances of those needs to the customer. Because the importances apply to perceived customer needs rather than product features or engineering solutions, the importance measurement task is closer to marketing's "expectancy value" tradition (e.g., Wilkie and Pessemier 1973) than to the conjoint tradition (e.g., Green and Srinivasan 1978), however recent hybrid techniques (Green 1984, Green and Srinivasan 1990, Wind, et. al. 1989) have blurred that distinction.

Customer perceptions. Customer perceptions are a formal market-research measurement of how customers perceive products that now compete in the market being studied. If no

[6] When necessary the hierarchy can go to deeper levels. For example, when Toyota developed a QFD matrix to help them eliminate rust from their vehicles, the hierarchy had eight levels (Eureka 1987). For example, the lowest level included a customer need relating to whether the customer could carry rotten apples in the bed of a pick-up truck without worrying about the truck body rusting.

product yet exists, the perceptions indicate how customers now fulfill those needs. (For example, existing patterns of medical care served as generic competition for health maintenance organizations in a study by Hauser and Urban 1977.) Knowledge of which products fulfill which needs best, how well those needs are fulfilled, and whether there are any gaps between the best product and "our" existing product provide further input into the product-development decisions being made by the QFD team. Furthermore, if the team compares its perceptions to those of the customer, the team can identify and overcome organizational biases. In the voice of the customer, customer perceptions are measured by any of a variety of standard market-research scales. We have seen applications of Likert-like scales and semantic differential scales. We refer the reader to market research texts such as Green, Tull and Albaum (1988) or Lehmann (1985) for examples of such scales.

Segmentation. In many applications, the product-development team will focus on one particular segment of the customer population. A complete "voice" will be obtained for each segment. In other applications, only the importances will be different for different segments. The issue of segmentation is an important research topic, however, for the purposes of this paper, we assume that the team has already decided to focus on a particular customer segment.

Engineering Input

Design attributes. After the product-development team identifies the customer needs, the team lists those measurable aspects of the product or service which, if modified, would affect customer's perceptions. Often a particular design attribute can affect many customer needs. For example, the illumination of a monitor screen can affect the clarity of text, the clarity of graphics, eye strain, and even power requirements of the computer system.

Engineering measures. The QFD team obtains objective measures of existing products (their product and competitors) on the design attributes.

Relationship matrix. The QFD team judges which design attributes affect which customer needs and by how much. Normally, the team specifies only the strongest relationships leaving most of the matrix blank (60-70% blank). While it is possible to undertake experiments to determine the strength of the relationship for some key elements of the matrix, the sheer number of relationships usually means that judgment is used for the majority of the entries. For example, a matrix with 200 customer needs and 100 design attributes would require 6,000 entries even if only 30% of the entries were necessary. The hierarchical structure of the customer needs is used in some cases to reduce further the number of entries that the team must judge. Naturally, when "hard" data is available from experiments or conjoint analysis, it is used.

Roof matrix. Finally, the "roof matrix" specifies the engineering relationships among the design attributes. For example, engineering realities might mean that increasing the illumination of the screen decreases the life of the screen material or the speed of screen refreshes. Such design interactions are quantified in the roof matrix. The roof matrix gives the "house" its distinctive shape as indicated by cross-hatched lines in figure 1.

Other data. Most applications include rows in the matrix which summarize the projected costs and technical difficulty of changing a design attribute.

Using the House of Quality

By collecting in one place information on both customer needs and engineering data on fulfilling those needs, the House of Quality forces the interfunctional product-development team to come to a common understanding of the design issues. In theory, the goal of a House-of-Quality analysis is to specify targets values for each of the design attributes. However, different teams use the house in different ways. In some cases it is central to the design process and is used to make every decision, in others its primary function is communication, and in still others formal arithmetic operations provide formal targets for the design attributes. For example, some teams multiply the importances times gaps in customer perceptions (best competitor vs. our product) to get "improvement indices." Other teams multiply importances times the coefficients in the relationship matrix to get imputed importances for the design attributes[7]. For the purposes of this paper, we accept the structure of the House of Quality as given. We attempt to get the best customer input for use in the house.

In closing this section, we note that the QFD seems to work. In a study of 35 projects Griffin (1991) reports that QFD provided short-term benefits (reduced cost, reduced time, increased customer satisfaction) in 27% of the cases and long-term benefits (better process or better project) in 83% of the cases. Griffin and Hauser (1991a) report that, in a head-to-head comparison with a traditional product-development process, QFD enhanced communication among team members. Collections of articles by Akao (1987) and the American Supplier Institute (1987) contain many case studies of successful applications. In the final section of this paper we provide nine application vignettes to illustrate how the voice of the customer (through QFD) has been applied in industry. We now present data to address some methodological questions that we have encountered.

IDENTIFYING CUSTOMER NEEDS

Identifying customer needs is primarily a qualitative research task. In a typical study between 10 and 30 customers are interviewed for approximately one-hour in a one-on-one setting. For example, a customer might be asked to picture himself (herself) viewing work on a computer. As the customer describes his or her experience, the interviewer keeps probing, searching for better and more complete descriptions of viewing needs. In the interview the customer might be asked to voice needs relative to many real and hypothetical experiences. The

---

[7] In this paper we do not discuss these formal operations other than to note that they assume certain scale properties of the importances and the perceptions. This is clearly an opportunity for further research.

interview ends when the interviewer feels that no new needs can be elicited from that customer. Interviewers might probe for higher-level (more strategic) needs or for elaborations of needs as in the laddering and means-ends techniques (Gutman 1982, Reynolds and Gutman 1988). Other potential techniques include benefit chains (Morgan 1984), subproblem decomposition (Ruiz and Jain 1991), and repertory grids (Kelly 1955). While many applications use one-on-one interviews, each of these techniques can be used with focus groups (Calder 1979) and with mini-groups of two-to-three customers.

The three questions which we have heard most often are: (1) Do group synergies identify more customer needs? (2) How many people (groups) must be interviewed? and (3) How many team members should analyze the data?

Groups vs. One-on-One Interviews

Many market research firms advocate group interviews (see also Calder 1979) based on the hypothesis that group synergies produce more and varied customer needs as each customer builds upon the ideas of the others. However, Calder also cautions that focus groups tend to produce intersubjectivity, that is, the *shared* belief structure among customers rather than intrasubjectivity, the beliefs that may be special to some but not all customers. Another concern about focus groups is that "air-time" is shared among the group members. If there are eight people in a two-hour group then each person talks, on average, for about 15 minutes. Some market research firms have argued that this is not sufficient time to probe for a complete set of customer needs.

We were able to compare focus groups to one-on-one interviews in a proprietary QFD application. The product category was a complex piece of office equipment. In this application, the QFD team obtained customer needs from eight two-hour focus groups and nine one-hour interviews. (The data were collected by an experienced, professional market research firm.) The entire set of data was analyzed by six professionals to produce a combined set of 230 customer needs. Silver and Thompson (1991) then analyzed the data to determine, for each customer need and for each group or individual, if that group or individual voiced that need.

Figure 2 plots the data. For example, the first point for the one-on-one plot indicates that, on average, a single one-on-one interview identified 33% of the 230 needs. The second point indicates that, on average, two one-on-one interviews identified 51% of the customer needs. The average is taken over all combinations of two interviews.

The data in figure 2 suggest that while a single two-hour focus group identifies more needs than a one-hour one-on-one interview, it appears that two one-on-one interviews are about as effective as one focus group (51% vs. 50%) and that four interviews are about as effective as two focus groups (72% vs. 67%). As one manager said when he examined the data, it is almost as if an hour of transcript time is an hour of transcript time independently of whether it comes from a one-on-one interview or a focus group. If it is less expensive to interview two consumers for an hour each than to interview six-eight customers in a central facility for two hours, then figure 2 suggests that one-on-one interviews are more cost-efficient. At minimum, figure 2 suggests that group synergies do not seem to be present in this data.

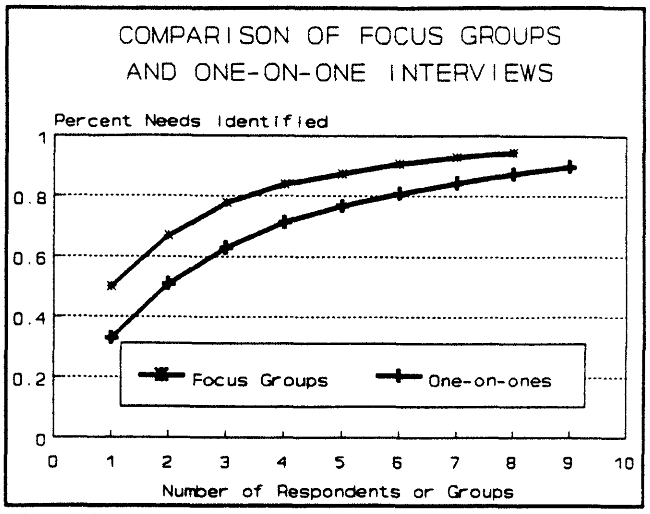

Figure 2. Focus Groups vs. One-on-One Interviews for Office Equipment (from Silver and Thompson 1991)

How many customers?

We would like to know how many customers need be interviewed to identify 90% of the customer needs. Besides intellectual curiosity, there are many reasons for industry to seek an answer to this question. First there is the cost. While the field costs per interview are moderate, the implicit analysis costs are quite high. Due to the communications aspects of QFD, it is typical for team members to observe the interviews and for four or more team members to read the transcripts. One major US firm estimates that the typical out-of-pocket costs for 30 interviews are only $10-20,000 but that the implicit team costs include over 250 person-hours to observe the interviews, read the transcripts, and summarize the customer needs. Even based on a low estimate of $100 per person-hour (fully-loaded) for professional personnel, this means that the total costs per interview are in the range of $1-2,000. If you multiply this by 5-10 segments (typical in a complex category) and 5-10 major product lines within a firm, then the cost savings of setting a policy of 20 customers per segment rather than 30 customers per segment can be substantial ($250,000-$2,000,000).

Another cost incurred if too many interviews are used is the time delay. Because the timely introduction of new products is considered important in today's competitive environment, product-development teams seek to avoid unnecessary delays in data collection. Some of these delays are market research time (recruiting and interviewing), but much of the delay is the time the team devotes to observing and analyzing the transcripts. There is a high opportunity cost for the teams' time.

