COMPARISON OF IMPORTANCE MEASUREMENT METHODOLOGIES AND THEIR RELATIONSHIP TO CONSUMER SATISFACTION

Professor John R. Hauser
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

M.I.T. Marketing Center Working Paper 91-1

January 1991

COMPARISON OF IMPORTANCE MEASUREMENT METHODOLOGIES AND THEIR RELATIONSHIP TO CONSUMER SATISFACTION

ABSTRACT

In the context of Quality Function Deployment (QFD) and the House of Quality, consumer needs provide the goals for product development. Importances of those consumer needs enable QFD teams to focus and prioritize strategies and tactics. In this report we compare four alternative methods to measure consumer-needs importances. Based on the analysis of data obtained from 5600 consumers in a major consumer-product category, we conclude that at least one direct measure, self-stated importances, has sufficient validity.

Specifically, (1) the self-stated direct measure gives believable and interpretable priorities that are consistent with the beliefs of the research sponsor, the product development division of one of the premier consumer-packaged-good companies, and (2) the importances predict consumer interest and preference for new-product concepts. The other two direct measures, anchored and constant-sum, provide similar implications for the importances of the primary needs and similar implications when the secondary and tertiary needs are worded such that they reflect more detailed needs. However, their accuracy depends strongly on such wording. Both the research sponsor and the author believe that such wording is difficult to guarantee. The direct measure, revealed preference, does not give believable results and does not predict consumer interest and preference for new-product concepts.

"Satisfaction indices" based on importances are compared to measured satisfaction. These indices correlate with satisfaction, but not significantly better than indices based on simply summing consumer evaluations. This lack of improvement is due, in part, to concerns about the satisfaction measure which was used in the data collection, in part, to the robustness of the linear model, and, in part, to the fact that it was difficult to classify consumers to segments based on demographic characteristics alone.

An important, but serendipitous, finding is that the satisfaction measures used in the study may be misleading because they confound variation among consumers with variation in the chosen product. In fact, the measures do not correlate with primary-brand share. We strongly recommend experimentation with alternative measures. For example, *relative* satisfaction measures, where consumers rate all products that they consider, show promise as do share-adjusted measures. In particular, the relative-satisfaction measures that the research sponsor collects among representative panels correlate well with primary-brand share. Until this issue is resolved, evaluation of product programs should not be based on monadic measures of consumer satisfaction.

A NOTE ON THE RESEARCH SPONSOR AND CONFIDENTIALITY

This document reports the results of joint research by M.I.T. and an unnamed research sponsor (RS). Many of the ideas expressed in this report are the basis of long discussions between M.I.T. and the RS and many reflect statistical analysis undertaken by the market research department at the RS. I hereby acknowledge this important input and thank them for their insight. (Naturally, all mistakes and opinions are my responsibility.)

Because the RS believes in the advancement of science, they are allowing M.I.T. to publish all summary statistics and conclusions of a methodological nature. However, they have asked that all information about the product category, the specific consumer needs, and the specific segment names, and even the name of the RS be disguised. This does not affect the scientific content, but in a few cases it means that the reader can not apply his (her) own judgment to the qualitative results. For this I ask you to bear with me and trust the RS's judgment when they say that "primary need x" is ranked better by one methodology than another.

The RS is a consumer-products company with experience in many markets. The product category is one with which I am sure you are familiar. I can say no more.

Finally, I would like to thank the M.I.T. students who put in long hours on the statistical analysis. They are Abbie Griffin, Ken Chay, Ning Peng, and Jae-kue Park.

COMPARISON OF IMPORTANCE MEASUREMENT METHODOLOGIES AND THEIR RELATIONSHIP TO CONSUMER SATISFACTION

TABLE OF CONTENTS

FOUR ALTERNATIVE METHODS 3
Self-stated Ratings 3
Anchored Scales 3
Constant-sum Scales 5
Revealed Preference 5

DEFINING WHAT IS "BEST" 7
"Stretching" Concepts 8
Correlation Between Stated Satisfaction and Implied Satisfaction 8
Unit-weights Satisfaction Index 9
Rescaling the importances 10
Face Validity 10

PREDICTING CONSUMER INTEREST IN CONCEPTS 13

SATISFACTION VS. THE SATISFACTION INDEX 13
Primary Needs 15
Secondary and Tertiary Needs 15
Rescaling Constant 16

ISSUES IN THE MEASUREMENT OF SATISFACTION 16
Satisfaction vs. "Primary-brand Share" 17
Potential Scales 18
A Relative Scale used by the RS 19

SEGMENTATION 20
Discriminant Analysis 21
Correlation of Satisfaction and the Satisfaction Index by Segment 21
Reverse Engineering 22

RECOMMENDATIONS 23

Appendix: REVEALED PREFERENCE. A1

Many leading American corporations focus on designing and delivering "quality" to their consumers. But defining quality is not an easy task. Quality has many meanings, meanings which vary by organization and by functional areas within organizations. Furthermore, definitions of quality are elaborated to different levels of detail depending upon the actions (strategic, tactical, or detailed implementation) being considered. Responding to evidence[1] that interfunctional communication enhances new-product success, a consumer-products company, which has experience in new-product development, has been experimenting with a management technique known as Quality Function Deployment (QFD). A key component of the QFD approach is to identify and prioritize the "voice of the consumer," that is, consumer needs. These consumer needs provide a common language with which marketing and product development teams discuss actions and strategies. A set of well-developed consumer needs assures that products are developed, modified, and marketed based on the consumer's perspective.

The voice of the consumer can become quite detailed. For example, in the product category being studied, the research sponsor (RS) identified a set of 198 consumer needs. To work with such a detailed list and to encourage coordination among the various functions represented on the QFD team, the consumer needs are structured into a hierarchy of primary, secondary, and tertiary needs. (The structure was based on the statistical analysis of the patterns by which consumers sorted needs. The labels were drawn from those that consumers highlighted. Naturally, the structure and labels were supplemented with managerial judgment.)

Primary needs provide strategic direction for the QFD process. In this study there were seven primary needs numbered 1000 through 7000. For example, the core benefit of one product might be based on need 1000, while the core benefit of another product might be based on need 7000. However, such strategic direction depends upon any segmentation strategy and upon the accurate wording of the primary need. For example, if need 5000 does not represent to the detailed needs, then the consumer may not interpret the primary need correctly and strategic directions may not be accurate. In this study one of the seven needs, need 5000, did not reflect the detailed needs.

Secondary needs elaborate the consumer's voice. For example, three secondary needs elaborate primary need 1000 and in doing so suggest specific actions. These tactical needs are useful for making decisions on the allocation of research effort and in setting marketing tactics such as advertising copy.

