RENEWING MARKET SEGMENTATION

Some new tools to correct old problems

Steve Cohen
Paul Markowitz

The intent of this paper is to present practicing researchers with an innovative use of state-of-the-art tools to solve problems that are too often glossed-over. This paper critically examines the usual and standard tools and methods of benefit segmentation as practiced by most researchers and suggests new ways of renewing market segmentation through methods of measuring benefit importance and performing segmentation analysis on the resulting data.

The authors introduce Maximum Difference Scaling, a more powerful method for measuring benefit importance that is *scale-free* and thus very applicable to international segmentation research. The paper describes how Maximum Difference Scaling can be combined with Latent Class Analysis to obtain international benefit segments. Finally the paper describes an example of how these methods were used in a cross-national business-to-business study.

Introduction

Segmentation continues to be one of the most powerful tools available to marketers to understand markets better and to deliver the right offers to customers at the right price. In a complex world, increased competition and market fragmentation has led to more product variety. Multiple product offerings thus lead to multiple marketing programs. Designing and managing these programs requires logic and skill. Deft implementation of this logic allows the manufacturer to recognize and cater to differences in customer needs. And if the marketer does so successfully, greater value will be created for the customer and for the manufacturer.

Many customers have shown that they are willing to pay more for products and services that better meet their needs. If priced and delivered properly, these higher prices should lead to higher profits for the manufacturer. Furthermore, well-marketed products and services will reduce competitive pressure and will provide the manufacturer some insulation from price wars.

Since the mid-1950s, marketing researchers have responded to the needs of management by conducting market segmentation studies. These studies are characterized by the collection of descriptive information about benefits sought, attitudes and beliefs about the category, purchase volume, buying styles, channels used, self, family, or company demographics, and so on. When designing a market segmentation study, the researcher chooses to look at the data through the lens of a *segmentation basis*. A *basis* is the effectively the organizing principle by which the information will be analyzed and presented to management.

Figure 1
A TAXONOMY OF SEGMENTATION BASES

	General Items	Category or Product-specific Items
Observable Characteristics	Geography Company size	Purchase volume Product & brand usage
Unobservable Characteristics	Psychographics Buying style	Benefits sought Brand perceptions Marketing elasticities

Segmentation bases are divided along two main axes (Kamakura and Wedel, 1999). The first axis defines whether the basis is a general basis or a product- or a category-specific basis. The second axis defines whether the basis is, on the one hand, easily observable, or, on the other, if it is unobserved, hard to measure, and thus must be uncovered during a *post hoc* analysis of the segmentation data. (See figure 1.)

The crossing of these two axes leads to a four-cell table. Observable general items are things like self and family demographics, or easily obtainable characteristics of the company under study. Unobservable general items include attitudes that span product categories, like psychographics, general buying styles, and the like. Category-specific observables relate to product and brand usage, purchase volumes, and usage occasions. The final cell is unobservable category-specific items. Included here are brand perceptions, marketing elasticities, and benefits sought from the category.

Alas, both practicing and academic researchers have often found that preexisting segments – the ones based on observables – when different, are well-distinguished in obvious ways and not much else (Kamakura and Wedel, 1999). Wealthier consumers purchase more goods and services, women purchase and use certain product categories more than men, smaller companies are more frugal in their spending patterns, families with children purchase and consume differently than those without. When looking however at buying motivations, benefits sought, and their sensitivity to the tools of marketers (e.g. price, promotions, channel strategies), members of preexisting groups are often found to be indistinguishable from one another.

Measuring and segmenting buyers using the bases in the fourth cell – unobservable category-specific bases – has attracted the attention of academic and practicing marketers, and with good reason.

In fact, this lens of segmentation has been endorsed by marketing strategists:

> *"If there is a 'most useful' segmentation variable, it would be benefits sought from a product, because the selection of benefits can determine a total business strategy."* (Aaker, 1995, p.52).

Using Aaker's prescription as a starting point, this paper critically examines the usual and standard tools and methods of benefit segmentation as practiced by most researchers and finds them sorely lacking. We suggest new ways of renewing market segmentation through methods of measuring benefit importance and performing segmentation analysis on the resulting data.