On the other hand there are benefits to more interviews. The goal of total quality management and the philosophy of QFD is to base product development on customer needs. In one application a $1 billion investment in a new-car program depended upon the choice of which strategic (primary) customer need to stress. In another application, a service firm was able to gain an additional $150 million in profit by reallocating operating procedures. They had been stressing the fulfillment of a customer need that was less important than they had thought. They reallocated resources to fulfilling other more-important, but less-costly-to-fulfill customer needs. In both applications the product-development team had to defend their recommendations to managers who would bear responsibility for the profit implications of the decisions. The teams were asked to certify that the initial list of needs was based on a sufficient number customers.

There is also the issue of "exciting" needs. The fulfillment of exciting needs can make or break a product-development program. For example Kao developed a highly-concentrated laundry detergent, Attack, that fulfills the needs of Japanese customers and retailers for a product that takes significantly less space to store. This product (and imitators) now command a significant fraction of Japanese sales. After the fact such exciting needs are often obvious. Most customers, once such needs are pointed out to them, will agree that they are important. But, by definition, such needs are voiced by relatively few customers. Usually such needs are not voiced because customers do not consider their fulfillment a realistic possibility. Thus, firms want to be confident that they have interviewed enough customers to uncover most of the exciting needs.

Firms would like to balance the cost of additional interviews with the benefits of identifying a complete set of needs. However, the analysis cost (in person-hours) to make this tradeoff is quite high. A complete coding of transcripts to identify, for every interview and for every need, whether that customer voiced that customer need is tedious and time-consuming. We have found few product-development teams willing to undertake such analyses for a typical QFD project. However, because of the general interest in obtaining a "ballpark" estimate of the number of customers required, we have obtained funding[8] for two applications -- the office equipment application described above and a low-cost durable application described below. We first describe the data and the analysis for the low-cost durable and then review the results for office equipment.

The data. We interviewed 30 potential customers of portable food-carrying and storing devices (coolers, picnic baskets, knapsacks, bike bags, etc.). The interviews were transcribed and each interview was read by seven analysts. The needs were merged across analysts and customers and redundancy was eliminated to obtain a core list of 220 needs. We recorded which customers and which analysts identified each need. Naturally, some needs were mentioned by more than one customer. See figure 3. For example, 38 needs were identified by one customer out of thirty, 43 needs were identified by two customers out of thirty, 29 needs by three customers out of thirty, etc. One need was identified by 24 of the thirty customers.

To get an idea of how many needs we would have obtained from interviewing fewer customers, we consider all possible orderings of the thirty customers and determine the average percent of non-redundant needs we would have obtained from n customers for n = 1 to 30. (Note that we are temporarily defining 100% as that obtained from 30 customers. We address missing needs below.) Because the number of possible orderings, 30!, is a very large number, we randomly sampled 70,000 orderings. The results, plotted in figure 4 as "observed," show that interviewing 20 customers identifies over 90% of the needs provided by 30 customers.

To generalize to more than thirty customers we need a model. We draw upon a model developed by Vorberg and Ulrich (1987, p. 19) and define for a given customer, c, and a given

[8] Funds for the low-cost durable study were obtained from M.I.T.'s program for the Management of Technology, the Kirin Brewing Company, the Marketing Science Institute, and the Industrial Research Institute. The office equipment manufacturer funded that study but wishes to remain anonymous.

customer need, i, the probability, p_i, that customer c voices need i at least once during the interview. In our data we observe the outcome of this binomial process. That is, we observe whether or not customer c voices customer need i. This model is related to Morrison's (1979) search model[9] and to concepts developed by Dawkins (1991) and Efron and Thisted (1976).

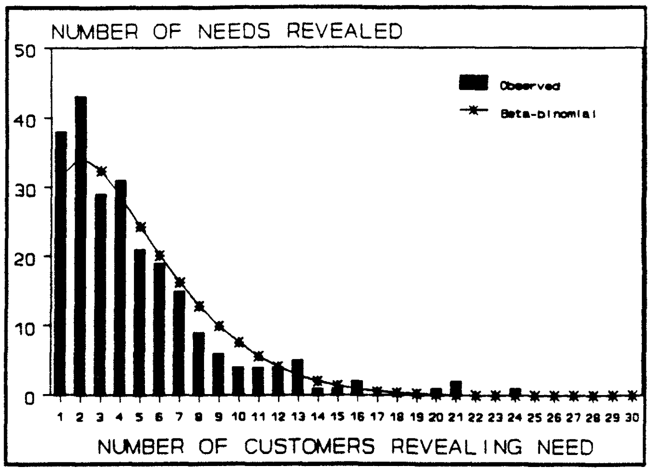

Figure 3 Number of Customers who Identify a Need

For thirty customers we observe the outcome of thirty binomial processes. Thus, for thirty interviews we observe how many customers voiced need i. We simplify the model by assuming that customers are more or less equivalent in their ability to articulate needs. Then for each need, i, we can consider our customers as thirty successive random draws from the same binomial distribution. We now assume that the probabilities, p_i, are described by a Beta distribution across customer needs[10]. This assumption, combined with the binomial processes, gives a Beta-binomial distribution for the number of times that needs are voiced in the thirty interviews. The best-fit Beta-binomial distribution[11] is plotted in figure 3. While not perfect it does appear to be a reasonable model[12].

Analysis. We use the Beta-binomial model in figure 3 to estimate the average number of needs obtained from n customers. Consider need i with probability p_i. For n independent draws from a binomial distribution, the probability that customer need i is identified is simply $1 - (1-p_i)^n$. (Each customer interview is considered an independent draw.) However, the probabilities, p_i, are distributed by the Beta distribution. Thus, if the Beta-binomial distribution[13] has parameters α and β, then the expected value, E_n, of the probability of observing a need from n customers is:

---

9 Morrison (1979) assumes that needs are voiced with Poisson rate, λ_i. Then the probability that a need is voiced at least once is $p_i = 1 - e^{-\lambda_i t}$.

10 Note that we "flip" the normal Beta-binomial analysis. In most applications (e.g., Greene 1982) the customer probabilities are Beta distributed across customers; in our model customers are replications. In our model the probabilities, p_i, are Beta distributed across customer needs, i.

11 Morrison (1979) shows that there exists a $G(\lambda_i)$ such that p_i is Beta distributed. Because the Beta distribution appears to fit the data we prefer to work directly with p_i rather than λ_i.

12 If we smooth the small "lump" at 21 customers, the observed frequencies are not statistically different than the Beta distribution (Kolmogorov-Smirnov test). We feel this "lump" does not seriously impair the model. Note that we can also assume that customers are heterogeneous in their abilities to voice needs. However, we feel that the assumption of two forms of heterogeneity complicates the model needlessly. Our data are available should anyone wish to extend the model in this direction.

13 The Beta distribution is given by $f(p) = p^{\alpha-1}(1-p)^{\beta-1}/B(\alpha,\beta)$ where $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$. Method of moments estimation gives $\alpha = 1.45$ and $\beta = 7.64$.

$$E_n = 1 - \frac{\Gamma(n+\beta)\Gamma(\alpha+\beta)}{\Gamma(n+\alpha+\beta)\Gamma(\beta)} \qquad (1)$$

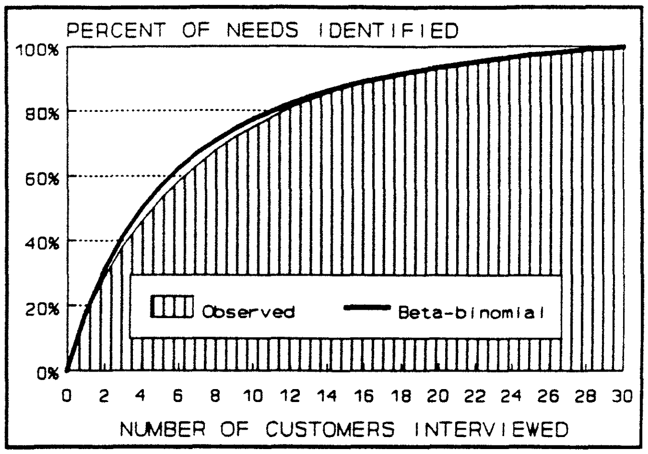

Figure 4 Percent of the Customer-Needs Identified by N Customers (where 30 Customers = 100%)

Figure 4 plots equation 1 for α and β estimated for our data. For comparison in figure 4, we have normalized equation 1 to correspond to a percentage of the thirty customer needs. A Kolmogorov-Smirnov test for goodness of fit between the actual and modeled cumulative distributions indicates that they do not differ at a statistical significance level of *0.05*. The analysis is slightly optimistic in the range of two to twelve customers, but fits quite well beyond twelve customers. Since most decisions will be made in the range above twelve customers, the model appears accurate enough for our purposes.

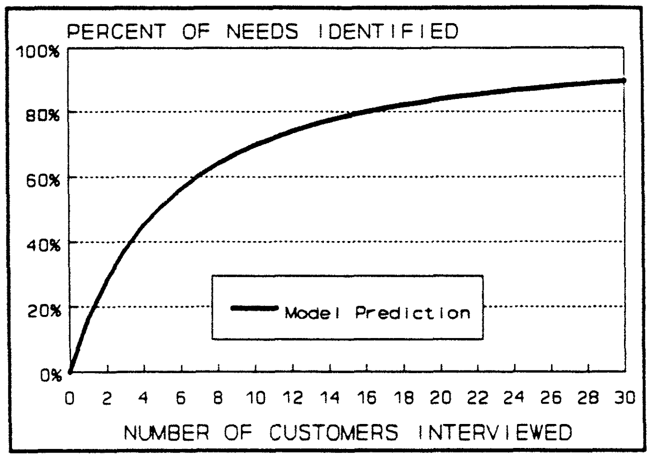

Figure 5 Predictions based on the Beta-binomial Model

What are we missing? While thirty customers produce 100% of our data, they may not produce 100% of the needs. We may have missed those needs which have a low p_i. Fortunately, equation 1 gives a means by which to estimate the magnitude of our error. That is, we estimate the number of needs that were given zero times out of thirty tries. The model estimates that our thirty customers gave us 89.8% of all the needs. The complete plot of E_n is given in figure 5.