But if product development is to improve their products, they need more direction than that provided by the secondary needs. For example, there were 15 tertiary needs that provided detail with respect to secondary need 1100. These needs provide specific goals for the product development process. Based on these tertiary needs, QFD, through the "House of Quality" relates specific actions -- formulations, packaging, distribution systems, etc. -- to assure that the

---

[1] See for example Cooper, Robert G. (1984), "New Product Strategies: What Distinguishes the Top Performers?," *Journal of Product Innovation Management*, 2, 151-164.; Hauser, John R. and Don P. Clausing (1988), "The House of Quality," *Harvard Business Review*, 66, 3, (May-June), 63-73.; Griffin, Abbie (1989), "Functionally Integrated New product Development," Unpublished Ph.D. Thesis, Sloan School of Management, M.I.T., Cambridge, MA 02139 (June).; or American Supplier Institute (1987), *Quality Function Deployment: A Collection of Presentations and QFD Case Studies*, (Dearborn, MI: American Supplier Institute, Inc.).

product and marketing programs respond to consumer needs.

In the product category being studied there are 7 primary needs, 36 secondary needs, and 155 tertiary needs. To allocate effort among these needs, the team must establish priorities. In QFD these priorities are based on evaluations of competitive products, on the cost and difficulty of improving the product with respect to a consumer need, and on the importance that consumers place on the needs. For example, if it is much more important to a segment of consumers that a product excels on need 1000 than it is that the product excels on need 7000, then product development will make the appropriate tradeoffs balancing cost, feasibility, and competitive advantage. (Or, perhaps, identify a breakthrough technology to meet both of the primary needs.) Relative importances do not provide the final answer, but they are key data in the product-development decision.

Because product importances play a key role in QFD, the RS and M.I.T. undertook a research project to understand better how to measure the importances of consumer needs. Our primary methodological goal was to establish whether or not the measured importances had sufficient validity to be used in the allocation of research effort. Within this context we hoped to explore (and solve) the challenges of data collection and analysis so that we could provide acceptable estimates of the importances of consumer needs.

Because QFD uses a large number of consumer needs, the measurement of relative importances strains traditional data collection. Methods that work well for a relatively short questionnaire containing 20-30 consumer needs may or may not translate to a questionnaire containing 198 consumer needs. Furthermore, the hierarchical nature of the needs may provide new opportunities to collect data while at the same time raising new issues in the relationships among levels in the hierarchy. Thus, the RS and M.I.T. designed a research study in which four different methods were tried -- a traditional importance scale, two methods based on the hierarchy, and a method by which importances were "revealed" through statistical analysis of consumer satisfaction ratings. By trying four different methods we felt we had a better chance to identify one method that was reasonable We also hoped to identify one method which would be superior to the others.

The remainder of this report addresses these issues of the validity and the comparison of the four methods. We provide evidence that, at the primary-need level, the methods have validity in their ability to predict interest in product concepts and that they provide believable outputs. However, we were not able to establish that one method dominated the rest with respect to this criterion, only that the three direct consumer measures appear to be superior to revealed preference (for this data). We also examined the qualitative implications of the alternative measures. Based on their experience and expertise in the product category, the RS identified a critical concern with the anchored and constant-sum scales. In particular, both measures cascade primary-need importances to lower levels in such a manner that the importance of a detailed need is limited by the importance of the primary need (secondary need) in which it is grouped. If the group labels do not reflect these detailed needs, then the measured

importances of the labels do not apply to the detailed needs. Self-stated measures are not subject to this criticism.

An important, but serendipitous, discovery was that measures of consumer satisfaction could be misleading if they were based on only the chosen product. We discuss this phenomenon and suggest alternative satisfaction measures.

FOUR ALTERNATIVE METHODS

The four methods were self-stated ratings, anchored scales, constant-sum scales, and revealed preference.

Self-stated Ratings

In the self-stated rating technique, each primary, secondary, and tertiary need was rated on a nine-point scale. Four different question orders were used to minimize order effects. Tertiaries were not grouped with secondaries and secondaries were not grouped with primaries. Thus, consumers evaluated each need by its own merits independent of the hierarchy.

Table 1 Example Self-Stated Ratings

	Not at all Important		Somewhat Important		Important		Very Important		Extremely Important
How important is it or would it be if: The products	1	2	3	4	5	6	7	8	9
How important is it or would it be if: Another consumer need	1	2	3	4	5	6	7	8	9
How important is it or would it be if: Still another consumer need	1	2	3	4	5	6	7	8	9

The advantages of the self-stated technique are that it has an established track record with the RS's market research department and that consumers find it easy to answer. It does not use the explicit structure of the hierarchy, but *a priori* that may be good or bad. Consumers are not constrained by the hierarchy, but the hierarchy contains additional information with which to interpret consumer response.

Anchored Scales

The anchored scales exploit the special structure of the hierarchy. Consumers are shown first the primary needs and asked to assign a rating of 10 to the most important need. The other

Table 2 Anchored Ratings for Primary Needs

When thinking about .... overall, how important is it that:

[ ] Primary need 1000. (Actual wording of consumer need
[ ] Primary need 2000. was used in the questionnaire.)
[ ] Primary need 3000.
[ ] Primary need 4000.
[ ] Primary need 5000.
[ ] Primary need 6000.
[ ] Primary need 7000.

primaries are rated from 0 to 10 depending upon their relative importance. Consumers are next given the secondary elaborations of each primary and asked to rate them on the same type of scale -- 10 to the most important secondary need *as it relates to the primary need*, and a number from 0 to 10 for each of the other related secondary needs. Finally, consumers complete the scales for tertiary needs as they relate to secondary needs.

Table 3 Anchored scale for secondary elaboration.

Considering primary need 1000, how important is it that:

[ ] Secondary need 1100.
[ ] Secondary need 1200.
[ ] Secondary need 1300.

(Naturally, the consumers' wording of the primary and the secondary needs was used in the question.)

To interpret the ratings the data is cascaded. Primary ratings remain unchanged. Each secondary rating is multiplied by the relevant primary rating. For example, if the consumer gives a '9' to primary 1000 in table 2 and an '8' to secondary need 1100 in table 3, then the importance of secondary need 1100 is 7.2, that is 9 times 8, divided by 10 to keep the importance in the range of 0-10. Similarly, a tertiary importance is the cascaded product of the tertiary rating times the relevant secondary rating times the relevant primary rating, divided by 100. Note that because the correction factor, division by 10 or 100, is chosen arbitrarily, one can not make comparisons between levels. It is valid (analytically) to compare two tertiaries, even if they correspond to different secondary needs, but one can not compare a tertiary need with a secondary need. Given the nature of the hierarchy in QFD, this restriction should pose no problems in using the House of Quality.

The advantage of the anchored method is that it uses the information contained in the hierarchy and provides answers that are consistent with the use of strategic, tactical, and detailed needs. It also asks the consumer to make explicit tradeoffs among a relatively small group of needs. However, this advantage may be illusionary if the labels of the primary needs are not descriptive of the integrating need or if the consumer feels that a secondary need associated with an unimportant primary need is really important (or visa versa). In our data set this was particularly true for need 5000 which the RS feels does not reflect fully the implied tertiary needs. The mismatch in wording remained despite a careful procedure which included consumer input and managerial judgment and which was subject to a series of pre-test analyses. (However, once the mismatch was uncovered the RS was able to interpret the importances within need 5000 and was able to interpret the importance of need 5000 as measured.)

Constant-sum Scales

Constant-sum scales are similar to anchored scales in that they rely upon and exploit the hierarchy. They differ in that consumers are asked to allocate 100 points among the needs. The theory is that by asking consumers to make a tradeoff among a fixed number of points, the consumers will provide more accurate importances.