This paper is organized as follows. We first review the standard practices of benefit segmentation and, along the way, point out their deficiencies. We then introduce the reader to Maximum Difference Scaling, a method that we believe is a much more powerful method for measuring benefit importance – a method that is *scale-free* and thus very applicable to international segmentation research. The next section describes how Maximum Difference Scaling can be combined with Latent Class Analysis to obtain international benefit segments. We then describe an example of how these methods were used in a cross-national business-to-business study. We conclude with some final thoughts and suggestions for use.

STANDARD SEGMENTATION METHODS

We suspect that we would not be exaggerating if we stated that over three-fourths, if not more, of benefit segmentation studies use Factor Analysis of rating scales followed by a Cluster Analysis of the factor scores. This two-stage or tandem method has been popular for over twenty years (Haley, 1985). In short, this practice may be summarized in four steps (Myers, 1996):

- Administer a battery of rating-scale items to a group of consumer, buyers, customers, etc. These rating scales typically take the form of agree - disagree, describes - does not describe, important - not important ratings.
- The analyst then seeks to reduce the data to some smaller number of dimensions. Factor Analysis of the rating scale data, using either the raw ratings or some transformation of the ratings to obtain better statistical properties, is most often performed. The analyst then outputs the factor scores, one set of scores for each respondent. Some analysts will first use Correspondence Analysis rather than Factor Analysis, but the basic data reduction idea is the same.
- The factor scores then serve as the inputs to a Cluster Analysis, with the k-means Cluster Analysis method being the most preferred. Standard statistical routines such as SAS Fastclus or SPSS Quickcluster are used.
- The derived clusters will then be profiled. Essentially a cross-tabulation of group, cluster, or segment membership is created against all the other significant items in the survey. The most interesting, applicable, and important findings are reported to management.

What do we find lacking in these procedures? Quite a bit.

1. Despite our best efforts, rating scales have proven time and again to be non-discriminating. Many readers will recognize creating a chart of average scores on a five-point rating scale of each item on a survey that shows the mean ratings ranging from a high of 4.3 to a low of 3.5. While this may be satisfying to some, we contend that the amount of variation in these rating scales is too often low. This is not because the items are incorrect or have been poorly written. On the contrary, the content of the

items would have not been in the study in the first place if they were not important to investigate. It is just that rating scale usage does not permit people to distinguish among items. All too often – for whatever reason – people will not distinguish among items and so they will rate all items the same. At the other extreme, some people will bounce between the polar ends of a rating scale to indicate agreement or disagreement. Making more scale points, like ten-points or 100-points, just covers up the problem.

Furthermore, it is also well known that people in different countries, on average, use different parts of rating scales to the detriment of other parts (Baumgartner and Steenkamp, 2001). For example, South Americans and Southern Europeans tend to use the positive ends of scales. North Americans are likewise positive, but a little less so. Some say Germans are naysayers and the Japanese are middle-of-the-roaders. While it is possible to resort to the "post hoc" statistical methods described by ter Hofstede, Steenkamp, and Wedel (1999) or Rossi et al (1999), we would prefer to have a measuring instrument that allows easy cross-item comparison and yet does not require fancy statistical methods to untangle.

2. Rote use of Factor Analysis can also present problems. Most analyses that we have seen use Principal Components Factor Analysis followed by Varimax Rotation. Why these? That's the way we have always done it. While much academic research has extolled the benefits of these algorithms, their utility and efficacy has been investigated in controlled data conditions where the characteristics of the data are known. Some characteristics of ratings data are known beforehand. For example, the inputs for Factor Analysis are expected to be normally distributed variables. And yet the discussion in the prior paragraph makes it abundantly clear that ratings data are rarely normally distributed and are often skewed to the positive ends of the rating scale. And why don't we investigate solutions from Oblique factor rotations more often? We don't know why, but we are certain that it is the rare marketing researcher who does.

   Deriving patterns from Factor Analysis and making cross-country comparisons becomes problematic when scales have such built-in, systematic, scale use biases and correlations. We have found, for example, that when using a 5-point scale in a typical segmentation analysis, the first dimension in a Factor Analysis is often a general factor. Using this factor in a Cluster Analysis will often uncover a "high rater" segment or a "general" segment. Additional partitions of the data may uncover meaningful groups who have different needs, but only after separating out a group or two defined by their response patterns. This approach is especially dangerous in multi-country studies, where segments often break out on national lines, more often due to cultural differences in scale use, than to true differences in needs.

