MARKETING SCIENCE
Vol. 12, No. 1, Winter 1993, 1–27 (original article)
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

http://dx.doi.org/10.1287/mksc.12.1.1

The Voice of the Customer

Abbie Griffin
University of Chicago

John R. Hauser
Massachusetts Institute of Technology

In recent years, many U.S. 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. This paper focuses on the "Voice-of-the-Customer" component of QFD, 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 or one-on-one interviews superior? 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 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 also address the question of whether frequency of mention can be used as a surrogate for importance. Finally, 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 a brief application to illustrate how a product-development team used the voice of the customer to create a successful new product.

Keywords: new product research; product policy; measurement

Introduction

Many leading US firms are focusing on total quality management techniques. For example, in 1991, 106 firms applied 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, Stratton 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). 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 U.S. in 1986 for initial applications at Ford and Xerox. By 1989 approximately two dozen U.S. firms had adopted QFD for some or all of their product and service development.[2]

[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.

[2] Among the U.S. and Japanese firms reporting QFD applications in 1989 were General Motors, Ford, Navistar, Toyota, Mazda, Mitsubishi, Procter & Gamble, Colgate, Campbell's Soup, Gillete, 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.

We estimate that in 1991 well over 100 firms use 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 assigned to it a priority which indicates its importance to the customer. Developing products based on the voice of the customer becomes a key criterion in total quality management. The first key concept in the Baldrige Award criteria is 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 cannot compare all the possible techniques for any given step in the customer input. Instead, based on experience over the past four years with over twenty-five U.S. corporations[3] and based on discussions with market research suppliers, we focus on those techniques that are applied most often within the QFD framework. Because 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 additional techniques and provide improvements relative to the techniques we report.

We begin with a review of QFD and the voice of the customer. We define customer needs and indicate briefly how they are tied to design goals and decisions. We then focus on each of 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 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).

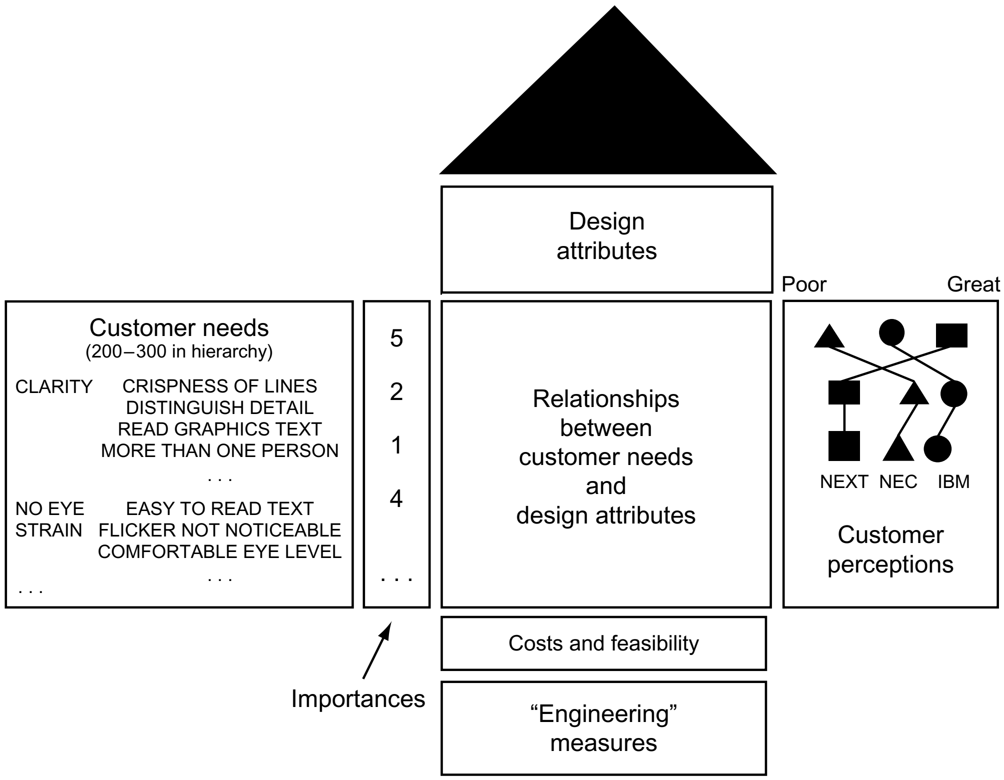
Quality Function Deployment—A Brief Review

Well-established research in the management of technology suggests that cooperation and communication among marketing, manufacturing, engineering, and R&D leads to greater new-product success and more profitable products.[4] QFD improves communication among these 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 1986, Shocker and Srinivasan 1979, Urban and Hauser 1992, Wind 1982), the Lens model (Brunswick 1952, Tybout and Hauser 1981), and benefit structure analysis (Myers 1976). Like these marketing processes QFD uses perceptions of customer needs as a lens by which

[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, portrait services, tools, retirement plans, movie theaters, health insurance, financial services, telephone service, gas and electrical service, distribution networks, automobiles and automobile subsystems and components.