Table 4 Constant-sum Ratings for Primary Needs

When thinking about .... overall, how important is it that:

[ ] Primary need 1000. (Actual wording of consumer need
[ ] Primary need 2000. was used in the questionnaire.)
[ ] Primary need 3000.
[ ] Primary need 4000.
[ ] Primary need 5000.
[ ] Primary need 6000.
[ ] Primary need 7000.
100

However, the constant-sum technique raises an issue of scaling. For example, suppose that there are two primary needs, A and B, and that A is more important, receiving 70 of the 100 points. Suppose further that there are 10 secondaries under primary A and that the most important secondary receives 15 points. Its importance would be (70 x 15)/100 = 10.50 points. (We divide by 100 to maintain a two digit number.) Suppose that there are only two secondaries under primary B and that the least important receives a score of 40. Its importance would be (30 x 40)/100 = 12.00. The least important secondary under the least important primary is more important than the most important secondary under the most important primary! To avoid this anomaly, we correct for the variation in the number of secondaries under primaries by multiplying the result by the number of secondaries. By this technique we force the average importance of a group of secondary needs to be equal to the importance of the corresponding primary need. We scale tertiary needs by a similar method.

The strengths and weaknesses of the constant-sum scale are similar to those of the anchored scale. The same caveats apply, especially the caveat about accurate wording of the primary and secondary needs.

Revealed Preference

The fourth method of identifying importances is fundamentally different than the other three methods. While the other methods depend on the ability of consumers to provide direct indications of importances, revealed preference attempts to infer importances from patterns of overall satisfaction as they relate to consumer evaluations of their current product process.

As part of the background information, consumers are asked to describe their current product process and to indicate their overall satisfaction with the way in which the product is used. Satisfaction was measured by two methods -- the nine-point scale shown in table 5 and a 100-point scale which asked for a direct assignment of a number between 0 and 100. In addition, consumers were asked to evaluate the product process on a nine-point *like or dislike*

scale.

Table 5 Nine-point Satisfaction Scale

1. Think about the way you NOW .... from an overall point of view, that is, the steps you have to take and the results you get. On the scale below, please circle the number that indicated how you would rate the way you NOW ....

Excellent		Very Good		Good		Fair		Poor
9	8	7	6	5	4	3	2	1

Consumers are then asked to evaluate their current product (or the process in which it is used) with respect to the consumer needs. The measurement scale that was chosen was a six-point agree-disagree scale. Because some scales were stated positively, "I feel great about ...," and some scales were stated negatively, "The products cause ... problems.," the RS rescaled the data so that a larger number implies a greater need. (An alternative interpretation is that a smaller number implies that the consumer is more satisfied with respect to that need.)

Table 6 Evaluation of Consumer Needs

THE WAY I ....:

	Strongly agree	Agree	Slightly agree	Slightly disagree	Disagree	Strongly disagree	Not Apply
The products	6	5	4	3	2	1	
Another consumer-need	6	5	4	3	2	1	
Still another consumer-need	6	5	4	3	2	1	

To see how revealed preference would work, suppose that there are two products (or product-based processes), X and Y, and two primary needs, A and B. Suppose that product X does very well on need A and poorly on need B and that product Y does very well on need B and poorly on need A. Then, if consumers are more satisfied with product X than with product Y, we might infer that need A, the need on which X excels, is more important than need B, the need on which Y excels. With multiple products and with multiple needs, the statistical problem requires multiple regression, but the concept is similar.

The theoretical advantage of revealed preference is that we can infer importances from overall satisfaction rather than asking consumers to provide them directly. But there are some problems with the technique. Two of these problems relate to the satisfaction scale and one is technical.

The first issue with the satisfaction scale is that it relates to the product process rather than a specific brand. This requires that any analyses must use evaluations at the level of the product process. In our questionnaire we were careful to use consistent definitions for the

satisfaction measures and the evaluation measures, they both referenced the product process, but this issue should be kept in mind in any comparisons among brands. For example, any evaluations would apply to a product process based on product 1, not just to product 1.

The second issue is more fundamental. The only variation in satisfaction ratings is among different consumers, but different consumers also use different product processes. If consumers using different processes also vary in the importances they place on different consumer needs -- not an unreasonable expectation if they chose their product process based on how it fulfills their needs, then the satisfaction scale confounds choice, relative importances, and satisfaction. Let's take an example. Suppose that demanding consumers who care very much about need 1000 choose a product process based on product 1. A product 1-based process may be best relative to their needs, but, because they are demanding, they may not be satisfied with even the best process. By the same token, casual users may be less discriminating and choose a process based on a competitive product, say product 2, for other reasons, perhaps availability in the store or because a parent used it. Because they have lower expectations, they may be quite satisfied with a product 2-based process. If this is the case, then product 2 may have a higher satisfaction rating even though product 1 is actually the better product. We show evidence later in this report that this confounding effect might be present in the data. Note that the confound would be less had we obtained *relative* satisfaction by asking consumers to evaluate all brands they consider or all brands of which they are aware.

The technical issue is called "multi-collinearity." For multiple regression to work properly, the evaluations of consumer needs must vary independently among consumers. By contrast, if products that are good on primary need 1000 are always low on primary need 7000 and visa versa, and if consumers choose products based on primary need 1000, then the statistics would also imply that consumers like products are poor on primary need 7000. Unfortunately, if the hierarchy is an accurate representation of consumer needs, then secondaries within primaries and tertiaries within secondaries should be highly intercorrelated and multi-collinearity will be a problem at the secondary and tertiary levels. This is the case in our data.

DEFINING WHAT IS "BEST"

The RS's goal is long run profit and profit depends upon sales (market share) and costs. Thus, the "best" importance measurement method is the method that helps the RS's product-development and marketing teams make the tradeoffs among consumer needs that lead to more sales and/or lower costs. The best *scientific* comparison of alternative importance measurement techniques would be to appoint four matched QFD teams, assign a set of importances to each, and have each work independently to develop and manage a new product. *Practically* this is impossible, not only because it would be difficult to carry out the experiment, but because it would be an expensive, inefficient use of scarce management resources. In light of this challenge, the RS and M.I.T. used three proxies for the definition of "best:" correlation of primary importances with consumer interest in product concepts, correlation between derived

satisfaction and stated satisfaction, and face validity. We describe each in turn.

"Stretching" Concepts

The RS created seven, one-paragraph product concepts such that each concept "stretched" one of the primary needs. The stretchers were written and tested carefully to ensure that they emphasized the target primary-need while emphasizing that the concept did not affect the other needs.

In the background information section of the questionnaire, prior to any importance ratings or need evaluations, the consumers were given each of the seven stretching concepts and asked to indicate their interest in the concepts on a nine-point interest scale. After seeing all of the concepts they were asked to rank them in terms of preference.

If the importances reflect consumer tradeoffs, then we would expect that consumers would be more interested (and show a higher preference for) the concepts that stretched important needs. Thus, one measure of "best" is the correlation of the rank order interest (preference) with the rank order of primary importances[2].