3. Cluster Analysis has its own set of problems in practice. With so many different clustering algorithms and options, which one should you use? There are many who claim to have rules of thumb for which to use, but the real unspoken fact is that, with the crush of client deliverables, few alternatives are explored.

   Compounding this issue is that some analysts will use Cluster Analysis inappropriately by entering nominal or ordinal variables into the algorithm. When needing these types of items in a Cluster Analysis, analysts will dummy-code or effects-code the item and expect that all will turn out well. To illustrate this problem, consider a three-level nominal item (coded as 1, 2, 3) that will be used, along with other items that will be fed into the clustering algorithm. The analyst knows that Cluster Analysis uses the distance between items to classify people. Because this is a nominal variable, the analyst takes the three-level item, creates two dummy-coded variables, and feels safe and secure that he has done well. But all is not well. By recoding the item, all he has done is to substitute one set of nominal scores (1, 2, 3) for another (the two dummy-coded items). If the model being estimated was a regression model, then dummy coding makes sense, since it has a meaningful interpretation, namely differences in category means as compared to the left-out category. But without a predictor-outcome relationship being estimated, the dummy-coded items are just nominal descriptors of the respondents, and no amount of recoding will transform nominal items into interval items whose differences imply distances.

   Finally, recent academic research has pointed out the deficiencies of the two-stage or tandem approach of Factor Analysis followed by Cluster Analysis [see DeSarbo et al (1990); Dillon, Mulani, & Frederick (1989); Green & Krieger (1995); Wedel & Kamakura (1999); and, Arabie & Hubert (1994)]. While the frequent use of the tandem method speaks eloquently to its ease of implementation with off-the-shelf software, most practicing researchers have failed to heed these warnings. The admonition against the tandem method cannot be said more bluntly than the disturbing assessment of Arabie & Hubert (1994):

   "Tandem clustering is an out-moded and statistically insupportable practice." (italics in original).

4. Finally, the analyst may often find himself or herself in a final quandary. Very often elegant benefits-based or attitude-based segments are well separated from one another on the benefits or attitudes, and yet often show

> little differentiation on descriptors that can be used for segment targeting. This point bears repeating. Even though the segments may pass the statistical test of within-homogeneity and between-separation when examining the basis variables, quite often benefit-segment membership does not predict key behaviors or characteristics of interest. Thus the segmentation results are rendered virtually useless to management.

The next sections describe the use of Maximum Difference Scaling to measure the *relative importance* of benefits and Latent Class Analysis for grouping people – uncovering market segments – with similar benefit importances. The issue of ensuring that the so-derived segments are predictive of key behaviors or characteristics of interest is discussed in the final section.

MAXIMUM DIFFERENCE SCALING

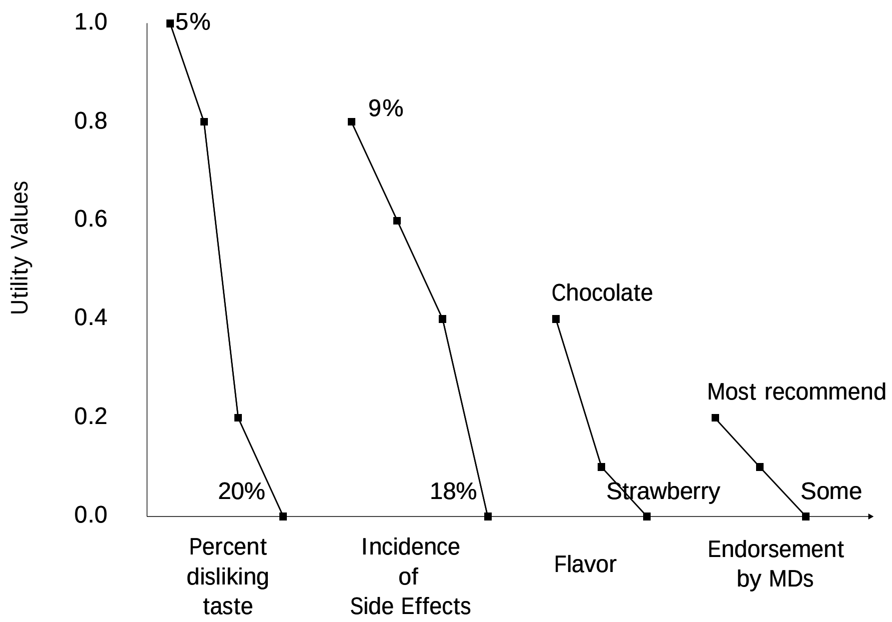
Maximum Difference Scaling (MaxDiff) is a measurement and scaling technique originally developed by Jordan Louviere and his colleagues (Louviere, 1991, 1992; Finn and Louviere, 1995; Louviere, Swait, and Anderson, 1995). Most of the prior applications of MaxDiff have been for use in Best-Worst Conjoint Analysis. We apply this scaling technique instead to the measurement of the importance of product benefits and uncovering segments. This is the first such discussion of the use of these methods for benefit segmentation that we are aware of in the marketing literature.