Office equipment. The low-cost durable application was completed in 1988. In the past three years interviewing techniques have evolved so that interviewers are more effective in eliciting customer needs. For example, interviewers attempt to keep track of the customer needs voiced by the customers who have been interviewed already. Thus, when they interview a new customer they focus their questions to probe for new customer needs. With the improved interviewing techniques, we expect that fewer customers need be interviewed. Indeed, in the 1991 analysis of office equipment (review figure 2) the Beta-binomial analysis ($\alpha = 1.88$, $\beta = 2.88$) suggests that the nine customers and eight focus groups identified 98% of the customer needs. If one group is equivalent to two interviews then this means that twenty-five "customers" identified 98% of the office-equipment needs. However, we caution the reader that this difference may also be due to the difference in product categories. Hopefully, subsequent applications will supplement the data in figures 2 and 3.

How many analysts?

While many applications assign 4-6 team members to read and analyze transcripts, other applications rely on qualitative expert(s) to read transcripts and identify needs. To test this strategy we asked seven "analysts" to code the transcripts in the low-cost durable application. One was a qualitative expert, two were undergraduate students, and four were engineering development teams who would be using the customer needs in their development efforts. The students and teams, which split the transcripts among themselves, were provided with about 30 minutes of training in identifying customer needs. (This is typical of the amount of training given to corporate product-development team members who use these techniques to identify customer needs.)

On average, the analysts were able to identify 54% of the customer needs with a range of 45%-68% across analysts. The qualitative expert was at the low end of the range while the engineering teams were at the high end. The students were in the middle of the range. The "observed" area in figure 6 represents the average cumulative percent of attributes identified as more analysts read the transcripts. Figure 6 also shows that a Beta-binomial ($\alpha=22$, $\beta=19$) model provides a reasonable fit to the data. Based on the model, we estimate that the seven analysts identified 99% of the customer needs obtainable from the transcripts.

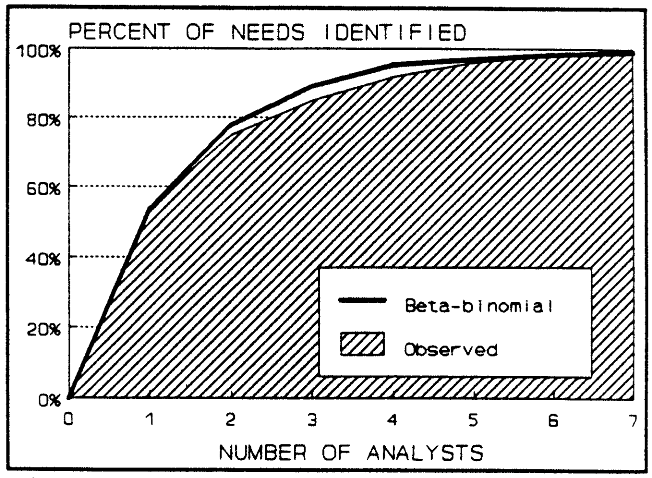

Figure 6 Ability of Analysts to Identify Customer Needs

Besides the low-cost durable study, we have observed many multiple-analyst applications. Analysts with different backgrounds interpret customer statements differently. This variety of perspectives leads to a larger set of customer needs and a richer understanding of the customer than is feasible with a single expert. One hypothesis is that experts may have preconceived notions of what constitutes a customer need. This may cause them to miss surprising or unexpected statements of needs.

If figure 6 is representative of other categories, then (1) more than one analyst should read the transcripts, and (2) it might be more cost-effective to replace "qualitative experts" with a greater number of less-experienced, but trained and motivated, readers. We have observed that the use of product-development-team members brings the added value of team buy-in to the data and greater internalization of the "voice" for later design work. Such ancillary benefits are lost if the team relies on outside experts to interpret the data.

Summary

Based on our data we hypothesize that (1) one-on-one interviews may be more cost-effective than focus groups, (2) that 20-30 interviews are necessary to get 90-95% of the customer needs, and (3) that multiple analysts or team members should read and interpret the raw transcripts. We now compare techniques to structure the customer needs into hierarchies of primary, secondary, and tertiary needs.

STRUCTURING CUSTOMER NEEDS

In this paper we compare the dominant structure-generating method, a group consensus process (affinity charts and tree diagrams), with a proposed customer-based structure-generating method, customer sorting and clustering.

Group Consensus Process

In most American and Japanese applications, customer needs are structured by group consensus using affinity charts (K-J diagrams[14]) and tree diagrams, two of the "Seven New Tools" used in Japanese planning processes (King 1987, Imai 1986). This group consensus process uses the product-development team to impose structure on the customer needs. The advantage of a consensus process is that it assures group buy-in to the structure; the disadvantage is that there is no assurance that the team's structure represents how customers think about their needs or make decisions.

The process we used in our comparison is typical of both American and Japanese applications. To create the affinity chart each team member is given a roughly equal number of cards each bearing one customer need. One team member selects a card from his (her) pile, reads it aloud, and places in on the table (or wall). Other members add "similar" cards to the pile with a discussion after each card. Sometimes the card is moved to a new pile, sometimes it stays. The process continues until the group has separated all the cards into some number of piles of similar cards, where each pile differs from the others in some way. The team then structures the cards in each pile into a hierarchical tree diagram with more detailed needs at lower levels, and more tactical and strategic needs at the upper levels. To select a higher-order need, say a secondary need to represent a group of tertiary needs, the group can either select from among the tertiary needs or add a new card to summarize the group of relevant tertiary needs. Throughout the process the team can rearrange cards, start new piles, or elaborate the hierarchy.

Customer Sort and Cluster Process

Green, Carmone and Fox (1969) and Rao and Katz (1971) applied a technique known as

---

[14] K-J is the registered trademark of Jiro Kawakita for his version of the affinity chart. For the remainder of the paper we use the more generic name.

subjective clustering in which subjects sort stimuli (e.g., television programs) into piles, a similarity matrix is calculated, and a 2- or 3-dimensional similarity map is derived. We modify that data collection procedure to apply to customer needs and then analyze the data to obtain the hierarchical customer-need structures.

In a customer sort process, customers are given a deck of cards, each bearing one customer need. They are asked to sort the cards into piles such that each pile represents similar needs and differs from the other piles in some way. The number of piles and the exact definition of similarity is left unspecified. After completing the sort, each respondent is asked to choose a single need from each pile, called an exemplar, which best represents the customer needs in the pile. From the sort data we create a co-occurrence matrix[15] in which the i-j-th element of the matrix is the number of respondents who placed need i in the same pile as need j. We also label each need with the number of times it was chosen as an exemplar.

To develop a structured hierarchy we cluster[16] the co-occurrence matrix. To name the clusters we use the exemplars. When there is no clearly dominant exemplar within a cluster, we either choose from among the exemplars in the cluster or add a label to the data.

The use of exemplars rather than labels (as in subjective clustering) is an attempt by the product-development teams to maintain as close a link as possible to the actual words used by customers. For example, one might label a group of statements about computer viewing devices as "appropriate ergonomics," but this may be misleading if the customer really said "everything is blurred after a day using my computer." The "blurred-vision" statement provides the product-development team with more realistic clues about product use which the sanitized label does not[17].

The Data

The group-consensus chart was constructed by a team of engineering managers, chosen from M.I.T.'s Management of Technology Program. The team had studied the product category, had read all of the interview transcripts, and had reviewed the list of customer needs. The team was lead by Griffin who had observed and/or participated in almost twenty industry

---

[15] If the number of piles varies dramatically across respondents one can weight the data by a monotonic function (e.g., log[•]) of the number of piles that a respondent uses. This gives a greater weight to respondents who are more discriminating in their sorting task. To assure a simpler and more straightforward comparison we have not included this complication for food-carrying devices.

[16] We have found that Ward's method, the average linkage method, and the complete linkage (farthest neighbor) provided similar structures in our data. See Griffin (1989). For example, when comparing a Ward's-based cluster solution and an average-linkage-based cluster solution, only 3% of the customer needs appeared in different primary groupings. Single linkage (nearest neighbor) led to "chaining" in which customer needs were merged to a large cluster one at a time. Because the difference between the three clustering algorithms is slight, we chose Ward's method for the comparisons in this paper. It is used more often in industry (Romesburg 1984) and, when shown the three solutions, the management team believed that the Ward's structure was slightly superior in terms of face validity to other two. (In Ward's method, clusters are merged based on the criterion of minimizing the overall sum of squared within-cluster distances.) Deciding where to cut the hierarchy remains an exercise in qualitative judgment. However, exemplars help identify the cuts.

[17] In another example, a airline-service team might react one way to the sanitized label of "organized boarding procedures" and a different way to the customer's statement that "we were herded like cattle when we got on the plane."

applications of group-consensus charts at that time. Sixty M.I.T. graduate students who use food-carrying devices participated in the customer sort. Because we funded this data collection ourselves, we report the actual customer needs.

In addition we compared group-consensus charts and customer-sort hierarchies for a major consumer good with almost 200 customer needs. Two group-consensus charts were developed. One by a team at the consumer-products company who had worked on the product category and another by a team of graduate students from M.I.T.'s engineering school. The customer-sort hierarchy was based on a sample of sixty consumers chosen randomly from active users of the product category. Because the data are proprietary, we report summary statistics and our qualitative impressions only.

Finally, we report on a computer-product application in which a team-based consensus chart was compared to a customer-based consensus chart and we report the qualitative experience of approximately fifteen proprietary applications of the customer-sort methodology.

Food-carrying Device Structures

Table 1 compares the top levels of the group-consensus-chart and customer-sort hierarchies for food-carrying devices. (The complete hierarchies are available in Griffin 1989.) Consider first the number of secondary and tertiary needs and the number of exemplars within each primary grouping. The customer-sort technique provides a more even distribution. (The exemplars are used to identify primary and secondary needs. The numbers in table 1 refer to needs which were classified as exemplars by at least 20% of the customers. Fortunately, within any group of needs, there is often a dominant exemplar.) While an even distribution is no guarantee that a hierarchy is better, an even distribution is one of the desirable features for which product-development teams look. An even distribution makes it easier to assign responsibilities. Notice also that twenty-seven labels were added to the group-consensus chart by the development team (247 total needs) while only ten labels were added to the customer-sort hierarchy (230 total needs). This means that more of the customers' semantics are used directly in the primary and secondary levels of the customer-sort hierarchy.