One potential concern with this measure of "best" is that primary need 5000 did not capture all of the corresponding secondary and tertiary needs. Fortunately, this concern applies equally to both the primary need and the corresponding stretcher. While neither the primary need nor the stretcher fully reflects the secondary and tertiary needs, the primary need and the stretcher are consistent in what they do reflect. The *statistical comparison* is appropriate because of this consistent correspondence.

Correlation Between Stated Satisfaction and Implied Satisfaction

QFD uses the importances as if they were linear weights that could be attached to the consumer needs. Thus, we can define a satisfaction index (SI) as a weighted combination of the importances and the consumer needs. Mathematically, for the primary needs, SI is given by equation 1.

$$SI = \sum_{i=1}^{7} I_i * E_i \qquad (1)$$

where I_i = importance of the i-th primary need

[2]We could report the actual correlation with interest but not with preference (preference data is ranking data). Because, for our data, the basic conclusions are the same for the correlation as for the rank correlation, we report here only the rank correlation.

E_i = evaluation of the product process with respect to the i-th primary need[3]

Similar equations apply for the secondary and the tertiary[4] consumer needs.

To get an understanding of equation 1 suppose that there were only two primary needs, need A and need B. Consider two product processes, one based on product 1 and one based on product 2. Suppose further that a consumer's evaluations and importances are given below.

PRIMARY NEED	EVALUATION OF PRODUCT 1 PROCESS	EVALUATION OF PRODUCT 2 PROCESS	IMPORTANCE
Consumer need A	2	1	3
Consumer need B	1	3	1

If we apply equation 1 we obtain,

PRODUCT 1 PROCESS: SI = 3 x 2 + 1 x 1 = 7
PRODUCT 2 PROCESS: SI = 3 x 1 + 1 x 3 = 6

If the satisfaction index (SI) were a valid measure of satisfaction, we would infer that the consumer should be more satisfied with the product 1-based process than with the product 2-based process.

In the data, as described later in this report, we face the issue of segmentation and the challenge of matching consumers who evaluated their product process with consumers who provided importances. However, the conceptual idea is the same. We use equation 1 (or its analog for secondary and tertiary needs) to compute a satisfaction index and we compare it to satisfaction as measured. (Review table 8[5].)

Unit-weights Satisfaction Index. In a unit-weights index we weight all needs equally. For the primary needs, SU is given by equation 2. (SU is called a unit weights index because equation 2 implicitly assumes that all importances are equal to 1.0.)

SU should be correlated with satisfaction. However, if the importances are valid indicators of consumer tradeoffs, then SI should correlate with measured satisfaction better than

---

[3] For ease of exposition we assume here that a better evaluation with respect to a primary need corresponds to a higher number. The actual data are coded such that a *lower* number corresponds to a better evaluation (a higher number means a greater need). Thus, for equation 1, we must either rescale the data by multiplying by (-1) or place a minus sign before the summation in equation 1.

[4] In some cases a secondary need had no tertiary elaborations. In this case we used the relevant secondary as if it were a tertiary. We ran it both ways (using the secondary and not using the secondary) with no discernable change in the conclusions.

[5] We also compare the satisfaction index to the 100-point satisfaction scale and to the like-or-dislike scale. There is some variation in the specific results, but the overall conclusion about the comparison of methods does not change.

$$SU=\sum_{i=1}^{7} E_i \qquad (2)$$

SU correlates with measured satisfaction. Thus, we judge importances (SI) relative to unit weights (SU).

Rescaling the importances. In concept the comparison of satisfaction and a satisfaction index is easy. But we face a technical issue of "rescaling." When we ask the consumer for importances, he can not know that we intend to use them in equation 1. He answers to the best of his ability, but we have no guarantee that her answer on a nine-, ten-, or 100-point scale matches our equation. To illustrate this point, suppose that he uses the upper-end of the scale, that is, he adds '2' to every importance. This does not change the relative order of the measured importances and it will not affect the correlation with stretching concepts. But it will affect equation 1. In particular, if we redo the example above we get:

PRIMARY NEED	EVALUATION OF PRODUCT 1 PROCESS	EVALUATION OF PRODUCT 2 PROCESS	IMPORTANCE+2
Consumer need A	2	1	5
Consumer need B	1	3	3

If we now apply equation 1 we obtain,

PRODUCT 1 PROCESS: SI = 5 x 2 + 3 x 1 = 11
PRODUCT 2 PROCESS: SI = 5 x 1 + 3 x 3 = 14

By rescaling we have reversed the order of the satisfaction index!

To avoid this rescaling problem, we allow the data to tell us the value of a rescaling constant rather than applying equation 1 blindly. To do this we estimate a rescaling constant to maximize the correlation between satisfaction and SI. Technically, we do this by regressing satisfaction on SI and SU. The rescaling constant is then the ratio of the regression weights[6]. This rescaling method has the added advantage that we can perform a statistical test to determine whether the importances add significantly to the correlation with measured satisfaction.

Face Validity

There are several sources of face validating information for the self-stated measures.

---

[6] If I_i' are the raw importances, then the rescaled importances, I_i, are given by $I_i = I_i' + \text{constant}$. If SI' corresponds to the raw importances and SI to the rescaled importances, then substitution into equations 1 and 2 gives SI = SI' + constant*SU. Thus, the correlation of satisfaction and SI equals the correlation based on the regression of SI' and SU on satisfaction.

1. Several major projects are under way at the RS that were initiated based on independent data. The preliminary results of these projects are consistent with the implications of the self-stated measures. (Some of these projects have reported preliminary market-place success.)
2. In a segmentation analysis (described later in this report) the RS used self-stated importance measures to classify consumers to segments. They then used key needs to identify consumers to participate in qualitative focus groups. The vast majority of the qualitative information supported and amplified the segmentation results. Thus, "real people" were identified to confirm the statistical-analysis-based hypotheses generated from the importances.

3. The RS has found that there is a matching of categories of products to the needs-based segments. (This matching is more difficult for brands.) The importances of needs within segments are consistent with the benefits delivered by the corresponding product type.

4. The ranking of tertiary needs is consistent with the RS's experience with delivering improved benefits that result in consumer purchase. This was particularly true of tertiary needs associated with primary need 5000. The RS believes these needs were ranked appropriately with the self-stated importances but not with the anchored and constant-sum measures.

DATA COLLECTION AND SCREENING

National representative samples of female product users were recruited via random digit dialing. Respondents who agreed to participate in the study were mailed a background-information questionnaire plus either an importance questionnaire (self-stated, anchored, or constant-sum) or an evaluation questionnaire. In addition to the description of their product process, the interest scales, and the stretching concepts, the background questionnaires obtained data on brands, habits and practices, and media habits. In total, 5600 questionnaires were mailed out -- 1400 for each cell. To encourage returns a $5 bill was enclosed with the questionnaires and respondents who returned completed questionnaires within a week were entered in a drawing for $100.

Subsequently, 1400 questionnaires (background plus anchored importances) were sent to names supplied by a national mail panel. The questionnaires were identical to those used in the random-digit samples. Response rates were very good for the random-digit sample (75-78%) and excellent for the panel sample (90%).