MaxDiff finds its genesis in a little-investigated deficiency of Conjoint Analysis. As discussed by Lynch (1985), additive conjoint models do not permit the separation of importance or weight and the scale value. Furthermore, with the utilities of the levels of each attribute being measured on separate scales with different origins, only ratio comparisons of differences are strictly permissible. Let us illustrate this problem concretely with the example in figure 2.

This figure shows the average utilities from a conjoint study of a new nutritional supplement. From these data, we would conclude that percent disliking taste is the most important attribute (its range is the largest) and doctor recommendation is the least important (smallest range). Within an attribute like flavor, chocolate is preferred to strawberry. This is the typical interpretation of the importance of items.

But let's look at the utility scale. All items have been coded such that their least popular level is equal to zero.[1)] But note however that an important comparison *cannot* be made. Comparing the zero level across attributes is impossible. You cannot tell whether "20% disliking the taste" is more or less important than "18% incidence of side effects." Thus, as Lynch says, the importance of product features has been confounded with the scaling of the items.

Figure 2
UTILITIES FROM STUDY OF NEW PHARMACEUTICAL



Utilities from Conjoint Model (best and worst levels labeled)

Maximum Difference Scaling permits intra- and inter-item comparison of levels by measuring attribute level utilities on a common, interval scale. Louviere, Swait, and Anderson (1995) present the basics of MaxDiff, or Best-Worst scaling. To implement maximum difference scaling for benefits requires these steps.

- Select a set of benefits to be investigated.
- Place the benefits into smaller subsets using an experimental design (e.g. 2^k, BIB, or PBIB are most common). Typically over a dozen such sets of three to six benefits each are needed.
- Present the sets one at a time to respondents. In each set, the respondent chooses the most salient or important attribute (the best) and the least important (the worst). This best-worst pair is *the pair* in that set that has the Maximum Difference.
- Since the data are simple choices, analyze the data with a multinomial logit or probit model. An aggregate level model will produce a total sample benefit ordering.

- Analyze pre-existing subgroups with the same statistical technique.
- To find benefit segments, use a Latent Class multinomial choice model (see below for discussion).

The MaxDiff model assumes that respondents behave as if they are examining every possible pair in each subset, and then they choose the most distinct pair as the best-worst, most-least, *maximum difference* pair.

Properly designed, MaxDiff will require respondents to make trade-offs among benefits. By doing so, we do not permit anyone to like or dislike all benefits. By definition, we force the relative importances out of the respondent. A well-designed task will control for order effects. Each respondent will see each item in the first, second, third, etc. position across benefit subsets. The design will also control for context effects: each item will be seen with every other item an equal number of times.

The MaxDiff procedure will produce a unidimensional interval-level scale of benefit importance based on nominal level choice data. As such, MaxDiff is particularly valuable in international segmentation research. Because there is only one way to choose something as “most important,” there is no opportunity to encounter bias in the use of the rating scale. Hence, there is no opportunity to be a constant high/low rater or a middle-of-the-roader. The method forces respondents to make a discriminating choice among the benefits. In our experience in over three dozen applications, the task is simple to complete, fast, easily understood by respondents, and, most importantly, it travels well across countries.

LATENT CLASS ANALYSIS

Combining MaxDiff derived utilities with a Latent Class choice model (DeSarbo, Ramaswamy, and Cohen, 1995; Cohen and Ramaswamy, 1998) leads easily to identifying segments with differing needs. All of this occurs in a scale-free and statistical model-based environment. For readers not familiar with Latent Class Analysis, we present this short description of its advantages. Interested readers are referred to Wedel and Kamakura (1999) for a more detailed discussion.