[4] See Cooper (1983, 1984a, b), Cooper and de Brentani (1991), Cooper and Kleinschmidt (1986), Dougherty (1987), de Brentani (1989), Griffin and Hauser (1992b), Gupta et al. (1985), Hise et al. (1990), Moenaert and Souder (1990), Pelz and Andrews (1976), Pinto and Pinto (1990), Souder (1978, 1987, 1988), and others.

Figure 1 The House of Quality from Quality Function Deployment



to understand how product characteristics and service policies affect customer preference, satisfaction, and ultimately, sales. One advantage of QFD is that it 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.[5] 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.

[5] For greater detail see Clausing (1986), Eureka (1987), Griffin (1989), Hauser and Clausing (1988), King (1987), Kogure and Akao (1983), McElroy (1987), and Sullivan (1986, 1987), as well as collections of articles in Akao (1987) and the American Supplier Institute (1987).

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 such as the number of pixels, the size of the screen, the intensity of the pixels, or the refresh rate. The third house of QFD links actions to *implementation* decisions such as manufacturing process operations, and the final house of QFD links the implementation (manufacturing process operations) to *production planning*.

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 to be 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, 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).

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.

Each primary need is elaborated into three-to-ten secondary needs. (Secondary needs are also known as tactical 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" in marketing research that often underlie perceptual maps. (See Green et al. 1988, Lehmann 1985, or Urban and Hauser 1992.)

The tertiary needs, also known as operational needs, provide 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.

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

Segmentation. In many firms, each product-development team focuses on one particular segment of the customer population. A complete "voice" is obtained for each segment. In other firms, only the importances differ among 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.

[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). The most-detailed level included a customer need relating to whether the customer could carry rotten apples in the bed of a pickup truck without worrying about the truck body rusting.

Engineering Input
The team identifies those measurable aspects, called *design attributes*, of the product or service which, if modified, would affect customer's perceptions. Objective *engineering measures* of existing products (the team's and competitors') on the design attributes are obtained and displayed. The *relationship matrix* displays judgments (or experiments) indicating which design attributes affect which customer needs and by how much. 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.) Finally, 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 target 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. (For these operations we require strong scale properties of the measures.)

In closing this section, we note that QFD seems to work. In a study of 35 U.S. projects Griffin (1992) 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 (1992a) 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.

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 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 detailed elaborations 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), Echo techniques (Barthol and Bridge 1968, Barthol 1976), 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?

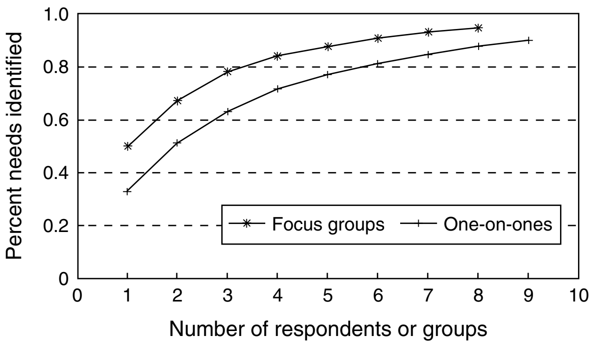
Groups vs. One-on-One Interviews
Many market research firms advocate group interviews (see also Calder 1979) based on the hypothesis that group synergies produce more and varied customer needs as each customer builds upon the ideas of the others. A concern about focus groups is that "airtime" is shared among the group members. If there are eight people in a two-hour group then each person talks, on average, for about 15 minutes.

We compared focus groups to one-on-one interviews in a proprietary QFD application. The product category was a complex piece of office equipment. 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. With our students (Silver and Thompson 1991) we 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. On average, a single one-on-one interview identified 33% of the 230 needs and 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

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



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, an hour of interviewing is an hour of interviewing independently of whether it comes from a one-on-one interview or a focus group.[7] If it is less expensive to interview two consumers for an hour each than to interview 6–8 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.

How Many Customers?

We would like to know how many customers need be interviewed to identify most of the customer needs. Besides intellectual curiosity, there are many reasons for industry to seek an answer to this question. First there is the monetary cost. While the field costs per interview are moderate, analysis costs are quite high. It is typical for some team members to observe each interview and for four or more team members to read each transcript. One major U.S. 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 $1,000–$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).

Time delays accrue if too many interviews are used. Because the timely introduction of new products is 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 service firm was able to gain an additional $150 million in profit by reallocating operating procedures from fulfilling one customer need to fulfilling a different, more important customer need. When the product-development team defended their recommendation to top-level managers, the team was asked to certify that the initial list of needs was based on a sufficient number customers to justify the decision.

Firms also want to be confident that they have interviewed enough customers to uncover most of the exciting needs. Exciting needs, if fulfilled, provide important competitive advantages. For example, Kao developed a highly-concentrated laundry detergent, Attack, that fulfills the need 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.

Firms seek to balance the cost of additional interviews with the benefits of identifying a more complete set of needs. However, the cost (in person-hours) to analyze 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 for two applications—the office equipment application described above and a low-cost durable application described below.

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

[7] In related research, Fern (1982) compared the ability of focus groups and individuals to generate ideas. Reanalysis of his data suggest that approximately 1.2 ideas were generated per minute, independently of whether subjects worked in focus groups or individually.