The data was screened by the RS's market research department. Basically, the following screening criteria were used to eliminate respondents:

Self-stated: More than 30 importances not provided, standard deviation of importance ratings below 0.6.

Anchored: More than 21 question sets (out of 36) with a rating greater than 10; or more than 10 question sets unanswered; or more than two question sets with all zeros; or more than 30 question sets with equal ratings; or primaries not rated; or at least 17 sets with one question unanswered; or gave a highest rating of less than 10 in at least 17 sets.

Constant Sum: Two or more questions not answered in every one of the 36 question sets; or any rating more than 100; or rated all needs the same in 25 or more question sets; or ratings totalled less than 80 or more than 120 in 10 or more question sets.

Evaluations: 60 or more needs not rated.

With these criteria, 94% of the self-stated, 88% of the anchored, 92% of the constant-sum, and 97% of the evaluation data were usable. Missing data were handled as follows:

Self-stated: Missing values were replaced by the respondent's mean rating where the mean is computed across all needs.

Anchored: If a complete question set was not answered, the mean rating for all qualified respondents was used. If at least one need in a question set was rated, a zero (0) was substituted for missing data.

Constant-sum: Same as for anchored except that ratings were normalized to add to 100.

Evaluations. Missing values were replaced by the midpoint of the rating scale[7].

As detailed in an internal RS report, these rules were likely to provide the least bias due to missing data while enabling us to use a sufficient portion of the returned questionnaires.

The RS also analyzed the data for convergent validity. The rank correlations between the random-digit and the national panel samples were 1.00 for the primary needs, 0.995 for the secondary needs, and 0.994 for the tertiary needs implying excellent test-retest validity across samples. These numbers suggest than any differences among methods (correlations well below 0.99) are real and not due to sampling error.

The rank correlation between the two hierarchy-based methods is quite high, 0.89 or better suggesting that while there are some differences, there is also a good amount of similarity[8]. There is less correlation between the self-stated scale (which is not based on the hierarchy) and the hierarchy-based scales, but the correlations are still quite reasonable. Much of these differences are due to the secondary and tertiary needs associated with primary need 5000. For example, need 5600 was ranked 3 of 36 secondary needs by the self-stated importances, but 22 of 36 and 25 of 36 by the anchored and constant-sum importances.

---

[7] This missing-data procedure is a standard RS procedure. An alternative approach would be to replace missing values with the respondent's mean rating. This decision should have minimal effect on the results.

[8] M.I.T. has also computed the actual correlations among the methods. The correlations have the same implications as the rank correlations.

PREDICTING CONSUMER INTEREST IN CONCEPTS

Table 7 Correlations Between Ranks of Mean Importances

	Self-stated	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

Table 8 reports the results of the analysis of interest and preference in the stretching concepts. The ranks reported are based on the averages across all usable respondents. All three of the direct measures of importances correlate well with both interest and satisfaction[9]. While the anchored and the constant-sum measures do slightly better on interest, we do not believe that this difference is significant because some of the rank differences are based on insignificant cardinal differences.

Revealed preference does poorly relative to the direct measures. While it does highlight primary need 1000 as the most important need, it identifies primary need 6000 as the next most important need -- a need for which consumers show neither interest nor preference. Overall, revealed preference importances actually exhibit a negative correlation with both interest and preference for the concepts. This negative correlation is probably due to both multi-collinearity[10] among the primary needs as well as the concerns with the validity of the measure of overall satisfaction upon which the revealed preference regression is based. See the Appendix for tables of the revealed preference results.

Based on table 8, we believe that at least at the primary level, all three direct measures of importance have reasonable validity in terms of predicting consumer interest in new-product concepts. We do not believe that the revealed-preference-based importances should be used to guide QFD.

SATISFACTION VS. THE SATISFACTION INDEX

The satisfaction index (equation 1) assumes that the better a product meets consumer needs, the more satisfied the consumers will be with the product. It further assumes that meeting or exceeding important consumer needs has a greater effect on satisfaction than meeting or exceeding less important consumer needs. Thus, the correlation between the satisfaction

---

[9] We obtain similar results if the correlations rather than the rank correlations are reported. For example, the correlations with concept-interest are 0.93 for self-stated, 0.91 for constant-sum, 0.91 for anchored, and -0.01 for revealed preference.

[10] The average correlation is 0.24 and 71% of the correlations are above 0.20.

Table 8 Comparison of Interest in Stretchers with Importances

	Interest	Preference	Self-St.	Anchored	Cons.-sum	Revealed
Primary need 1000	2	1	1	1	1	1
Primary need 2000	1	2	3	2	2	7
Primary need 3000	4	4	4	4	4	4
Primary need 4000	6	6	6	5	5	5
Primary need 5000	5	5	5	6	6	3
Primary need 6000	7	7	7	7	7	2
Primary need 7000	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

index and true satisfaction is one indicator of the validity of a set of importances. Of course any empirical comparison depends on the validity of our satisfaction measures.

Ideally, we would like to measure importances *and* evaluations for each respondent. However, in pre-tests such measurement proved to be infeasible for the typical consumer. Fortunately, the House of Quality uses a single set of importances (per segment) to represent the typical consumer in that segment. Thus, we compute averages based upon responses to the importance questionnaires and use those averages with respondent-specific data from the evaluation questionnaire. In other words, in equation 1, the importances (I_i) are segment (or overall) averages while the evaluations (E_i) are different for each respondent. Satisfaction is measured for each respondent. To provide a fair evaluation, we obtain importances from the analysis of 70% of the data, the calibration sample, and compute correlations based on the remaining 30% of the data, the holdout sample[11].

We report in this section the unsegmented results, that is, results based on importances obtained by averaging over all respondents in the calibration sample. We discuss segmented results in a later section. (The methodological conclusions do not change when different sets of importance values are used for different segments.) We also discuss in that section further issues dealing with the use of segment-based importances.

[11] The calibration and holdout samples are drawn randomly from the screened data. To check for inadvertent bias we compared key results to those obtained from alternative sets of calibration/holdout samples. There were no differences in the comparisons among methods even though the actual correlations varied due to sampling error.

Primary Needs

Table 9 reports results, based on the primary needs, for the holdout sample. The unit weights model does reasonably well indicating that consumers report that they are more satisfied with products which meet their needs and that they like them better. The correlation is slightly higher for liking than satisfaction suggesting either that the satisfaction index is a better indicator of liking than of satisfaction or calling into question the satisfaction measures.

Table 9 Comparison of Satisfaction Index and Satisfaction

MODEL	9-pt. Satisfaction	100-pt. Satisfaction	Liking
Unit Weights	0.37	0.29	0.48
Satisfaction Index			
Self-stated	0.37	0.30	0.48
Anchored	0.37	0.30	0.48
Constant-sum	0.37	0.30	0.48
Revealed Preference	0.39	0.28	0.47

None of the techniques for measuring importances do significantly better than unit weights for either satisfaction or for liking. This lack of discrimination is disappointing -- we had hoped to identify a "best" technique, but it is not surprising. Linear models such as equation 1 tend to be robust, especially when the consumer needs are intercorrelated as they are in our sample.

Secondary and Tertiary Needs

We obtain comparable results for the secondary and tertiary needs. See table 10 for results based on a holdout sample[12]. Results are similar to tables 9 and 10 when the importances are standardized[13].