The more interesting comparison is based on qualitative impressions. (Primary labels are shown in table 1.) We have shown these hierarchies to a number of people including the team that created the consensus chart and executives at firms which use the voice of the customer in their product-development processes. In all cases, including the team that did the consensus chart, judgments were that the customer-sort hierarchy provided a clearer, more-believable, easier-to-work-with representation of customer perceptions than the group-consensus charts. Only one of the five group-consensus primary groupings is specific to the category (not generic), while four of the seven customer-sort groupings are specific to the category. The qualitative reaction seems to be summarized by: "The group-consensus chart is a good systems-engineering description of the problem while the customer-sort hierarchy is really the customer's voice."

Table 1 Comparing Group-consensus and Customer-sort Food-Carrying-Device Hierarchies

AFFINITY CHART				CUSTOMER SORT			
PRIMARY NEED	SECD. NEEDS	TERT. NEEDS	EXEMPLARS	PRIMARY NEED	SECD. NEEDS	TERT. NEEDS	EXEMPLARS
Price	4	0	2	Attractiveness	4	20	9
Container Utility	2	14	21	Carries Many Things	4	21	9
Phys. Characteristics	10	30	10	Maintains Temps.	5	39	6
Thermal Attributes	4	34	3	Right Size	3	29	7
Convenience	5	139	6	Easy to Move	2	23	6
				Convenience	2	29	1
				Works as Container	2	30	4
Total	25	217	42		22	201	42
Coeff. of Variation	0.6	1.4	0.9		0.4	0.2	0.5

To compare the hierarchies formally we report two statistical measures of structure similarity. The first is Kruskal's λ (Goodman and Kruskal 1954) which is a measure of association of nominal variables equal to the reduction of error when one hierarchy is used to predict the other. Its maximum value is 1.0. The second measure is a variation of an information theoretic measure to compare different probability matrices (Hauser 1978). Let n_{ij} be the number of needs that are in group i from one hierarchy and group j from another hierarchy. The two hierarchies are identical if the number of groups are equal and the matrix of the n_{ij}'s can be rearranged to be a diagonal matrix. The information measure, U^2, measures the lack of deviation from such a diagonal matrix. It's maximum value is also 1.0. For details see an appendix that is available from the authors. For the primary needs we calculate $\lambda = 0.28$ and $U^2 = 0.30$. For the secondary needs we calculate $\lambda = 0.51$ and $U^2 = 0.63$. Notice that while the two hierarchies differ, the group-consensus chart agrees more with the customer sort at the tactical (secondary) level than at the strategic (primary) level.

Consumer-Product Structures

Qualitatively, the customer-sort hierarchy seems to be superior to the group-consensus chart for food-carrying devices. We sought to replicate this comparison for another category. In this category we were fortunate that an experienced product-development group at a world-class new-product organization developed a group-consensus chart and then tested it with a customer-sort analysis. While similar in most aspects to the above comparison, this comparison differs because (1) the group-consensus chart was developed by category experts and (2) the products in the category are less complex and more familiar to consumers than food-carrying devices. To separate these effects, we had "non-expert" engineering students develop a second group-consensus chart.

As before, the distribution of tertiary needs is more uniform for the customer-sort hierarchy than for the product-development-team consensus chart. Furthermore, the product-development-team consensus chart contained twenty labels that were not in the customer-sort chart. (The student team added fourteen labels.) In retrospect some of these labels obscured the true customer voice.

The statistical comparisons in table 2 suggest that there is more agreement between group-consensus charts and customer-sort hierarchies for the consumer product than for the food-carrying device. We hypothesize that this is due to the less complex nature of the consumer product. It is not totally attributable to the expertise of the professional product-development team because the student team did almost as well as the professional team in their agreement with the customer-sort.

Table 2 Statistical Comparison for Consumer-Product Hierarchies

	PRIMARY NEEDS		SECONDARY NEEDS	
	Kruskal's λ	Information Test	Kruskal's λ	Information Test
Consensus (Develop. Team)--Customer Sort	0.60	0.65	0.53	0.65
Consensus (Student)--Customer Sort	0.59	0.57	0.53	0.62
Consensus (Develop. Team -- Student)	0.62	0.69	0.56	0.68

The most compelling evidence of the customer-sort method's utility is its face validity. The product-development team felt that the customer-sort hierarchy was a better representation of consumer perceptions than either group-consensus chart. After looking at all three structures, the product-development team concluded that the in-house structure reflected the way the firm developed the product (technology by technology). The customer-sort structure, on the other hand, reflected the way customers use the product (function by function). The product-development team chose to use the customer-sort hierarchy for product-development and segmentation activities[18].

Other Applications

In an application to a computer product with 469 customer needs, we compared team-based and customer-based consensus charts. The team sorted the needs into 14 primary and 57 secondary groups while the customers sorted the needs into 11 primary and 50 secondary groups. The coefficients of variation were comparable, 0.6 for the team and 0.5 for the customers, but the team added more labels (50% vs. 18% of the primary needs were labels). Qualitatively, the team consensus chart structured the needs to reflect an engineering view while the customers sorted the needs to reflect product use. After seeing the customer-consensus chart, the team accepted it as a better structural representation. The resulting change in

[18] As is to be expected in real applications, the product-development team did make some adjustments in the raw output from the cluster analysis. These adjustments were based on their experience in the category.

organizational emphasis led to a number of fundamental changes in product development.

The customer-sort hierarchies have been applied over fifteen times by one supplier[19]. That supplier reports that in every application the product-development team accepted the customer-sort data as a better representation of the customer's voice and that, in some cases, the customer-sort structure changed dramatically the philosophy of the product-development effort. Some of these examples are given in the vignettes at the end of this paper.

Team Buy-in

One argument that has been advanced in favor of the team-based consensus charts is that they result in greater team buy-in to the hierarchical structure. Recent applications of customer-sort and customer-consensus structures have addressed this issue by having the team complete the customers' task in parallel with the customers. As the team sorts the cards they begin to ask themselves: "I sort the cards like this, but how would the customer sort the cards?" Indeed, while the customer instructions state that there is no right or wrong answer, the team begins to realize that for them there is a right answer -- how the customer sorts the cards.

In the end, the QFD philosophy of focusing on the customer and the scientific evidence that products are more successful if marketing input is understood by engineering and R&D, both suggest that the customer's perspective on the structure of customer needs should be given serious consideration. Note also that while we focus on the customer hierarchy for the customer's voice, the design attributes (engineering inputs to the House of Quality) can be (and often are) structured as the product is built. The relationship matrix (figure 1) provides the necessary link.

Summary

While the customer-sort analyses have not enjoyed the popularity of group-consensus charts, we feel that they deserve serious consideration for developing the hierarchical structure of customer needs that is used in QFD. We now address methods to measure or estimate importances for the primary, secondary, and tertiary needs.

MEASURING OR ESTIMATING IMPORTANCES

The next step in QFD's voice of the customer is to establish priorities for the customer needs in the form of importance weights. These priorities aid in allocating engineering resources and guide the team when it is forced to make tradeoffs among needs. For example, if a product-development team increases the thickness of the insulation in a food-carrying device, then they are likely to improve satisfaction relative to the primary need of *maintains food temperatures*

---

[19] Private communication with Robert Klein of Applied Marketing Science, Inc.

while degrading *carries many things*. Naturally, we prefer engineering strategies that stretch the frontier and improve satisfaction relative to both primary needs (such as changing the insulting material to obtain more insulting power per inch), but at times tradeoffs must be made and priorities set.

We begin with a brief review of the literature.

Previous Literature

The goal in QFD is to obtain importances that relate to customer *satisfaction*. This goal is similar to obtaining importances for *attitude*, *preference*, or *utility*. For each of these dependent measures, researchers have tested a variety of scales where customers are asked to state their importances directly (self-explicated importances) and a variety of techniques where the analyst attempts to infer statistically the importances (revealed importances) by relating customer ratings of products to measures of the dependent measure (attitude, preference, or utility). Extensive reviews have been published in attitude theory (Wilkie and Pessemier 1973), information integration (Lynch 1985), concept development (Shocker and Srinivasan 1979), conjoint analysis (Green and Srinivasan 1978, 1990), behavioral decision theory (Huber 1974), and the analytic hierarchy process (Wind and Saaty 1980).

While a few applications in hybrid conjoint analysis have dealt with large numbers of attributes[20], e.g., Wind, et. al. (1989) use 50 product features, the norms in these academic literatures are for far fewer attributes than the 200-300 customer needs that are typical in QFD. For example, 2-9 attributes are typical in multi-attribute attitude studies (Wilkie and Pessemier 1973), 5-7 are typical in hybrid conjoint analysis (Green 1984, pp. 165-167), 20-30 are typical in concept development (Shocker and Srinivasan 1979, p. 170), and 6 were used by Hoepfl and Huber (1970) in decision theory. Large numbers of customer needs mean that we do not have the luxury of estimating non-linear transforms (Lynch 1985) for each of the customer needs and that we must use a relatively parsimonious procedure -- usually a linear model.

There have been some explicit comparisons of the abilities of different techniques to predict the dependent measure. However, each of the comparative studies was based on between four and sixteen attributes. For example, Hoepfl and Huber (1970) report similar fits with direct measures and linear regression; Lehmann (1971) reports agreement between rank-order and 6-point bipolar scales; Schendel, Wilkie, and McCann (1971) report agreements between yes/no, rank-order, 6-point, and 100-point constant-sum scales; Hauser and Urban (1979) report comparable fits among the revealed techniques of regression and logit and a self-explicated "utility assessment" procedure; and Hauser and Koppelman (1977) report comparable fits between regression and logit, but slightly better fits with self-explicated scales. However, Hauser and Koppelman (1977) caution that an equal-weighting scheme fit as well as the self-explicated scales. See also Einhorn and Hogarth (1975). Green (1974, p. 166) and Akaah and

[20] We use the word attribute to refer to the independent measures reported in the literature. In some cases they are customer needs, but in others they are product features or components of attitude.