Figure 3 Number of Customers Who Identify a Need

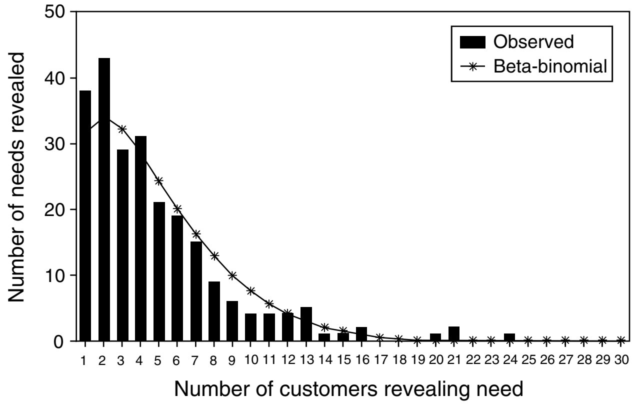
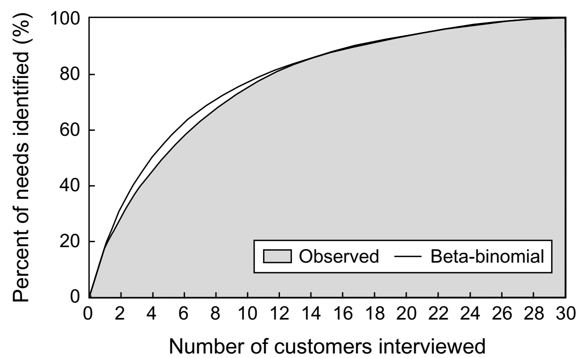


Figure 4 Percent of the Customer Needs Identified by N Customers (Where 30 Customers = 100%)



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 30, 43 needs were identified by two customers out of 30, 29 needs by three customers out of 30, etc. One need was identified by 24 of the 30 customers.

To calculate how many needs we would have expected to obtain from interviewing fewer customers, we consider all possible orderings of the 30 customers and determine the average percent of nonredundant 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 30 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 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 and to concepts developed by Dawkins (1991) and Efron and Thisted (1976).

For 30 customers we observe the outcome of 30 binomial processes. Thus, for 30 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 30 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.[8] This assumption, combined with the binomial processes, gives a beta-binomial distribution for the number of times that needs are voiced in the 30 interviews. The best-fit beta-binomial distribution[9] is plotted in Figure 3. While not perfect it does appear to be a reasonable model.[10]

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. Because each customer is considered an independent draw, the probability that customer need i is identified in a sample of n customers is simply $1-(1-p_i)^n$. However, the probabilities p_i are distributed by the beta distribution. Thus, if the beta-binomial distribution[11] has parameters α and β, then the expected value E_n of the probability of observing a need from n customers is

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

[8] 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.

[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 $p_i = 1 - e^{-\lambda_i t}$. Morrison 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.

[10] 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 that 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.

[11] The Beta distribution is given by

$$f(p) = p^{\alpha-1}(1-p)^{\beta-1}/B(\alpha,\beta) \quad \text{where } B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta).$$

Method of moments estimation gives $\alpha = 1.45$ and $\beta = 7.64$.

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 30 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 2–12 customers, but fits quite well beyond 12 customers. Since most decisions will be made in the range above 12 customers, the model appears accurate enough for our purposes.

What Are We Missing? While 30 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 30 tries. The model estimates that our 30 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 and 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. Based on the transcript-hour equivalence discussed above, this means that 25 hours of interviews 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 portable food-carrying device application. One was an experienced analyst of qualitative data, 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. Figure 6 plots the average cumulative percent of attributes identified as more analysts read the transcripts (observed) compared to a beta-binomial ($\alpha = 22$, $\beta = 19$) model. Based on the model, we estimate that the seven analysts identified 99% of the customer needs obtainable from the transcripts.

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. Sometimes readers who claim category expertise have preconceived notions which causes them to miss surprising or unexpected statements of needs.

If Figure 6 is representative of other categories more than one analyst should read the transcripts. The use of product-development team members brings the added value of team buy-in to the data and greater

Figure 5 **Predictions Based on the Beta-Binomial Model**

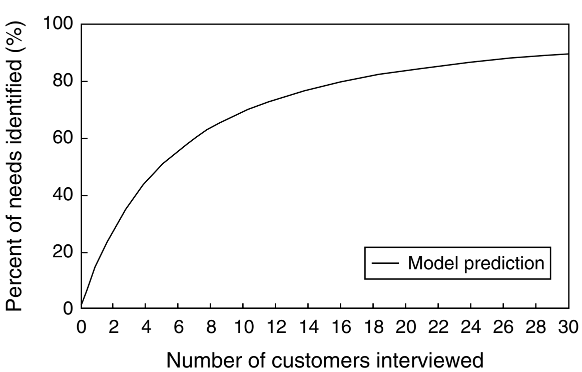
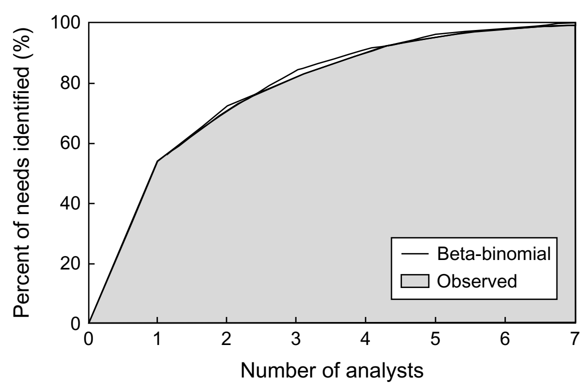


Figure 6 **Ability of Analysts to Identify Customer Needs**



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.