---

[12] Revealed preference can not be applied (in this data) to the tertiary needs because of the severity of the multi-collinearity problem.

[13] In standardization, a mean and standard deviation are computed across all importance scales for each individual. By subtracting the mean and dividing by the standard deviation, the data are corrected for scale biases, that is, for the tendency of respondents to use the basic 9-, 10-, or 100-point scales differently. However, standardized data should not be used in equation 1 without rescaling because standardization leads to some importances with negative values. Negative importances are misleading with respect to equation 1.

Table 10 Comparison of Satisfaction Index and 9-point Satisfaction Scale

MODEL	Primary Needs	Secondary Needs	Tertiary Needs
Unit 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	–

Rescaling Constant

The theory of the rescaling constant is valid if the satisfaction index provides significant explanatory power above and beyond unit weights. This was not the case in tables 9 and 10 nor was it the case in the regressions based on the estimation sample. The t-statistic for the satisfaction index was not significant. For example, we obtain a scaling constant of 4.8, (0.1616/0.00336), by dividing the coefficient of SU by the coefficient of SI, but because the coefficient of SI was not significant, this scaling constant can not be used reliably[14].

ISSUES IN THE MEASUREMENT OF SATISFACTION

Tables 9 and 10 do not identify a "best" importance measurement methodology. Furthermore, while correlations in the range of 0.30 to 0.40 are positive and significant, they are not at the level of 0.60 to 0.90 that we have seen in proprietary studies in a durable goods industry. On the other hand, at least for primary needs, importances seem to correlate highly with consumer interest in new product concepts. Faced with this seeming inconsistency, we examined the satisfaction ratings themselves.

[14] For the record, the scaling constants were 9.1 for the secondary needs and 12.2 for the tertiary needs. However, in a result quite consistent with the counter-intuitive revealed-preference regressions (review table 8), SI had a negative sign in the regressions. This is just one more indictment of the monadic measure of satisfaction.

Satisfaction vs. "Primary-brand Share"

In table 11 we compare reported satisfaction to "primary-brand share" where primary-brand share[15] is based on the primary brand that the consumer used in her product process. Naturally, this is not a true share because consumers often use multiple brands, but it does provide an indicator. Because the data were based on open-ended self-reports, we could not assign respondents to specific product variations. The correlation of "primary-brand share" and reported satisfaction is 0.08 (not significant at the .05 level) suggesting that reported satisfaction is not an indicator of primary-brand share. For example, product 1 has the highest share, but is only eleventh in reported satisfaction. Consumers are most satisfied with product 9, a brand with only a 2.9% share. (The rank-order correlation is an insignificant -0.08.) It doesn't get better with the 100-point satisfaction and 9-point liking scales which have insignificant 0.13 and -0.14 correlations with "primary-brand share."

Table 11 Comparison of Primary-brand Choice and Satisfaction

BRAND	PRIMARY-BRAND SHARE (%)	9-POINT SATISFACTION
Brand 1	37.3	6.8
Brand 2	8.1	7.0
Brand 3	6.0	6.5
Brand 4	5.9	6.8
Brand 5	4.0	6.8
Brand 6	3.9	6.5
Brand 7	3.8	7.1
Brand 8	3.2	6.2
Brand 9	2.9	7.1
Brand 10	2.6	6.9
Brand 11	1.9	6.4
Brand 12	1.3	6.6
Brand 13	1.1	5.7
Brand 14	1.0	6.2
Brand 15	0.9	6.9
Brand 16	0.9	7.0
Brand 17	0.7	6.9
Brand 18	0.3	7.0

We can not explain this lack of correlation from the data at hand, but the hypothesis advanced earlier in this report is one potential explanation. That is, product 1 consumers rate product 1, product 2 users rate product 2, etc. Product 1 may be the product that satisfies product 1 users the most but such consumers may be more demanding and, hence, less satisfied with their product.

To understand this phenomenon imagine a salesman who is compensated on sales and on satisfaction. Suppose he is closing a sale with a customer who is very demanding, who will choose his product because it is the best product on the market, but who is unlikely to be satisfied fully with any product. If the salesman is rewarded only on sales he will seek the sale, but if he rewarded only on satisfaction he will forego the sale. (On the other hand if he is only rewarded on sales, he may not have the incentive to provide after sales service. It would depend upon the likelihood of repurchase, the discount rate, and the salesman's other opportunities.) The same phenomenon could happen in a consumer market. The RS might have a brand, say product 1, which attracts the marginal, but critical, consumer -- the switching consumer. Such a brand could be very profitable but have a low monadic satisfaction rating.

---

[15] The numbers in table 15 do not add to 100% because of other brands and because some respondents did not report a brand. These numbers are based on the self-stated questionnaire; we get similar methodological implications for the other four questionnaires.

Stated another way, is it better for the RS to attract 1 million extremely satisfied customers or 2 million extremely satisfied consumers *and* 2 million mildly satisfied consumers? The latter are more profitable (we hope), but the former would have a higher *average* monadic satisfaction score. The manner in which satisfaction is measured can have a major influence on product-development and marketing incentives.

Potential Scales

At this point in time I do not have the data with which to suggest a "best" satisfaction scale[16]. Monadic scales, such as were used in this study, have problems, but other scales have potential problems as well. Two concepts that show promise are:

- *Relative scales.* A relative scale asks a consumer to state her satisfaction *compared with other brands*. The comparison set could be the brands that are seriously considered or just familiar brands. I do not know. Such a relative scale would mitigate the salesman's dilemma, but it would not eliminate it. (The salesman might now avoid a sale to a customer who would give a low relative satisfaction score.) However, if product 1 were the best brand in a set of brands, it would have a high relative score even for critical consumers.

- *Share-weighted satisfaction.* The theoretical concept that is illustrated with the salesman's dilemma is that he chooses not to serve the marginal customer because the marginal customer lowers the average. We can overcome this dilemma if we always *add* the marginal customer rather than *average* the marginal customer. This suggests that we use the sum (not average) of satisfaction over all consumers or, in a sampling context, share times average satisfaction. I have no experience with this scale.

- *Repurchase intentions.* Some firms and some market-research suppliers are suggesting that repurchase intentions are good measures of satisfaction. I can not dispute this, but repurchase intentions fall prey to the salesman's dilemma as would any monadic scale.

- *Market share.* Clearly market share is important, but market share is dependent upon more than consumer satisfaction. For example, price-off promotions, trade deals, coupon drops, sampling campaigns, and other marketing actions all influence market share. One goal of a satisfaction measure is to separate the product-quality/product-design issues from these marketing actions. One would never want a satisfaction scale to be correlated perfectly with share, but I feel that correlations, such as those in table 11, are much too low.

---

[16] For example, Douglas R. Hausknecht, "Measurement Scales in Consumer Satisfaction/Dissatisfaction," *Consumer Satisfaction/Dissatisfaction & Complaining Behavior*, 3, pp. 1-11., reviews 34 satisfaction scales that have been used in the academic literature.