Korgaonkar (1983) report that hybrid conjoint analysis, which combines self-explicated and revealed techniques, can predict holdout profiles better than pure conjoint or self-explicated techniques. However, Cattin, Hermet and Pioche (1982) find the opposite to be true.

The literature cautions us that many techniques may provide comparable fits to a dependent measure, that it may be difficult to distinguish the comparative accuracy of self-explicated and revealed measures based on the fit to a dependent measure, and that we should consider seriously a model of equal weights as a basis for the comparison of fits. However, it is an open question whether these hypotheses apply to large number of needs. For example, a large number of needs usually means severe collinearity which makes accurate statistical estimates difficult to obtain. This collinearity also suggests that equal-weighting schemes are likely to provide good fits to the dependent measure. (Indeed the hierarchical structure suggests inherent collinearity at any level below primary needs.) On the other hand, respondent fatigue may be a factor for the self-explicated measures.

To the best of our knowledge, there have been no published comparisons of techniques within the QFD framework of primary, secondary, and tertiary needs, no published comparisons of techniques for large numbers (200-300) of customer needs, and no published comparisons where products or product concepts were developed based on the customer needs and then customer reactions to the products (product concepts) were compared to the measured or estimated importances[21].

We now report some new data collected within the QFD framework that attempts to address four questions that we have heard from industry: (1) Do survey measures of importances have any relation to customer preferences among products designed based on customer needs?, (2) What is the best survey measure?, (3) Can we avoid data collection for importances by using frequency of mention in the qualitative research as a surrogate for importance?, and (4) Are revealed techniques (with satisfaction as the dependent measure) superior to survey measures?

Do customers prefer product concepts that emphasize the fulfillment of "important" customer needs?

The following analysis is based on data collected by an unnamed consumer products firm.

The data. The consumer-products firm measured or estimated customer's importances for 198 customer needs using four different methods:

- 9-point *Direct-rating* scale in which customers answered for each need "How important is it or would it be if: ...?".
- *Constant-sum* scale in which customers allocated 100 points among the seven primary

[21] The have been conjoint comparisons where preferences for new profiles of features were compared to predictions (Green and Srinivasan 1990).

needs, then allocated 100 points to each set of secondary needs within each primary need group, and finally allocated 100 points among each set of tertiary needs within each secondary need group[22].

- *Anchored* scale in which customers allocated 10 points to the most important primary need and up to 10 points to the other six primary needs. Similarly up to 10 points were allocated to secondary needs corresponding to each primary need and to tertiary needs corresponding to each secondary need.

- *"Revealed" satisfaction* in which customers rated overall satisfaction with their current product on a 9-point satisfaction scale and evaluated their current product on a 6-point Likert scale for each customer need. Satisfaction was regressed on evaluations to reveal the importances[23].

Questionnaires were mailed to 5600 randomly selected consumers (1400 for each method). Response rates were very good (75-78%). (All recipients of the questionnaires were given a $5 incentive. Those that responded in a week were entered in a lottery for $100.) In addition, the constant-sum questionnaire was mailed to an additional 1400 consumers from a national panel. The response rate for that sample was 90%. The rank-order correlation of the importances as measured by the random sample and the panel sample was 0.995.

Customer reactions to product concepts. To test whether the importances made sense for setting priorities among product-development programs, the professional product-development team in the consumer-products company created seven product concepts. Each concept was created to emphasize one of the primary customer needs while stressing that the other six customer needs would not be any better or worse than existing products. The concepts went through two pre-tests with actual consumers and were modified until the firm felt that they did indeed "stretch" the consumer needs. (The actual concept statements are proprietary.) Consumers were asked to express their interest (9-point scale) and preference (rank order) for the concepts. Table 3 indicates that consumers' interest and preference is highly correlated with the self-stated measures of primary needs. (Revealed estimates are discussed below.)

We asked the product-development team at the consumer product company to judge the face validity of the importance measures. They felt that the measured importances (direct, constant-sum, and anchored) corresponded to their beliefs about the category -- beliefs based on experience and a large number of other market studies. Based on the data in table 3 and the face validity of the measures, the consumer-products company felt that self-explicated measures provided accurate importance measures for the QFD process. See Hauser (1991) for details.

---

[22] The importance of a tertiary need reflects the cascaded allocations to the primary and secondary needs. Cascading is used also for the anchored scale.

[23] We also tried a 100-point satisfaction measure and a 9-point like/dislike scale. Results were similar methodologically. Due to collinearity we only estimate primary- and secondary-need regressions. We did not test the analytic hierarchy process (AHP) because it would have required 602 pair-wise judgments by each respondent.

Table 3 Comparison of Interest and Preference with Importances

	Interest	Preference	Direct	Anchored	Cons.-sum	Revealed
Primary Need A	2	1	1	1	1	1
Primary Need B	1	2	3	2	2	7
Primary Need C	4	4	4	4	4	4
Primary Need D	6	6	6	5	5	5
Primary Need E	5	5	5	6	6	3
Primary Need F	7	7	7	7	7	2
Primary Need G	3	3	2	3	3	6
RANK Correlation with		Interest	0.89	0.93	0.93	-0.36
		Preference	0.96	0.96	0.96	-0.14

Note that, by definition, a null model of equal importances for the primary needs would have a 0.00 correlation with interest and preference for the product concepts. Thus, in this case, the importance measures do significantly better than the null model. Based on the data in table 3 the product development team was able to focus on primary needs A and B. Had they relied on the null model they would have missed this opportunity entirely.

Table 4 Correlations Between Ranks of Mean Importances

	Direct	Anchored
PRIMARY NEEDS		
Anchored	0.96	
Constant-sum	0.96	1.00
SECONDARY NEEDS		
Anchored	0.78	
Constant-sum	0.67	0.94
TERTIARY NEEDS		
Anchored	0.84	
Constant-sum	0.71	0.89

Which survey measure is best?

Table 3 suggests that all three self-explicated survey measures are comparable in their ability to predict how customers will react to product concepts. Furthermore the three measures give similar rank-order results. As table 4 indicates the measures are also similar at the secondary and tertiary levels, particularly between the constant-sum and the anchored scales. (Tables 3 and 4 report rank correlations. We get similar results for Pearson correlations.) We have also completed comparisons for two other product categories, the portable food-carrying device described earlier (Griffin 1989) and a proprietary application to a high-cost durable product[24]. In both cases there was rank-order agreement between the survey measures of importance. Qualitatively we prefer the anchored scale[25], but the scientific data to date suggest that any of the three scales could be used to measure importances.

Is frequency of mention a surrogate for importance?

It is a reasonable hypothesis that customers will mention most those needs that are most important. If this were true, then we could save time and money by using frequency of mention as a surrogate for importance. To test this hypothesis we measured importances for the primary, secondary, and tertiary customer needs identified for the portable food-carrying device. We used a nine-point direct-rating importance scale. We then reanalyzed data as described in figures 3, 4, and 5, but for only the most important needs. The results are plotted in figure 7, where, for comparison, we have normalized the data so that 30 customers equals 100%. Figure 7 suggests that important needs are no more likely to be mentioned by a customer than needs in general. (The distributions do not differ at the *0.05* level by a Kolmogorov-Smirnov goodness-of-fit test.) Regrettably, frequency of mention does not appear to be a good surrogate for importance.

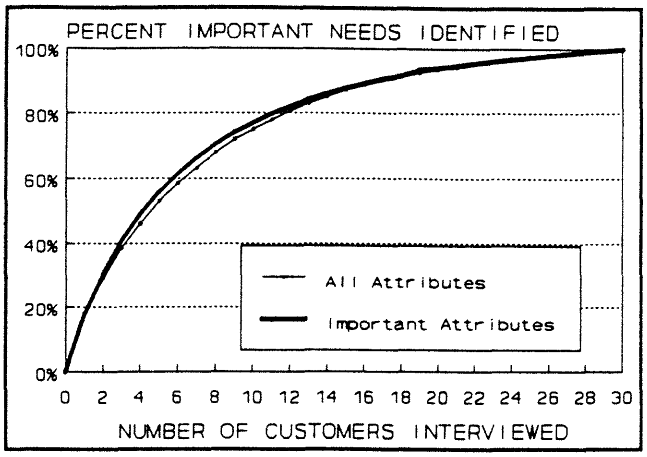

Figure 7. Important Needs vs. Total Needs (from Griffin and Hauser 1991b)

Are revealed techniques (based on satisfaction) superior to survey measures?

Econometricians advocate revealed preference measures where the importance weights of attributes are derived statistically (Manski and McFadden 1981, Ben-Akiva and Lerman 1985). For the consumer good we measured customer's perceptions of their chosen product and regressed those perceptions on customer's satisfaction with that product. The results are reported in the last column of table 3.

As reported, neither interest nor preference for product concepts correlates with the revealed importances. This poor predictive ability may be due to the collinearity among primary needs (71% of the correlations are above 0.20). This collinearity is even more severe at the

---

[24] Griffin's study was a pretest of 133 students for the 230 customer needs discussed earlier. She found direct, constant-sum, and anchored measures to be similar. The proprietary study compared direct ratings and constant-sum measures for almost 150 customer needs. The sample size was 350 customers.

[25] One must be cautious in using either anchored or the constant-sum scale. In both of these scales the rated importance of the primary need is cascaded down as a multiplying factor for the corresponding secondary and tertiary needs. If the primary need is poorly worded, then any measurement error affects *all* corresponding secondary and tertiary needs. For this reason, the consumer-goods company prefers the direct measures.

secondary and tertiary levels. The results may also be due to the fact that satisfaction for the chosen product may not be the best dependent measure. We address this result in the next section.

We have also attempted to estimate revealed importances for a high-cost durable product and for the portable food-carrying devices (Griffin 1989). In both cases the collinearity was severe. It was not uncommon that less that 20% of the importances were "revealed" to be significant and several had negative signs. In none of the applications did the revealed estimates have high face validity.

While we can not rule out revealed satisfaction techniques for the large numbers of customer needs in QFD, we do feel that collinearity poses formidable barriers to such estimation.