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[12]) 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 card 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.

[12] 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.

Customer Sort and Cluster Process

Green et al. (1969), Rao and Katz (1971), and Green and McMennamin (1973) applied a technique known as subjective clustering in which subjects sort stimuli (e.g., television programs) or activities (e.g., sunbathing) into piles, a similarity matrix is calculated, and either a similarity map or a hierarchical cluster is derived. We modify that data collection procedure to apply to customer needs.

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[13] in which the i-jth 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[14] 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 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

[13] If the number of piles varies dramatically across respondents one can weight the data by a monotonic function (e.g., $\log[\cdot]$) 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. Recent modifications collect importances and customer perceptions for the piles or exemplars.

[14] 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 Ward's structure was slightly superior in terms of face validity to the 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.

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.

The Data

The group-consensus chart for portable food-carrying devices 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 Abbie Griffin, who had observed and/or participated in almost 20 industry 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 60 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 20 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. 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 27 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."

To compare the hierarchies formally we report two statistical measures of structure similarity, Kruskal's λ (Goodman and Kruskal 1954) and an information theoretic measure, U^2 (Hauser 1978). For the primary needs we calculate $\lambda = 0.28$ and $U^2 = 0.30$ and for the secondary needs we calculate $\lambda = 0.51$ and $U^2 = 0.63$.

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	3	Attractiveness	4	20	9
Container utility	2	14	2	Carries many things	2	23	2
Phys. characteristics	10	30	10	Maintains temps.	2	29	6
Thermal attributes	4	34	6	Right size	3	29	6
Convenience	5	139	21	Easy to move	2	30	4
				Convenience	4	31	9
				Works as container	5	39	6
Total	25	217	42		22	201	42
Coeff. of variation	0.6	1.4	0.9		0.4	0.2	0.4

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 of 198 customer needs. 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 "nonexpert" 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 20 labels that were not in the customer-sort chart. (The student team added 14 labels.) In retrospect some of these labels obscured the true customer voice. Statistical analysis suggested that there was 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.

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 (with some minor modifications).

Other Applications

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

The customer-sort hierarchies have been applied over 20 times by one supplier.[15] 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. See also Cooper (1992), Roberts (1992), Ross (1992), and Yie (1992).

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

[15] Private communication with Robert Klein of Applied Marketing Science, Inc. for their Vocalyst™ technique.

they deserve serious consideration for developing the hierarchical structure of customer needs that is used in QFD.

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* 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 insulating material to obtain more insulating power per inch), but at times tradeoffs must be made and priorities set.

In the interest of brevity we resist reviewing the academic literature[16] which is rich in the study of obtaining importances for *attitude*, *preference*, *or utility*. (However, this literature has not addressed explicitly the challenges of obtaining importances for large numbers (200–400) of customer needs.[17])

In this section we focus on importances within the QFD framework of primary, secondary, and tertiary needs. We 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?

[16] 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), formal choice models (Corstjens and Gautschi 1983), and the analytic hierarchy process (Wind and Saaty 1980). For comparisons of methods see Akaah and Korgaonkar (1983), Cattin et al. (1982), Einhorn and Hogarth (1975), Green (1984), Hauser and Koppelman (1977), Hauser and Urban (1979), Hoepfl and Huber (1970), Lehmann (1971), and Schendel et al. (1971).

[17] While a few applications in hybrid conjoint analysis have dealt with large numbers of attributes, e.g., Wind et al. (1989) use 50 product features, the norms in these academic literatures are for far fewer attributes than the 200–400 customer needs that are typical in QFD.

Do Customers Prefer Product Concepts That Emphasize the Fulfillment of "Important" Customer Needs?

The Data. This analysis is based on data collected by an unnamed consumer products firm. The consumer-products firm measured or estimated customer's importances for 198 customer needs using three different methods:[18]

(1) 9-point *direct-rating scale* in which customers answered for each need, "How important is it or would it be if ...?".

(2) *Constant-sum scale* in which customers allocated 100 points among the seven primary 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.

(3) *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.

Questionnaires were mailed to 5,600 randomly selected consumers (1,400 for each method plus 1,400 who rated products on the customer needs). 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 1,400 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 pretests 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 evaluate the concepts by expressing their interest in (9-point

[18] For the constant-sum and anchored scales, the importance of a tertiary need reflects the cascaded allocations to the primary and secondary needs. We also tried a 100-point satisfaction measure and a 9-point like/dislike scale. Results were similar methodologically. We did not test the analytic hierarchy process (AHP) because it would have required 602 pairwise judgments by each respondent.