No scale is perfect. Each has its implications and each has its strengths and weaknesses, but I believe that the evidence is sufficiently compelling that the RS should consider seriously an experiment to compare their current measures of satisfaction with other measures such as relative satisfaction and share-weighted satisfaction.

A Relative Scale used by the RS

The RS does use a relative satisfaction scale for tracking purposes. The question asks for satisfaction compared to other brands. When computing averages, the RS weighs excellent by 100, very good by 75, good by 50, and fair by 25. Poor is given a weight of zero. Scores are reported for those consumers who have heard of the brand, year to date, and for those consumers who have used the brand in the last three months.

Table 12 Comparison of Satisfaction and Relative Satisfaction

BRAND	PRIMARY-BRAND SHARE	MONADIC SATISFACTION MEASURE	RELATIVE SATISFACTION (HEARD OF BRAND)	RELATIVE SATISFACTION (USED BRAND)
Brand 1	1	6	1	1
Brand 2	2	3	3	5
Brand 3	3	8	2	4
Brand 7	4	2	6	8
Brand 8	5	9	7	7
Brand 9	6	1	5	9
Brand 10	7	4	8	3
Brand 12	8	7	4	2
Brand 14	9	10	9	6
Brand 17	10	5	10	10
Rank Correlation		0.20	0.83	0.39
t-statistic		0.58	4.21	1.21

We must be careful when comparing the data from our study with the data from the RS's tracking surveys. However, we can get initial insight through a cautious comparison of the results. Table 12 compares rank share and rank satisfaction from our data to rank relative-satisfaction from the RS tracking data. The comparisons are limited to the ten brands that are reported in our data and in the tracking data.

For these ten brands, the relative satisfaction measure among consumers who have heard of the brand has a much larger (and significant) rank-order correlation with "primary-brand share" than does the monadic satisfaction measure. See "heard of brand" in table 12. We can now limit the sample to those consumers who have used the brand in the last three months and

compute average satisfaction among the limited sample (see "used brand" in table 12). For these consumers the relative-satisfaction measure correlates better with primary-brand share than does monadic satisfaction, but the correlation is not as high as that obtained for relative-satisfaction based on all consumers who have heard of the brand. Among consumers who have heard of the brand the correlation is significant at the 0.01 level; among consumers who have used the brand in the last 3 months the correlation is significant at the 0.15 level.

Table 12 highlights the dilemma in choosing an appropriate measure of satisfaction. If we were evaluating a product program we might prefer to base satisfaction on consumers who have used the brand rather than including only those that have heard of the brand. However, the used-brand sample is subject to the same criticism as the monadic measure -- it confounds people and 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 sample includes all consumers, those who used it and those who have not found it sufficiently attractive to use. It should not surprise us that the latter measure is more like a market share measure.

If I were evaluating the marketing program I might be tempted to use the heard-of sample since it is marketing's responsibility to inform consumers, even those who are not now users. On the other hand if I were evaluating the product program, I am not so sure. Arguments can be made either way. I might also recommend a sample based on those who have considered the brand. This sample would include some consumers who rejected the brand, but it would be limited to only those who had sufficient knowledge with which to judge the brand.

In summary, the means by which satisfaction is measured affects strongly the incentives given to the organization. This is an area of research that demands further study.

SEGMENTATION

Consumers vary in the importances that they place on different needs. To take this into consideration the RS used the self-stated data to identify 10 clusters of consumers who were homogeneous with respect to the importances they placed on the consumer needs. These clusters and the details of the methodology[17] used to identify them are contained in an internal RS report.

If there were no problem with the satisfaction measures, we would expect a better match between satisfaction and the satisfaction index if we perform the analysis by segment. (A segment-based analysis uses the importances as averaged over consumers in each of the ten

[17] Ward's hierarchical clustering algorithm was used to identify an initial set of ten clusters. The number of clusters was chosen based on variance explained and the interpretability of the resulting solutions. A K-means algorithm based on the hierarchical solutions was then used to adjust the assignment of respondents.

needs-based segments.)

However, in a segment-based analysis, it is a real challenge to match consumers who completed the evaluation questionnaires to the importance-based segments. Now that the RS has identified the segments based on the self-stated importances, they can use a "gear box" to classify consumers to segments. That is, in an evaluation questionnaire they can also ask the key importance questions that distinguish the segments. Based on the responses to these importance questions they can classify consumers to segments. This has already proven successful with qualitative focus groups as described earlier in this report. However, when the RS/M.I.T. study was undertaken we had not yet identified the consumer segments. A gear box was not feasible. As a surrogate, M.I.T. attempted to classify consumers to segments based on the demographic, usage-pattern, and media-habit variables that were included in the questionnaires. In total, there were 97 variables available. However, even with step-wise discriminant analysis, we were not able to obtain sufficient classification. In part because of this assignment problem, the segment-based satisfaction indices were not significantly better than unit-weight satisfaction indices. Perhaps future research with a well-developed gear box will make this feasible.

Discriminant Analysis

To compute a satisfaction index we must assign consumers who completed the evaluation questionnaire to the segments that are based on the self-stated importance questionnaire. To make these assignments we used discriminant analysis on the self-stated questionnaire data. The discriminant analysis attempts to identify a set of demographic, usage-pattern, and media-habit variables that can be used to assign consumers to one of the ten needs-based segments. In total, there were 97 variables available. By using all of the variables on a calibration data set we were able to correctly classify 57.4% of the respondents. But this ability did not hold up on a holdout sample; only 14.9% of the respondents were correctly classified. By using step-wise discriminant analysis we were able to identify a reduced set of 27 variables that classified 26.4% of a calibration sample and 20.6% of a holdout sample. (A random assignment would classify correctly 11.7%.) Because of this weak link, the correlations based on the segmentation analyses may appear artificially low. This should not be interpreted as an indictment of the segmentation analyses; we present no scientific evidence either way. However, in the interest of completeness and scientific documentation, we present the (null) results. Future research with a well-developed gear box or with measurements of importances *and* evaluations from the same consumers should improve these results.

Correlation of Satisfaction and the Satisfaction Index by Segment

The comparison is simple in concept. Based on the respondents who completed the self-stated questionnaire, we compute average importances for each segment. We use data on demographics/usage/media-habits with the discriminant function to classify to segments respondents who completed the evaluation questionnaire. This assignment tells us which set of

importances to use for each respondent. We use the appropriate set of importances with the respondents' evaluations to compute the satisfaction index in equation 1 and we correlate the satisfaction index with measured satisfaction. (We rescale as described earlier in this report.) The reported correlation is based on all respondents in the holdout sample.

Table 13 reports on a holdout sample the correlation between the 9-point satisfaction scale and the primary-needs satisfaction index. As with the unsegmented analyses none of the direct measures does better than unit weights. Revealed preference does do better, but suffers from the same collinearity problems cited above[18]. (The variation in the unit-weights correlations is due to the fact that different calibration-holdout samples were used for the segmented results. In fact, one interpretation of table 13 is that there is greater variation between samples than among methods.)