Latent Class Analysis (LCA) has a great deal in common with traditional Cluster Analysis, namely the extraction of several relatively homogeneous and yet separate groups of respondents from a heterogeneous set of data. What sets LCA apart from Cluster Analysis is its ability to accommodate both categorical and continuous data, as well as descriptive or predictive models, all in a common framework. Unlike Cluster Analysis, which is data-driven and model-free, LCA is model-based, true to the measurement level of the data, and can yield results which are a leap ahead in terms of the explanation of buyer behavior.

The major advantages of LCA include:

- Conversion of the data to a metric scale for distances is not necessary. LCA uses the data at their original level of measurement.
- LCAs can easily handle models with items at mixed levels of measurement. In Cluster Analysis, all data must be metric.
- LCA fits a statistical model to the data, allowing the use of tests and heuristics for model fit. The tandem method, in contrast, has two objectives, which may contradict one another: factor the items, then group the people.
- LCA can handle easily cases with missing data.
- Diagnostic information from LCA will tell you if you have overfit the data with your segmentation model. No such diagnostics exist in Cluster Analysis.
- Respondents are assigned to segments with a probability of membership, rather than with certainty as in Cluster Analysis. This allows further assessment of model fit and the identification of outliers or troublesome respondents.

Perhaps the biggest difference between Cluster Analysis and LCA is the types of problems they can be applied to. Cluster Analysis is solely a descriptive methodology. There is no independent-dependent, or predictor-outcome relationship assumed in the analysis. Thus, while LCA can also be used for descriptive segmentation, its big advantage lies in simultaneous segmentation and prediction. For example, an LCA prediction equation can be estimated –as in regression analysis – at the same time that the segments are uncovered. In the case of LCA regression, the segments consist of people whose regression coefficients (or conjoint part-worths) are relatively similar. So instead of having one aggregate regression equation describing the entire sample, a small number of equations capture several different predictor-outcome relationships – one equation for each latent segment.

If we think of a discrete choice model as another predictor-outcome relationship, then we can apply an LCA. Recognizing the need for conducting post hoc market segmentation with Choice-based Conjoint Analysis (CBCA), DeSarbo, Ramaswamy, and Cohen (1995) combined LCA with CBCA to introduce Latent Class CBCA, which permits the estimation of benefit segments with CBCA. LC-CBCA has been implemented commercially in a program from Sawtooth Software.

In our case, the predictor-outcome relationship is simple. Given the set of benefits in the subsets (the predictors), which items were selected as best and worst (the outcomes) in each?

To summarize this and the prior section:

- We advocate the use of Maximum Difference scaling to obtain a unidimensional interval-level scale of benefit importance. The task is easy to implement, easily understood by respondents and managers alike, and travels well across countries.
- To obtain benefit segments from these data, we advocate the use of Latent Class Analysis. LCA has numerous advantages over Cluster Analysis, the chief among them being that it will group people based on their pattern of nominal-level choices in several sets, rather than by estimating distances between respondents in an unknown or fabricated metric.

The next section discusses an empirical example of the application of these techniques.

AN EXAMPLE

Our client, a multinational company offering industrial products around the globe, wished to conduct a study of its global customers. The goal of the research was to identify key leverage points for new product design and marketing messaging. Previous segmentation studies had failed to find well-differentiated segments and thus the marketing managers and the researchers were amenable to the use of the techniques described above. For the sake of disguising the product category and the client, we present the category as file servers.

The survey was administered in the client‘s three largest markets: North America, Germany, and Japan. 843 decision-makers were recruited for an in-person interview: 336 in North America, 335 in Germany, and 172 in Japan. The questionnaire contained background information on the respondent’s company, their installed base of brands and products, and a trade-off task that examined new products, features, and prices. The benefit segmentation task is described next.

A list of thirteen product benefits was identified that covered a range of needs from product reliability to service and support to price. Prior qualitative research had identified these attributes as particularly desirable to server purchasers. The benefits tested were:

1. Brand name/vendor reputation
2. Product footprint
3. Expandability
4. Ease of maintenance and repair
5. Overall performance
6. Lowest purchase price
7. Redundant design
8. Reliability
9. Security features
10. Management tools
11. Technical support
12. Upgradeability
13. Warranty policy

A glossary was included with the survey so that respondents understood the meaning of each of these.