Table 2 Rank Correlations Between Concept Evaluations and Importances

	Importances		
	Direct	Anchored	Co.-sum
Concept evaluations			
Interest	0.89	0.93	0.93
Preference	0.96	0.96	0.96

scale) and preference for (rank order) the concepts. Table 2 indicates that consumers' interest and preference is highly correlated with the self-stated measures of primary needs.

Which Survey Measure Is Best?

The direct, anchored, and constant-sum measures give similar rank-order results (Table 3) and each correlates with interest and preference (Table 2).[19] 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.[20] In both cases there was agreement between the survey measures of importance. Qualitatively we prefer the anchored scale,[21] but the scientific data to date suggest that any of the three scales could be used to measure importances.

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, anchored, and constant-sum) corresponded to their beliefs about the category—beliefs based on experience and a large number of other market studies. They felt that the simpler self-explicated measures provided sufficiently accurate importance measures for the QFD process and used them to select customer needs upon which to focus. See Hauser (1991) for details.

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, with a nine-point direct scale, importances for the primary, secondary, and tertiary customer needs identified for the portable food-carrying device. 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.

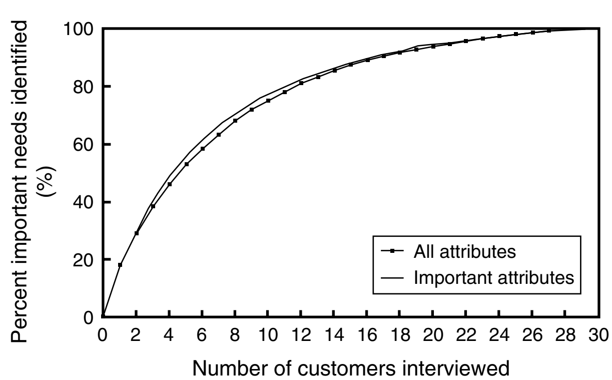
Table 3 Correlations Between Ranks of Mean Importances

	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

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 with respect to the primary needs and regressed those perceptions on customer's satisfaction with that product. Regrettably, the revealed importances did not correlate with either preference for or

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



[19] Tables 2 and 3 report rank correlations. We get similar results for Pearson correlations.

[20] Griffin's study was a pretest of 133 students for the 230 customer needs discussed earlier. She found direct, anchored, and constant-sum 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.

[21] One must be cautious in using either the anchored or 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.

interest in the concepts (−0.36 with interest, −0.14 with preference). This poor predictive ability may be due to the collinearity among primary needs (71% of the correlations are above 0.20). Also, as argued in the next section, monadic satisfaction may be a poor dependent measure.

We have also attempted to estimate revealed importances for a high-cost durable product, for the portable food-carrying devices (Griffin 1989), and for the secondary needs of the consumer product. In all cases collinearity was severe. It was not uncommon that less than 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 cannot 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

Industry accepts customer satisfaction as the goal of QFD because its advocates believe that, in the long-run, satisfied customers are an asset of the firm. Future short-run strategies can be adjusted to draw profitably on this asset.

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 20 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. 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.

Table 4 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

One initial test of this hypothesis is presented in Table 4. This table compares rank-order primary-brand share with monadic satisfaction and with relative-satisfaction from ongoing tracking data at the consumer-products company.[22] 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 4 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 products (albeit to a lesser

[22] 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.

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.

Implications for Technique Comparisons

Self-selection bias also has implications for academic research comparing different measures and estimates of importance weights.[23] 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.

Summary

Our data caution firms that monadic measures of satisfaction lead to counterproductive incentives when evaluating products and product programs. Furthermore self-selection biases (and/or collinearity) can give misleading results when evaluating alternative measures of importance.

[23] 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 applies to studies where subjects evaluate only those products that they would consider seriously.

Table 5 Customer Needs for Spirometry

• Product is affordable	• Easy to hold
• Easy to operate	• Right size for patient
• Easy to clean	• Easy to set up the first time
• Convenient-sized output	• Easy to calibrate
• Sanitary	• Availability machine/supplies
• Quick service response	• Good training/education
• Provides accurate readings	• Sleek appearance
• Eliminates technician variability	• Good printer quality
• Good printout quality	• Low cost of repairs/service
• Reliability	• Portability
• Diagnostic information meets needs	• Effective data storage/retrieval
• Easy-to-interpret diagnostic information	• Environmentally safe
• Fast to use	

Applications[24]

The development of Puritan-Bennett's Renaissance™ Spirometry System is typical of voice-of-the customer applications. (A spirometer is a medical instrument that measures lung capacity—an important indicator of general health.) Puritan-Bennett's (PBs) PB900A was a major player in the market but in 1989 its share slipped from 15% to 7% as a result of a new product by Welch Allyn (WA). WAs PneumoCheck was introduced at a dramatically lower price ($1,995 vs. $4,500 for WA) made possible by a reduced functionality—the PneumoCheck measured a person's ability to exhale while the PB900A measured both exhaling and inhaling. PB considered a cost-reduction program but felt that the basic design of the PB900A would make it impossible to come close to the WA price. Instead, they started the design from scratch based on QFD and the voice of the customer.