Summary

Based on the data examined to date we feel that survey measures of importance can predict how customers will react to product concepts. However, we have not yet identified a single "best" measure. On the other hand, frequency of mention does not appear to be a good surrogate for importance and revealed techniques suffer from collinearity in customer perceptions.

CUSTOMER SATISFACTION AS A GOAL

We based the revealed estimates on customer satisfaction because industry accepts customer satisfaction as the goal of QFD. This philosophy is based on the belief that in the short term the firm can affect sales by price, promotion, advertising, and other variables. While these are important decisions, QFD focuses on designing a product for long-term profitability. Total-quality advocates believe that, in the long-run, satisfied customers are an asset of the firm. Future short-run strategies can be adjusted to draw on the asset of satisfied customers.

Self-selection Bias

Given the academic interest in revealed importances, the poor showing of the revealed technique is sobering. While this may be due entirely to collinearity among customer perceptions, we (and the consumer-product firm) suspected that there was something more fundamental about the measure of satisfaction. For example, the firm's leading brand had been number one in the category for over twenty years, but its average satisfaction score was below that of many other brands. The brand with the highest satisfaction score was a small niche brand. (This phenomenon was also identified in Swedish data. See Fornell 1991.)

Recall that the satisfaction measure asks customers to rate the brand they have chosen[26]. We call such a measure a monadic measure. At minimum this measure contains a self-selection bias -- presumably customers prefer most (price and promotion considered) the brand they chose. Indeed, a niche brand may satisfy only a few customers, but it may satisfy them quite well. On the other hand, a market-share leader might satisfy its customers more than other brands, but, because its customers are diverse, its satisfaction score (for leading-brand customers) might be lower than the niche brand's score (for niche-brand customers). Thus, while satisfaction is a different construct than market share, a low correlation between measured satisfaction and market share would suggest the presence of such a self-selection bias.

One initial test of this hypothesis is presented in table 5. This table compares rank-order primary-brand share with monadic satisfaction and with relative-satisfaction from ongoing tracking data at the consumer-products company[27]. For the ten brands for which data from both studies is available, monadic satisfaction did not correlate with primary-brand share. However, the relative satisfaction measure did correlate with primary-brand share. The correlation was marginal (*0.15* level) among consumers who have used the brand in the last three months, but highly significant (*0.01* level) among consumers who have heard of the brand.

Table 5 Comparison of Monadic Satisfaction, Relative Satisfaction, and Primary-Brand Share

BRAND	PRIMARY-BRAND SHARE	MONADIC SATISFACTION MEASURE	RELATIVE SATISFACTION (BRAND USER)	RELATIVE SATISFACTION (HEARD OF BRAND)
Q	1	6	1	1
R	2	3	5	3
S	3	8	4	2
T	4	2	8	6
U	5	9	7	7
V	6	1	9	5
W	7	4	3	8
X	8	7	2	4
Y	9	10	6	9
Z	10	5	10	10
Rank Correlation		0.20	0.39	0.83
t-statistic		0.58	1.21	4.21

Table 5 highlights the dilemma in choosing an appropriate satisfaction measure. When evaluating a product program we prefer to base satisfaction on customers who have used the brand and, perhaps, not include those who have only heard of the brand. However, the used-brand sample is subject to the same criticism as the monadic satisfaction measure -- it confounds people and

[26] This is the most common measure of satisfaction that we have seen used in industry. There is no self-selection bias in the academic studies in which all customers evaluate hypothetical products (Oliver and DeSarbo 1988, Tse and Wilton 1988).

[27] Rank-order data preserves confidentiality better. The qualitative insights were similar for the interval-scaled data. Primary-brand share is the share of consumers who use the brand as their primary product. It is similar to, but not identical to, a market-share measure. The relative measures are relative in the sense of customers, all customers who have heard of [used] the brand rate it, and in the sense of brands, customers rate the brand relative to all brands that they have heard of (used). The ten brands reported comprise approximately 80% of the market.

products (albeit to a lesser degree). The used-brand sample includes only people who have evaluated a brand and who, at least once, believed that it would meet their needs. The heard-of-brand sample includes more consumers including many who have evaluated the brand and rejected it. It should not surprise us that the latter measure is more like a market-share measure.

The self-selection bias with respect to the commonly-used satisfaction measure is extremely important to designed-in-quality programs. Many American corporations are using measures of customer satisfaction as part of employee rewards and bonuses. For example, GTE and Montgomery Ward both tie management compensation to customer-satisfaction and quality measurements (Phillips, et. al. 1990). If our hypothesis about the satisfaction measure holds up, then there is a real danger that these corporate programs based on monadic satisfaction may be sending the wrong signals to product design. For example, suppose that a product-development group is rewarded only on monadic satisfaction. Then they might choose to design a product that gets extremely high satisfaction scores from a small niche of the target customers. They might avoid designs that capture a large market share of diverse customers. On the other hand, a relative measure of satisfaction would give better incentives. The niche product might satisfy its niche, but not the large set of diverse customers. The large-share product might satisfy many more customers (relative to alternative products). Naturally, satisfaction-based incentive systems raise many complex issues beyond the scope of this paper. However, the self-selection bias inherent in the most common measures of customer satisfaction deserves further research.

Implications for Technique Comparisons

Self-selection bias also has implications for academic research comparing different measures and estimates of importance weights[28]. Many of the comparisons in the literature (including some that we have published) are based on the correlation of a "preference index" with measured preference. When the preference, attitude, or utility measure is monadic, such correlations may confound the self-selection bias with differences in the predictive ability of the importances. To test this with our data on the consumer product we correlate a "satisfaction index" with measured satisfaction.

We create a satisfaction index based on the sum of importances times the evaluations of customer needs[29]. As a null model we use an "equal-weights" model which simply sums the evaluations (Einhorn and Hogarth 1975). However, there are two technical issues to consider in the comparison. First, if we are to treat the measured importances as ratio-scales, we must estimate a translation factor. Second, because the importances and the evaluations were measured in different questionnaires (in this data), we can not match consumers one-to-one. The

---

[28] This discussion applies to those studies that measure or estimate importances for groups of customers. It is mute on the many conjoint studies in which an individual's importance weights are used to predict preference among holdout profiles, that is, where the experimenter chooses the product profiles. It is also mute on studies where subjects evaluate products or product concepts chosen by the experimenter. It does apply to studies where subjects evaluate only those products that they would consider seriously.

[29] If W_k is the importance of the *k-th* primary need and E_k is the evaluation of a product on the *k-th* primary need, then a primary-level satisfaction index is $SI = \sum_k W_k E_k$. Secondary- and tertiary-level indices are computed in the same way. An equal-weights model is given by $SE = \sum_k E_k$.

best that we can do is to use average importances with each consumer's evaluations.

Table 6 Correlation of Satisfaction Index and 9-point Monadic Satisfaction

MODEL	Primary Needs	Secondary Needs	Tertiary Needs
Equal Weights	0.37	0.42	0.39
Satisfaction Index			
Self-stated	0.37	0.41	0.39
Anchored	0.37	0.41	0.39
Constant-sum	0.37	0.40	0.39
Revealed Preference	0.39	0.43	–

We address the ratio-scaled issue by recognizing that we can uncover an unknown translation factor for the importances by regressing satisfaction on the satisfaction index and an equal-weight index. The ratio of the regression weights is the scale factor. The t-statistic associated with the satisfaction index tests whether the importances explain satisfaction significantly better than the null model of equal importances[30]. For this consumer-product data set, the satisfaction indices had reasonable correlations with measured satisfaction, but the correlations were not statistically superior to an equal-weights model. Table 6 reports the correlations for a holdout sample (a random thirty percent of the sample).

As feared, although the measured importances do appear to predict customer interest in and preference for product concepts (table 3), the correlation between a satisfaction index and measured satisfaction can not distinguish the predictive ability of the measured importances from that of equal weights. This false rejection of the importance measures may be due to collinearity among customer perceptions or it may be due to the self-selection bias, but, at minimum, table 6 does caution us about the conclusions that can be drawn from such correlative measures.

Summary

Customer satisfaction is often cited as the goal of QFD. However, our data caution firms that monadic measures of satisfaction lead to counterproductive incentives when evaluating products and product programs. Our data also caution academic researchers that self-selection biases and/or collinearity can result in false negatives (relative to equal weights) when evaluating alternative measures of importance.

---

[30] If W_i' are the raw importances, then the rescaled importances, W_i, are given by $W_i = W_i' + constant$. If SI' corresponds to the raw importances and SI to the rescaled importances, then substitution gives $SI = SI' + constant*SE$.

APPLICATION VIGNETTES

The preceding sections have discussed some of the technical issues in identifying, structuring, and providing priorities for the voice of the customer. This section describes briefly some of the successful applications of QFD and the voice of the customer[31].

A **manufacturer of medical diagnostic equipment** was faced with a well-financed competitor whose new product was being sold at half the price. However, the competitive product did not have a feature that the manufacturer considered important. The voice of the customer suggested that the product could be redesigned totally to satisfy important customer needs. A modular design would enable customers in different segments to pick and choose those features that best satisfied their needs. Within a year, a new product was developed based on the voice of the customer. The basic module is priced below competition, but a fully featured product commands a high price. Sales are expected to be five times higher than last year.

A **consumer stationary-products manufacturer** observed a year-by-year decline in the sales of a key replacement component in their line. They did not know whether the decline was due to the perceived quality of the component or the appeal of the product itself. QFD identified the important customer needs which, in turn, identified key design attributes. Laboratory measures of the design attributes indicated that a modest improvement in the component's quality would reverse the decline in sales.

A **manufacturer of tools for the construction industry** was considering a new technology that promised to save significant manpower at construction sites. These tools would have applications in a variety of industries and market segments. Although the basic technology existed, the firm would need to invest in significant design efforts to create an actual product. The voice of the customer identified key customer needs that were important to a target segment and which would distinguish the new product if it were developed based on the technology. The manufacturer focused the product-development program to this segment and these needs.

A **financial institution** used the voice of the customer to evaluate their formal customer communications program. They identified eight important customer needs that were not being addressed effectively by current communications. They revised the communications programs and achieved increased sales.

The information-systems group in an **insurance company** had a substantial backlog of information-systems requests from other functions, each of which was labeled as high priority. By linking the benefits provided by the projects to important customer needs, the company was able to eliminate some projects, identify new projects which where important to the customer, and establish priorities for the remaining projects.