To develop the MaxDiff task, we used an experimental design. We created thirteen sets of four attributes each. Across the sets, every possible pair of items appeared together exactly once. Each benefit appeared once in each of the four positions in a set (first, second, third, and fourth). And, each benefit appeared exactly four times across the thirteen sets. When shown a set of four items, the respondents were asked to choose the item that was the most important and the least important when deciding which server to buy.

The results of the task are benefit utilities. In this study, the utilities for the benefits range from positive 3.5 to negative 3.5. We have found that looking at raw utilities may sometimes be unclear to managers. For ease of interpretation, we rescale the utilities according to the underlying choice model. Remember that the model estimated is a multinomial logit (MNL) model, where the sum of the choices after exponentiating is 100%. Hence, if we rescale the utilities according to the MNL model, we will get a "share of preference" for each benefit. If all benefits were equally preferred in this study, then each one's share of preference would be 7.7% (=1/13). If we index 7.7% to be 100, then a benefit with an index score of 200 would result from a share of preference of 15.4% (7.7% times 2). We have found that using this rescaling makes it much easier for managers and analysts to interpret the results. In this paper, we present only the index numbers and not the raw utilities.

By using the standard aggregate multinomial logit model, we obtained the results in table 1, after rescaling.

Table 1
OVERALL PRODUCT BENEFIT IMPORTANCES FROM MAXDIFF TASK

Reliability	571
Overall performance	277
Ease of maintenance and repair	84
Technical support	80
Expandability	59
Management tools	54
Upgradeability	50
Warranty policy	33
Brand name/reputation	27
Redundant design	24
Security features	27
Lowest purchase price	10
Product footprint	3

It is obvious that Product Reliability is the most important benefit followed by Overall performance. In this market, Lowest Purchase Price and Product Footprint are the least important items. Looking at the data by country, we find that the three countries follow the same pattern of importance.

Table 2
OVERALL PRODUCT BENEFIT IMPORTANCES BY COUNTRY

	North America	***Germany***	***Japan***
Reliability	579	410	421
Overall performance	272	192	264
Ease of maintenance and repair	72	114	75
Technical support	78	89	119
Expandability	63	77	57
Management tools	50	70	101
Upgradeability	48	69	62
Warranty policy	35	73	44
Brand name/reputation	46	31	39
Redundant design	31	41	23
Security features	15	95	67
Lowest purchase price	9	28	16
Product footprint	2	12	11

It is clear that there are only a few small differences by country. The overall picture of what is important in each marketplace is the same as found in the aggregate results.

We then conducted a segmentation analysis of the Maximum Difference data using the Latent Class Multinomial logit model. A six-segment solution was selected with the following segments emerging.

Table 3
OVERALL PRODUCT BENEFIT IMPORTANCES FROM MAXDIFF TASK BY BENEFIT SEGMENT

	Easy to buy and maintain	*Never breaks*	*Grows with me*	*Help me fix it*	*Brand's the clue*	*Managed and safe*
Reliability	**264**	**601**	**373**	**554**	**623**	**481**
Overall performance	**185**	**197**	**309**	**120**	**228**	**266**
Ease of maintenance and repair	100	33	71	**157**	23	51
Technical support	86	34	34	**305**	23	58
Expandability	81	30	**192**	33	21	30
Management tools	53	29	38	23	26	**190**
Upgradeability	58	12	**225**	16	10	31
Warranty policy	**100**	14	21	45	20	29
Brand name/reputation	45	28	10	20	**300**	7
Redundant design	56	**306**	11	16	8	10
Security features	31	12	10	6	10	**139**
Lowest purchase price	**213**	3	5	4	7	5
Product footprint	28	1	2	1	2	3
Percent of total sample	17%	11%	19%	14%	16%	24%
Percent of expected product purchases	31%	19%	23%	9%	9%	9%

In all segments, Reliability is the most important benefit, but its importance varies greatly from a low index number of 264 in the first segment to a high of 623 in the fifth. The second most important benefit is Overall performance, again ranging widely from 120 to 309. We would call these two benefits price of entry benefits in the server category. Respondents in all segments agree, in varying intensities, that Reliability and Performance are what a server is all about. Segment differences reveal themselves in the remaining benefits.