An interfunctional team drawn from marketing, customer service, sales, engineering, R&D, manufacturing, and management began with qualitative interviews and focus groups with their customers to identify the customer needs. Following procedures discussed in this paper PB structured the customer needs into a hierarchy, measured importances, and measured customer perceptions of both PB and competitive products. For example, the secondary needs are shown in Table 5.

By focusing development on these needs, within a year they designed an entirely new modular spirometry system that could be customized by each user segment (hospitals, large laboratories, small clinics, and general practitioners). For example, "affordability"

[24] For more details and for eight additional application vignettes see Hauser (1992). The application vignettes include a stationary-products manufacturer, a construction–tools manufacturer, a financial institution, an insurance company, an entertainment provider, a manufacturer of office equipment, a manufacturer of surgical instruments, and a manufacturer of a lightweight chemical mixing device.

was achieved with the modular prices, "good printout quality" by using the customers' existing printers rather than the traditional approach of a built-in thermal printer, "effective data storage and retrieval" by plug-in patient data cards, "easy to hold" by a detachable, rechargeable unit that makes and stores measurements without the base station, etc. Each customer need was met at a level that matched or exceeded competition. An occasional-screening system was priced $405 below WA, but heavy users could increase functionality and productivity with a system of three spirometers, two base stations, two charging stations, two memory cards, and a Canon Bubblejet™ printer at a cost of $4,088.

Discussion and Summary

Quality Function Deployment (QFD), which promises decreased product-development costs, decreased product-development time, and improved customer satisfaction, has been adopted widely by U.S. and Japanese product-development teams. QFD begins with the voice of the customer—a prioritized list of 200–400 customer needs in a hierarchical tree of primary strategic needs, secondary tactical needs, and tertiary operational needs.

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 be 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.

While group-consensus charts are the most popular method for obtaining a hierarchical structure, our data suggest that different structures are obtained by analyzing customer-sort data. 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.

Our comparison of importance-measurement techniques 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.

The stated goal of QFD is customer satisfaction. Our data suggest that a self-selection bias might be present in standard customer-satisfaction data collected by corporations. This bias could give counter-productive incentives encouraging firms to retreat to small segments of easy-to-satisfy customers, or to (inadvertently) implement policies that increase average satisfaction by getting rid of dissatisfied customers.

Many challenges remain. Perhaps other techniques will prove superior to those that we studied. 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. 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 research problems in industry. For example, 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. Fulfilling exciting customer needs leads to breakthrough products. Perhaps leading-edge user studies (von Hippel 1986) can be developed to identify these exciting needs.[25]

Acknowledgments

Funds for the research in the paper were provided by M.I.T.'s program for the Management of Technology, Kirin Brewing Company, the Marketing Science Institute, the Industrial Research Institute, an anonymous office-equipment manufacturer, and an anonymous consumer-products company. Abbie Griffin was partially supported by an AAUW American Fellowship.

References

Akaah IP, Korgaonkar PK (1983) An empirical comparison of the predictive validity of compositional, decompositional, and hybrid multiattribute preference models. *J. Marketing Res.* 20(May):187–197.

Akao Y (1987) *Quality Deployment: A Series of Articles*, Glen Mazur, trans. (GOAL, Inc., Lawrence, MA).

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

Barthol RP (1976) ECHO: Estimating values in familiar and unfamiliar cultures. Sinaiko HW, Broedling LA, eds. *Perspectives on Attitude Assessment: Surveys and Their Alternatives* (Pendelton Press, Champaign, IL).

Barthol RP, Bridge RG (1968) The echo multi-response method for surveying value and influence patterns in groups. *Psych. Rep.* 22:1345–1354.

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

[25] This paper was received February 5, 1991, and has been with the authors 7 months for 2 revisions. Processed by Anne T. Coughlan, Area Editor.

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

Calder BJ (1979) Focus groups and the nature of qualitative marketing research. *J. Marketing Res.* 14(August):353–364.

Cattin P, Hermet G, Pioche A (1982) Alternative hybrid models for conjoint analysis: Some empirical results. *Analytical Approaches to Product and Market Planning: The Second Conference*, October (Marketing Science Institute, Cambridge, MA), 142–152.

Clausing D (1986) *QFD Phase II: Parts Deployment* (American Supplier Institute, Dearborn, MI).

Cooper LK (1992) The structure of recruiter need at the Sloan School of Management: A quantitative assessment. Master's thesis, Sloan School of Management, MIT, Cambridge, MA.

Cooper RG (1983) The new product process: An empirically-based classification scheme. *R&D Management* 13(1):1–13.

Cooper RG (1984a) New product strategies: What distinguishes the top performers? *J. Product Innovation Management* 2:151–164.

Cooper RG (1984b) How new product strategies impact on performance. *J. Product Innovation Management* 2:5–18.

Cooper RG, de Brentani U (1991) New industrial financial services: What distinguishes the winners. *J. Product Innovation Management* 8:75–90.

Cooper RG, Kleinschmidt E (1986) An investigation into the new product process: Steps, deficiencies, and impact. *J. Product Innovation Management* 3:71–85.

Corstjens ML, Gautschi DA (1983) Formal choice models in marketing. *Marketing Sci.* 2(1, Winter):19–56.