---

[31] These applications were described to us by Robert Klein of Applied Marketing Science, Inc. They are representative of the applications we have observed and, we believe, comparable to many other applications that we have not observed directly.

A **manufacturer of a lightweight chemical mixing device** had positioned its product on "greater portability." However, the voice of the customer suggested that "ease of use" and "accuracy" were of much greater importance. Through QFD they modified the design attributes of training requirements and the predictability of the mixing result. They improved these attributes via a modification in the design of the mixing device and through a new sales package. Sales increased.

An **entertainment provider** with 350 separate locations discovered that customers viewed one specific area of its operation as very important. However, customers did not perceive that the entertainment provider satisfied their needs in that area. A crash program was undertaken to identify and implement programs that would impact that area and beat out competition.

A start-up **manufacturer of a new surgical instrument** had contracted with a design firm to develop prototypes for a new scalpel. The choice of the final design would depend upon accommodating the constraints of the technology in a package that fit the way surgeons operated. The voice of the customer reoriented the development and design effort and significantly improved product acceptability.

A **manufacturer of office equipment** was designing the next generation of a product that held a dominant market position. A competitor, using digital technology, was making rapid inroads in a related market segment. The voice of the customer suggested that the benefits of the digital technology were important for the related market segment. However, they were of only minor importance for the manufacturer's market segment. As a result the manufacturer refocused its development effort to more important customer needs and avoided the expense and delay of moving to the digital technology.

In each of these vignettes, the key lesson is not only that the data were timely, but that the product-development team was able to use the data successfully to make changes that resulted in either improved sales or profitability. Voice-of-the-customer analyses meet the needs of its customer, the product-development team.

DISCUSSION AND SUMMARY

Quality Function Deployment (QFD) has been adopted widely by US and Japanese product-development teams in an attempt to design products that are more profitable. QFD promises decreased product-development costs, decreased product-development time, and improved customer satisfaction. There is some evidence, albeit preliminary, that these goals are being achieved.

QFD begins with the voice of the customer -- a list of 200-300 customer needs in a hierarchical tree of primary strategic needs, secondary tactical needs, and tertiary operational needs. The voice of the customer is the list of customer needs, the hierarchical tree, and a set of importances. In this paper we have described common techniques used to identify, structure, and provide priorities for the customer needs. We have summarized some of the academic research and we have presented new data to compare and/or evaluate the techniques. Throughout the paper we have adopted the incremental-improvement philosophy of total quality management by building upon industry practice to make the voice of the customer a little more effective.

Our data suggest interviews with 20-30 customers should identify 90% or more of the customer needs in a relatively homogeneous customer segment. Both one-on-one experiential interviews and focus groups seem to effective at identifying needs, but the group synergies expected from focus groups do not seem to be present. Multiple analysts (4-6) should analyze the transcripts.

Group-consensus charts are the most popular method for obtaining a hierarchical structure. Our data suggest that different structures are obtained by analyzing data in which customers are asked to sort the customer needs into similar piles. The customer-sort hierarchies seem to group the needs to reflect how the customer uses the product while team-consensus charts group the needs to reflect how the firm builds the product. If we are to believe the scientific data on the advantages of the product-development team understanding the customer, then customer-sort hierarchies should lead to better products and services. At minimum, they deserve more attention from industry.

The measurement or estimation of importances (relative to preference, attitude, or utility) has a rich history in the academic literature. We build upon that literature and present some new data on the comparison of techniques to obtain importances relative to satisfaction. Our data suggest that if product concepts are created based on measured importances, then customers prefer and are interested in those products which stress important customer needs. However, for our data, estimated importances (regressing perceptions on satisfaction) do not seem to correlate with preference or interest. We suspect that this is due to the collinearity in the data (inherent in QFD) and/or the self-selection bias of the dependent measure, monadic satisfaction. Regrettably, frequency of mention does not appear to be a surrogate for importance.

We discuss measurement issues relative to the stated goal of QFD, customer satisfaction. Our data suggest that a self-selection bias might be present in standard customer-satisfaction data collected by corporations. This bias causes the relative satisfaction of low-share brands to be overstated and the relative satisfaction of high-share brands to be understated. Furthermore, this self-selection bias, when present, might lead to false negatives when measured or estimated importances are compared to the null model of equal importance weights.

Finally, the application vignettes suggest that the voice of the customer is used by product development teams and that it does affect product strategies, sales, and profit.

We feel we have made substantial incremental progress, but many challenges remain. There are many techniques in the literature which have not been compared to common practice in QFD. Perhaps, one of these techniques will prove superior. While data on two applications suggest that 20-30 customers per segment are sufficient, we do not know how this varies with the characteristics of product categories. Perhaps our focus-group result is a false negative. Satisfaction measurement is a complex issue. We have only indicated one potential bias. In these and in many other ways we hope that other researchers build upon the data presented in this paper.

We have also seen new research problems in industry. For example, when a voice of the customer is completed for many segments and for many related product lines there are many common needs. Furthermore, a representative structure of the customer needs allows for commonalities in product design. Industry is concerned with balancing the expense of multiple voice-of-the-customer studies, for each segment and for each product category, with the opportunity cost of doing a common voice-of-the-customer which has the same structure and mostly the same customer needs, but different importances for different segments. Finally, the search for the breakthrough exciting customer needs has received much attention in industry. Perhaps new elicitation techniques, such as leading-edge user studies (von Hippel 1986), can be developed to identify these exciting needs.

REFERENCES

Akaah, Ismael P. and Pradeep K. Korgaonkar (1983), "An Empirical Comparison of the Predictive Validity of Compositional, Decompositional, and Hybrid Multiattribute Preference Models," *Journal of Marketing Research*, 20, (May), 187-197.

Akao, Yoji (1987), *Quality Deployment: A Series of Articles*, (Lawrence, MA: G.O.A.L., Inc.) Translated by Glen Mazur.

American Supplier Institute (1987), *Quality Function Deployment: A Collection of Presentations and QFD Case Studies*, (Dearborn, MI: American Supplier Institute, Inc.)

Ben-Akiva, Moshe and Steven R. Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, (The MIT Press: Cambridge, MA).

Brunswick, E. (1952), *The Conceptual Framework of Psychology*, (Chicago, IL: University of Chicago Press).

Calder, Bobby J. (1979), "Focus Groups and the Nature of Qualitative Marketing Research," *Journal of Marketing Research*, 14, (August), 353-364.

Cattin, Philippe, Gerard Hermet, and Alain Pioche (1982), "Alternative Hybrid Models for Conjoint Analysis: Some Empirical Results," in *Analytical Approaches to Product and Market Planning: The Second Conference*, (Marketing Science Institute: Cambridge, MA), (October), 142-152.

Clausing, Don, (1986), "QFD Phase II: Parts Deployment," American Suppliers Institute Publication.

Cooper, Robert (1983), "The New Product Process: An Empirically-Based Classification Scheme," *R&D Management*, 13(1), 1-13.

Cooper, Robert G. (1984a), "New Product Strategies: What Distinguishes the Top Performers?," *Journal of Product Innovation Management*, 2, 151-164.

Cooper, Robert G. (1984b), "How New Product Strategies Impact on Performance," *Journal of Product Innovation Management*, 2, 5-18.

Cooper, Robert G. and Ulricke de Brentani (1991), "New Industrial Financial Services: What Distinguishes the Winners," *Journal of Product Innovation Management*, 8, 75-90.

Cooper, Robert, and Elko Kleinschmidt (1986), "An Investigation into the New Product Process: Steps, Deficiencies, and Impact," *Journal of Product Innovation Management*, 3, 71-85.

Dawkins, Brian (1991), "Siobhan's Problem: The Coupon Collector Revisited," *The American Statistician*, 45, 1, (February), 76-82.

de Brentani, Ulrike (1989), "Success and Failure in New Industrial Services," *Journal of Product Innovation Management*, 6, 239-58.

Dougherty, Deborah J. (1987), "New Products in Old Organizations: The Myth of The Better Mousetrap in Search of the Beaten Path," Unpublished Ph.D. Thesis, Sloan School of Management, M.I.T., Cambridge, MA 02139 (June).

Efron, Bradley and Ronald Thisted (1976), "Estimating the Number of Unseen Species: How many words did

Shakespeare know?," *Biometrika*, 63, 3, 435-447.

Einhorn, Hillel J. and Robin M. Hogarth (1975), "Unit-weighting Schemes for Decision Making," *Organizational Behavior and Human Performance*, 13, 171-192.

Eureka, William E., (1987), "Introduction to Quality Function Deployment," Section III of Quality Function Deployment: A Collection of Presentations and QFD Case Studies, American Suppliers Institute Publication, January.

Fornell, Claes (1991), "A National Customer-Satisfaction Barometer: The Swedish Experience," Working Paper, The Office for Customer Satisfaction Research, University of Michigan, Ann Arbor, MI 48109.

General Accounting Office, U. S. (1991), "Management Practices: U.S. Companies Improve Performance Through Quality Efforts," GAO/NSIAD-91-190, (May), Gaithersburg, MD 20877.

Green, Paul E. (1984), "Hybrid Models of Conjoint Analysis: An Expository Review," *Journal of Marketing Research*, 21,2, (May), 155-169.

Green, Paul E., Frank J. Carmone, and Leo B. Fox (1969), "Television Programme Similarities: An Application of Subjective Clustering," *Journal of the Market Research Society*, 11, 1, 70-90.

Green, Paul E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research*, 5, (September) 103-123.

Green, Paul E. and V. Srinivasan (1990), "Conjoint Analysis in Marketing Research: New Developments and Directions," *Journal of Marketing*, 54, 4, (October), 3-19.

Green, Paul E., Donald S. Tull, and Gerald Albaum (1988), *Research for Marketing Decisions, 5th Edition*, (Prentice-Hall, Inc.: Englewood Cliffs, NJ).

Greene, Jerome D. (1982), *Consumer Behavior Models for Non-statisticians*, (Praeger Scientific: New York, NY).

Goodman, L. A. and W. H. Kruskal (1954), "Measures of Association for Cross-Classification," *Journal of the American Statistical Association*, 49, 732-764.