Table 13 Comparison of Segmented and Unsegmented Analysis

	Segmented	Unsegmented
Unit weights	0.39	0.37
Satisfaction Index		
Self-stated	0.39	0.37
Anchored	0.38	0.37
Constant-sum	0.39	0.37
Revealed Pref.	0.45	0.39

We obtain similar methodological results for the secondary and tertiary needs. That is, even for the segmented analyses, the importance methods do not do better than unit weights. However, one weak link in our segmented analyses is the assignment procedure which assigns but 20% of the consumers to segments correctly. Thus, we must use extreme caution when interpreting these segmented results.

Reverse Engineering

To overcome the difficulties imposed by the inability of the discriminant analysis to classify respondents we looked at the data via "reverse engineering." That is, we used the data from the evaluation questionnaire to compute average evaluations for each brand. For example, we identified all consumers who reported using a product 1-based laundry process and averaged their evaluations. These evaluations were assigned to product 1. Then, for each consumer who completed the self-stated questionnaire, we identified the product upon which their process was based and assumed that they would evaluate their process in the same manner as the average consumer using a process based on that product. With reverse engineering there is no problem

[18] When used on the segments revealed preference does much better than when used on the unsegmented sample. However, there are still some problems with interpretation, there are many insignificant coefficients, and the coefficients do not match extremely well with the segment importances (based on the self-stated data). See table A2 in the appendix. Furthermore, revealed preference can only be used on the segments *after* they are identified by other means.

with assignment because each consumer reports the product upon which their process is based. But there are other problems. First, consumers are evaluating a process not a product and processes vary by more than just the brand of detergent used most often. Second, consumers do vary in their evaluations and we miss this variation by using averages.

The problems inherent in reverse engineering proved to be worse than those introduced by the discriminant analysis. As a result, the correlations between satisfaction and the satisfaction index were much less for reverse engineering than for the "forward" analyses. As a result, we did not pursue reverse engineering beyond the self-stated questionnaire.

RECOMMENDATIONS

Our initial goals in this project were to determine whether we could place confidence in the measures of importance and, if possible, to determine a "best" method. Based on the analysis of this data we can conclude that:

1. *For the primary needs, each of the three direct measures of importances (self-stated, anchored, and constant-sum) predicts consumer interest in and preference for product concepts. The indirect importance measure (revealed preference) does not predict interest and preference well.*

2. *The self-stated importances have high face validity. Their implications are consistent with the beliefs of the RS and are consistent with other information that the RS has on the product category.*

3. *The face-validity of the anchored and the constant-sum importances depend upon the ability of the RS to identify representative labels for the primary and secondary consumer needs.*

4. *None of the measures (direct or indirect) appear to form satisfaction indices that correlate with measured satisfaction better than a satisfaction index based on simply summing the consumer evaluations. However, this negative result may be due to concerns with the validity of the satisfaction measure.*

5. *The monadic measure of satisfaction may be misleading because it confounds variation in consumer needs with variation among consumers in the brands they use. Initial data on a relative measure of satisfaction is encouraging and further experimentation is recommended.*

Based on this data we would be comfortable with using the self-stated direct measures in QFD. We feel that the face validity and the ability to predict interest and preference in product concepts is sufficiently compelling to justify their use. We do not see the inability to

correlate with the satisfaction index as damning to the methods because of the potential problems with the satisfaction measures. While the statistical data provide no scientific basis to favor one of the direct measures over the others, the RS feels that face-validity issues recommend the self-stated measures.

Appendix

REVEALED PREFERENCE

Revealed preference importance measures are obtained by regressing consumers' evaluations of their product process on the direct measures of satisfaction. Table A1 provides the estimated importances for the primary needs. Different columns correspond to the different dependent measures. For ease of interpretation, table A1 displays only those importances that are significant at the 0.05 level or better. The t-statistics are shown in parentheses next to the importances. Table A2 reports the importances for the ten segments. Only those significant at the 0.10 level are reported. The t-statistics are available upon request.

Table A3 reports the importances (and t-statistics) for the secondary needs. Most are not significant due to the collinearity problems discussed in the text. It was not feasible to estimate revealed importances for the tertiary needs due to the large number (155) of such needs.

Table A1 Revealed Preference for Primary Needs

	DEPENDENT MEASURE		
	9-pt. Satisfact.	100-pt. Satisfact.	9-pt. Liking
Primary Need			
PRIMARY NEED 1	0.34 (5.55)	3.64 (5.13)	0.28 (4.99)
PRIMARY NEED 2		-1.38 (-2.65)	
PRIMARY NEED 3			
PRIMARY NEED 4			
PRIMARY NEED 5	0.17 (3.73)	2.03 (3.73)	0.08 (1.85)
PRIMARY NEED 6	0.25 (5.34)	2.65 (4.91)	0.30 (7.14)
PRIMARY NEED 7			
R-square	0.21	0.18	0.23

Table A2 Revealed Preference Analysis by Segment

Segments:	1	2	3	4	5	6	7	8	9	10
Primary Need										
PRIMARY NEED 1	0.41	0.42		0.28			0.33	0.40		
PRIMARY NEED 2			1.85				-0.24			
PRIMARY NEED 3					0.28					0.49
PRIMARY NEED 4							0.32		0.32	
PRIMARY NEED 5	0.18	0.21			0.66			0.20	0.30	
PRIMARY NEED 6	0.28			0.46			0.31		0.40	0.40
PRIMARY NEED 7						-0.38				-0.21
R-square	0.25	0.14	0.96	0.37	0.27	0.55	0.23	0.22	0.26	0.27

Table A3 Revealed Preference for Secondary Needs

Secondary Need	9–pt. Satisfaction	100–pt. Satisfaction	9–pt. Liking
SECONDARY NEED 1			0.08 (1.76)
SECONDARY NEED 2			
SECONDARY NEED 3			
SECONDARY NEED 4	0.15 (2.15)		
SECONDARY NEED 5			
SECONDARY NEED 6			
SECONDARY NEED 7			
SECONDARY NEED 8			
SECONDARY NEED 9			
SECONDARY NEED 10			
SECONDARY NEED 11			
SECONDARY NEED 12	-0.18 (-4.19)	-1.08 (-2.09)	
SECONDARY NEED 13	0.12 (1.76)	1.52 (1.94)	
SECONDARY NEED 14			
SECONDARY NEED 15			
SECONDARY NEED 16	0.12 (2.75)	1.94 (3.79)	0.22 (5.48)
SECONDARY NEED 17			
SECONDARY NEED 18			
SECONDARY NEED 19			
SECONDARY NEED 20	0.76 (1.68)	1.03 (1.91)	0.16 (3.69)
SECONDARY NEED 21			
SECONDARY NEED 22			0.07 (1.68)
SECONDARY NEED 23			
SECONDARY NEED 24			
SECONDARY NEED 25			
SECONDARY NEED 26			
SECONDARY NEED 27			
SECONDARY NEED 28			
SECONDARY NEED 29	0.16 (3.89)	1.96 (3.83)	0.07 (1.70)
SECONDARY NEED 30		1.70 (2.33)	
SECONDARY NEED 31			
SECONDARY NEED 32	0.14 (2.45)		0.17 (3.24)
SECONDARY NEED 33	0.25 (2.78)		
SECONDARY NEED 34			
SECONDARY NEED 35			
SECONDARY NEED 36			
R–square	0.32	0.23	0.27