Dawkins B (1991) Siobhan's problem: The coupon collector revisited. *Amer. Statist.* 45(1, February):76–82.

de Brentani U (1989) Success and failure in new industrial services. *J. Product Innovation Management* 6:239–258.

Dougherty DJ (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, MIT, Cambridge, MA.

Efron B, Thisted R (1976) Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* 63(3): 435–447.

Einhorn HJ, Hogarth RM (1975) Unit-weighting schemes for decision making. *Organ. Behav. Human Performance* 13:171–192.

Eureka WE (1987) Introduction to quality function deployment. *Quality Function Deployment: A Collection of Presentations and QFD Case Studies*, Section III (American Suppliers Institute, Dearborn, MI).

Fern EF (1982) The use of focus groups for idea generation: The effect of group size, acquaintanceship, and moderator on response quantity and quality. *J. Marketing Res.* 19(February): 1–13.

Fornell C (1991) A national customer-satisfaction barometer: The Swedish experience. Working paper, The Office for Customer Satisfaction Research, University of Michigan, Ann Arbor, MI.

General Accounting Office, U.S. (1991) Management practices: U.S. companies improve performance through quality efforts. GAO/NSIAD-91-190, (May), Gaithersburg, MD.

Goodman LA, Kruskal WH (1954) Measures of association for cross-classification. *J. Amer. Statist. Assoc.* 49:732–764.

Green PE (1984) Hybrid models of conjoint analysis: An expository review. *J. Marketing Res.* 21(2, May):155–169.

Green PE, McMennamin JL (1973) Market position analysis. Britt SH, Guess NF, eds. *Marketing Manager's Handbook* (Dartnell Press, Chicago, IL), 543–554.

Green PE, Srinivasan V (1978) Conjoint analysis in consumer research: Issues and outlook. *J. Consumer Res.* 5(September): 103–123.

Green PE, Srinivasan V (1990) Conjoint analysis in marketing research: New developments and directions. *J. Marketing* 54(4, October):3–19.

Green PE, Carmone FJ, Fox LB (1969) Television programme similarities: An application of subjective clustering. *J. Market Res. Soc.* 11(1):70–90.

Green PE, Tull DS, Albaum G (1988) *Research for Marketing Decisions*, 5th ed. (Prentice-Hall, Englewood Cliffs, NJ).

Griffin A (1989) Functionally integrated new product development. Unpublished Ph.D. thesis, Sloan School of Management, MIT, Cambridge, MA.

Griffin A (1992) Evaluating development processes: QFD as an example. *J. Product Innovatton Management* 9(September).

Griffin A, Hauser JR (1992a) Patterns of communication among marketing, engineering and manufacturing—A comparison between two new product teams. *Management Sci.* 38(3, March): 360–373.

Griffin A, Hauser JR (1992b) The marketing and R&D interface. University of Chicago, Working paper (October).

Gupta AK, Raj SP, Wilemon D (1985) R&D and marketing dialogue in high-tech firms. *Indust. Marketing Management* 14:289–300.

Gutman J (1982) A means-end chain model based on consumer categorization processes. *J. Marketing* 46(Spring):60–72.

Hauser JR (1978) Testing the accuracy, usefulness, and significance of probabilistic choice models: An information-theoretic approach. *Oper. Res.* 26(3, May–June):406–421.

Hauser JR (1991) Comparison of importance measurement methodologies and their relationship to consumer satisfaction. MIT working paper, Sloan School of Management, MIT, Cambridge, MA.

Hauser JR (1992) Puritan-Bennett—The Renaissance™ Spirometry System: Listening to the voice of the customer. MIT working paper, Sloan School of Management, MIT, Cambridge, MA.

Hauser JR, Clausing DP (1988) The house of quality. *Harvard Bus. Rev.* 66(3, May–June):63–73.

Hauser JR, Koppelman FS (1977) Effective marketmg research: An empirical comparison of techniques to model consumers' perceptions and preferences. Shocker AD, ed. *Analyttc Approaches to Product and Market Planning* (Marketing Science Institute, Cambridge, MA).

Hauser JR, Urban GL (1977) A normative methodology for modeling consumer response to innovation. *Oper. Res.* 25(4, July–August):579–619.

Hauser JR, Urban GL (1979) Assessment of attribute importances and consumer utility functions. *J. Consumer Res.* 5(March): 251–262.

Hise RT, O'Neal L, Parasuraman A, McNeal JU (1990) Marketing/R&D interaction in new product development implications for new product success rates. *J. Product Innovation Management* 7(2, June):142–155.

Hoepfl RT, Huber GP (1970) A study of self-explicated utility models. *Behavioral Sci.* 15:408–414.

Huber GP (1974) Multi-attnbute utility models: A review of field and field-like studies. *Management Sci.* 20(10, June):1393–1402.

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

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

Kelly GA (1955) *The Psychology of Personal Constructs*, Vol. 1 (W. W. Norton, New York).

King R (1987) *Better Designs in Half the Time: Implementing Quality Function Deployment (QFD) in America* (GOAL, Lawrence, MA).