Griffin, Abbie (1989), "Functionally Integrated New product Development," Unpublished Ph.D. Thesis, Sloan School of Management, M.I.T., Cambridge, MA 02139 (June).

Griffin, Abbie (1991), "Evaluating Development Processes: QFD as an Example," University of Chicago Working Paper.

Griffin, Abbie, and John R. Hauser (1991a), "Patterns of Communication Among Marketing, Engineering and Manufacturing -- A Comparison Between Two New Product Teams," accepted for publication in *Management Science*.

Griffin, Abbie and John R. Hauser (1991b), "The Marketing and R&D Interface," forthcoming in *Handbook: MS/OR in Marketing*, Gary L. Lilien and Jehoshua Eliashberg, Eds., (Elsevier Science Publishers: Amsterdam, The Netherlands).

Gupta, Ashok K., S.P. Raj, and David Wilemon (1985b), "R&D and Marketing Dialogue in High-Tech Firms," *Industrial marketing Management*, 14, 289-300.

Gutman, Jonathan (1982), "A Means-End Chain Model Based on Consumer Categorization Processes," *Journal* *of Marketing*, 46, (Spring), 60-72.

Hauser, John R. (1978), "Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information-Theoretic Approach," *Operations Research*, 26, 3, (May-June), 406-421.

Hauser, John R. (1991),"Comparison of Importance Measurement Methodologies and their Relationship to Consumer Satisfaction," M.I.T. Working Paper, Sloan School of Management, Cambridge, MA, January.

Hauser, John R. and Don P. Clausing (1988), "The House of Quality," *Harvard Business Review*, 66, 3, (May-June), 63-73.

Hauser, John R. and Frank S. Koppelman (1977), "Effective Marketing Research: An Empirical Comparison of Techniques to Model Consumers' Perceptions and Preferences," in *Analytic Approaches to Product and Market Planning*, Allan D. Shocker, ed., (Marketing Science Institute: Cambridge, MA).

Hauser, John R. and Glen L. Urban (1977), "A Normative Methodology for Modeling Consumer Response to Innovation," *Operations Research*, 25, 4, (July-August), 579-619.

Hauser, John R. and Glen L. Urban (1979), "Assessment of Attribute Importances and Consumer Utility Functions," *Journal of Consumer Research*, 5, (March), 251-62.

Hise, Richard T., Larry O'Neal, A. Parasuraman and James U. McNeal (1990), "Marketing/R&D Interaction in New Product Development: Implications for New Product Success Rates," *Journal of Product Innovation Management*, 7, 2, (June), 142-155.

Hoepfl, Robert T. and George P. Huber (1970), "A Study of Self-explicated Utility Models," *Behavioral Science*, 15, 408-414.

Huber, George P. (1974), "Multi-attribute Utility Models: A Review of Field and Field-like Studies," *Management Science*, 20, 10, (June), 1393-1402.

Imai, Masaaki (1986), *Kaizen: The Key to Japan's Competitive Success*, (McGraw-Hill Publishing Company: New York).

Juran, Joseph M. (1989), *Juran on Leadership for Quality: An Executive Handbook*, (The Free Press: New York), 14-19.

Kelly, George A. (1955), *The Psychology of Personal Constructs, vol. 1.*, (W. W. Norton: New York, NY).

King, Robert (1987), *Better Designs in Half the Time: Implementing Quality Function Deployment (QFD) in America*, (Lawrence, MA: G.O.A.L., Inc.)

Kogure, Masao and Yoji Akao (1983), "Quality Function Deployment and CWQC," *Quality Progress*, 16:10, (October), 25-29.

Lehmann, Donald R. (1971), "Television Show Preference: Application of a Choice Model," *Journal of Marketing Research*, 8, (February), 47-55.

Lehmann, Donald R. (1985), *Market Research and Analysis, 2nd Edition*, (Irwin: Homewood, IL).

Lillrand, Paul and Noriaki Kano (1989), *Continuous Improvement*, (University of Michigan Press: Ann Arbor, MI).

Lynch, John G., Jr. (1985), "Uniqueness Issues in the Decompositional Modeling of Multi-attribute Overall Evaluations: An Information Integration Perspective," *Journal of Marketing Research*, 22, (February), 1-19.

Manski, Charles F. and Daniel McFadden (1981), *Structural Analysis of Discrete Data with Econometric Applications*, (The MIT Press: Cambridge, MA).

McElroy, John, (1987), "For Whom are We Building Cars?," *Automotive Industries*, (June), 68-70.

Moenaert, Rudy K. and William E. Souder (1990), "An Information Transfer Model for Integrating Marketing and R&D Personnel in New Product Development Projects," *Journal of Product Innovation Management*, 7, 2, (June), 91-107.

Morgan, Anthony I. (1984), "Point of View: Magic Town Revisited (A Personal Perspective)," *Journal of Advertising Research*, 24, 4, (August/September), 49-51.

Morrison, Donald G. (1979), "An Individual Differences Pure Extinction Process," *Journal of Mathematical Psychology*, 19, 3, (June), 307-315.

Myers, James H. (1976), "Benefit Structure Analysis: A New Tool for Product Planning," *Journal of Marketing*, 40, (October), 23-76.

National Institute of Standards and Technology (NIST) (1991), "1991 Application Guidelines: Malcolm Baldrige National Quality Award," Department of Commerce, Gaithersburg, MD 20899.

Oliver, Richard L. and Wayne S. DeSarbo (1988), "Response Determinants in Satisfaction Judgments," *Journal of Consumer Research*, 14, (March), 495-507.

Pelz, Donald C., and F.M. Andrews (1976), *Scientists in Organizations*, Revised Edition, (Ann Arbor, MI: University of Michigan Press).

Pessemier, Edgar A. (1986), *Product Management: Strategy and Organization*, (Robert E. Krieger Publishing Company, Inc.: Malabar, FL).

Phillips, Stephen, Amy Dunkin, James Treece, and Keith H. Hammonds (1990), "King Customer: At Companies That Listen Hard and Respond Fast, Bottom Lines Thrive," *Business Week*, (March 12), 88-94.

Pinto, Mary Beth and Jeffrey K. Pinto (1990), "Project Team Communication and Cross-Functional Cooperation in New Program Development," *Journal of Product Innovation Management*, 7, 200-212.

Rao, Vithala and Ralph Katz (1971), "Alternative Multidimensional Scaling Methods for Large Stimulus Sets," *Journal of Marketing Research*, 8, (November), 488-494.

Reimann, Curt (1991), "The Baldrige Award Criteria Speak for Themselves," *Quality Progress*, 24, 5, (May), 43-44.

Reynolds, Thomas J. and Jonathan Gutman (1988), "Laddering Theory, Method, Analysis, and Interpretation," *Journal of Advertising Research*, 28, 1, 11-31.

Romesburg, H. Charles (1984). *Cluster Analysis for Researchers*, (Lifetime Learning Publications: Belmont, CA).

Ruiz, Dirk and Dipak Jain (1991), "Designing and Developing New Products: The Subproblem Decomposition

Approach," Working Paper, Kellogg Graduate School of Management, Northwestern University, Evanston, IL 60208.

Schendel, D. E., William L. Wilkie, and John M. McCann (1971), "An Experimental Investigation of Attribute Importance," *Proceedings: Second Annual Conference*, Association of Consumer Research, 404-416.

Shocker, Allan D. and V. Srinivasan (1979), "Multi-attribute Approaches for Product Concept Evaluation and Generation: A Critical Review," *Journal of Marketing Research*, 16, (May), 159-180.

Silver, Jonathan Alan and John Charles Thompson, Jr. (1991), "Understanding Customer Needs: A Systematic Approach to the 'Voice of the Customer,'" Master's Thesis, Sloan School of Management, M.I.T., Cambridge, MA 02139.

Souder, William E. (1978), "Effectiveness of Product Development Methods," *Industrial Marketing Management*, 7.

Souder, William E. (1987), "Managing the R&D/Marketing Interface," Chapter 10 in *Managing New Product Innovations*, Lexington Books, Lexington, MA.

Souder, William E. (1988), "Managing Relations Between R&D and Marketing in New Product Development Projects," *Journal of Product Innovation Management*, 5, 6-19.

Stratton, Brad (1991), "The Value of Implementing Quality," *Quality Progress*, 25, 7, (July), 70-71.

Sullivan, Lawrence P. (1986), "Quality Function Deployment," *Quality Progress*, 19, 6, (June), 39-50.

Sullivan, Lawrence P. (1987), "The Power of Taguchi Methods," *Quality Progress*, 12:6, (June), 76-79.

Tse, David K. and Peter C. Wilton (1988), "Models of Consumer Satisfaction Formation: An Extension," *Journal of Marketing Research*, 25, (May), 204-212.

Tybout, Alice M. and John R. Hauser (1981), "A Marketing Audit Using a Conceptual Model of Consumer Behavior: Application and Evaluation," *Journal of Marketing*, 45, (Summer), 82-101.

Urban, Glen L., and John R. Hauser (1980), *Design and Marketing of New Products*, (Englewood Cliffs, NJ: Prentice-Hall, Inc.)

von Hippel, Eric (1986), "Lead Users: A Source of Novel Product Concepts," *Management Science*, 32, 7, (July), 791-805.

Vorberg, Dirk and Rolf Ulrich (1987), "Random Search with Unequal Search Rates: Serial and Parallel Generalizations of McGill's Model," *Journal of Mathematical Psychology*, 31, 1-23.

Wilkie, William L. and Edgar A. Pessemier (1973), "Issues in Marketing's Use of Multi-attribute Attitude Models," *Journal of Marketing Research*, 10, (November), 428-441.

Wind, Yoram (1982), *Product Policy: Concepts, Methods, and Strategy*, (Addison-Wesley: Reading, MA).

Wind, Yoram, Paul E. Green, Douglas Shifflet, and Marsha Scarbrough (1989), "*Courtyard by Marriott:* Designing a Hotel facility with Consumer-Based Marketing Models," *Interfaces*, 19, (January-February), 25-47.

Wind, Yoram and Thomas L. Saaty (1980), "Marketing Applications of the Analytic Hierarchy Process," *Management Science*, 26, 7, (July), 641-658.