Kogure M, Akao Y (1983) Quality function deployment and CWQC. *Quality Progress* 16(10, October):25–29.

Lehmann DR (1971) Television show preference: Application of a choice model. *J. Marketing Res.* 8(February):47–55.

Lehmann DR (1985) *Market Research and Analysis*, 2nd ed. (Irwin, Homewood, IL).

Lynch JG Jr (1985) Uniqueness issues in the decompositional modeling of multi-attribute overall evaluations: An information integration perspective. *J. Marketing Res.* 22(February):1–19.

Manski CF, McFadden D (1981) *Structural Analysis of Discrete Data with Econometric Applications* (MIT Press, Cambridge, MA).

McElroy J (1987) For whom are we building cars? *Automotive Indust.* (June):68–70.

Moenaert RK, Souder WE (1990) An information transfer model for integrating marketing and R&D personnel in new product development projects. *J. Product Innovation Management* 7(2, June):91–107.

Morgan AI (1984) Point of view: Magic town revisited (a personal perspective). *J. Advertising Res.* 24(4, August/September): 49–51.

Morrison DG (1979) An individual differences pure extinction process. *J. Math. Psych.* 19(3, June):307–315.

Myers JH (1976) Benefit structure analysis: A new tool for product planning. *J. 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.

Pelz DC, Andrews FM (1976) *Scientists in Organizations*, revised ed. (University of Michigan Press, Ann Arbor, MI).

Pessemier EA (1986) *Product Management: Strategy and Organization* (Robert E. Krieger Publishing, Malabar, FL).

Phillips S, Dunkin A, Treece J, Hammonds KH (1990) King customer: At companies that listen hard and respond fast, bottom lines thrive. *Bus. Week* (March 12):88–94.

Pinto MB, Pinto JK (1990) Project team communication and cross-functional cooperation in new program development. *J. Product Innovation Management* 7:200–212.

Rao V, Katz R (1971) Alternative multidimensional scaling methods for large stimulus sets. *J. Marketing Res.* 8(November):488–494.

Reimann C (1991) The Baldrige Award criteria speak for themselves. *Quality Progress* 24(May):43–44.

Reynolds TJ, Gutman J (1988) Laddering theory, method, analysis, and interpretation. *J. Advertising Res.* 28(1):11–31.

Roberts JW (1992) M.B.A. recruiters' needs: Voice of the customer analysis. Master's thesis, Sloan School of Management, MIT, Cambridge, MA.

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

Ross LG (1992) A voice of the customer analysis of M.B.A. schools: The student segment. Master's thesis, Sloan School of Management, MIT, Cambridge, MA.

Ruiz D, Jain D (1991) Designing and developing new products: The subproblem decomposition approach. Working paper, Kellogg Graduate School of Management, Northwestern University, Evanston, IL.

Schendel DE, Wilkie WL, McCann JM (1971) An experimental investigation of attribute importance. *Proc.: Second Annual Conf.*, Association of Consumer Research, 404–416.

Shocker AD, Srinivasan V (1979) Multi-attribute approaches for product concept evaluation and generation: A critical review. *J. Marketmg Res.* 16(May):159–180.

Silver JA, Thompson JC Jr (1991) Understanding customer needs: A systematic approach to the "voice of the customer." Master's thesis, Sloan School of Management, MIT, Cambridge, MA.

Souder WE (1978) Effectiveness of product development methods. *Indust. Marketing Management* 7.

Souder WE (1987) Managing the R&D/marketing interface. *Managing New Product Innovations*, Chap. 10 (Lexington Books, Lexington, MA).

Souder WE (1988) Managing relations between R&D and marketing in new product development projects. *J. Product Innovation Management* 5:6–19.

Stratton B (1991) The value of implementing quality. *Quality Progress* 25(7, July):70–71.

Sullivan LP (1986) Quality function deployment. *Quality Progress* 19(6, June):39–50.

Sullivan LP (1987) The power of Taguchi methods. *Quality Progress* 12(6, June):76–79.

Tybout AM, Hauser JR (1981) A marketing audit using a conceptual model of consumer behavior: Application and evaluation. *J. Marketing* 45(Summer):82–101.

Urban GL, Hauser JR (1992) *Design and Marketing of New Products* II (Prentice-Hall, Englewood Cliffs, NJ).

von Hippel E (1986) Lead users: A source of novel product concepts. *Management Sci.* 32(7, July):791–805.

Vorberg D, Ulrich R (1987) Random search with unequal search rates: Serial and parallel generalizations of McGill's model. *J. Math. Psych.* 31:1–23.

Wilkie WL, Pessemier EA (1973) Issues in marketing's use of multi-attribute attitude models. *J. Marketing Res.* 10(November): 428–441.

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

Wind Y, Saaty TL (1980) Marketing applications of the analytic hierarchy process. *Management Sci.* 26(7, July):641–658.

Wind Y, Green PE, Shifflet D, Scarbrough M (1989) Courtyard by Marriott: Designing a hotel facility with consumer-based marketing models. *Interfaces* 19(January–February):25–47.

Yie S (1992) The core curriculum at Sloan: Establishing a hierarchy of needs. Master's thesis, Sloan School of Management, MIT, Cambridge, MA.