- Segment 1, *Easy to Buy and Maintain* (17% of sample and 31% of future purchases), values Lowest Purchase Price (213), Ease of Maintenance & Repair (100), and Warranty Policy (100).
- Segment 2, *Never Breaks* (11% of sample and 19% of future purchases), values Redundant Design (306) even more than Performance (197). They have a high need for uptime.
- Segment 3, *Grows with Me* (19% and 23%), values Upgradeability (225) and Expandability (192). They want to leverage their initial investment over time.
- Segment 4, *Help Me Fix It* (14% and 9%), values Technical Support (305) and Ease of Maintenance & Repair (157) even more than Performance.
- Segment 5, *Brand's the Clue* (16% and 9%), uses the Brand Name/Reputation (300) to help purchase highly reliable (623) servers. As the old saying goes, "No one ever got fired for buying IBM."
- Segment 6, *Managed and Safe (*24% and 9%), looks for Management Tools (190) and Security Features (139) when purchasing servers.

Note that Lowest Price, the second lowest index number overall is very important to the first segment, with an index score of 213. The benefits have very large variations across segments, indicating good between-segment differentiation. By looking at the number of servers expected to be purchased, we also provided guidance to management on which segments to target.

How do the three countries fall into the segments? Table 4 shows these results.

Table 4
INCIDENCE OF RESPONDENTS FROM EACH COUNTRY BY BENEFIT SEGMENT

	Total	***Easy to buy and maintain***	***Never breaks***	***Grows with me***	***Help me fix it***	***Brand's the clue***	***Managed and safe***
North America	40%	12%	65%	51%	59%	52%	17%
Germany	40%	70%	21%	30%	28%	18%	55%
Japan	20%	17%	14%	19%	13%	30%	28%

Relative to the total sample, respondents from Germany are overrepresented in two of the segments: Easy to Buy and Maintain, and Managed and Safe.

Respondents in Japan are overrepresented in Brand's the Clue and Managed and Safe.

From these data, we can conclude that:

- Looking just at country totals does not advance our understanding more than a three-country aggregate analysis.
- Six segments exist in the market for this product, with some segments being more prevalent in certain countries than others.
- Three of the segments – Easy to Buy and Maintain, Never Breaks, and Grows with Me – represent just under half of the respondents but almost three-fourths of expected purchases.

SUMMARY AND CONCLUDING THOUGHTS

The intent of this paper has been to present practicing researchers with an innovative use of state-of-the-art tools to solve problems that are too often glossed-over. Problems of attribute scaling, misuse of Factor Analysis, and blind use of Cluster Analysis have been described and criticized.

We suggest using Maximum Difference scaling for developing a unidimensional scale of benefit importance. The clever reader will realize that this tool can be used in any situation where there is a need to order a set of objects. The objects may be brands, products, positioning statements, product designs, product names, and so on. We have used the method with as few as nine objects and as many as 36. The tool is easy to implement, relatively easy to analyze with standard software, and easy to explain to respondents and managers alike.

To obtain benefit segments, we suggest using Latent Class Analysis. LCA has numerous advantages over Cluster Analysis. The disadvantages of this latter method are well known but not often heeded. The benefits of LCA are also well known, however its use is limited but growing. We hope that this paper will spur the frequent use of these two methods.

We alluded earlier to the fact that some segmentation results, while elegant and methodologically correct, may suffer from a lack of relationship to key behaviors and background variables of interest. Alas, the methods described here may also suffer the same failings. Our experience shows, however, by forcing the respondent to make benefit trade-offs, relationships with key exogenous variables are strong. For those interested in more rigorous methods of ensuring that these relationships are strong, we refer the reader to Krieger and Green (1996), Forsyth et al (1999), Brusco et al (2002), and Wedel and DeSarbo (2002).

To quote from Wedel and Kamakura (1999):

"The identification of market segments is highly dependent on the variables and methods used to define them."

We hope that we have shown that current research practice can be improved, so that the methods commonly used today have less of an effect on segment discovery and identification. By describing these methods, we trust that we have planted the seeds of a renewal in market segmentation methods.

FOOTNOTES

1. Other coding schemes will center the utility levels around zero, so that their sum is zero. Either way the point is the same.

REFERENCES

Aaker, David A. (1995). Strategic Market Management. New York: John Wiley & Sons.

Arabie, Phipps and Lawrence Hubert (1994). Cluster analysis in marketing research. *Advanced Methods of Marketing Research*. Richard J. Bagozzi (Ed.). London: Blackwell Publishers. pp. 160-189.

Baumgartner, Hans and Jan-Benedict E.M. Steenkamp (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research,* 38 (May).

Brusco, Michael J., J. Dennis Cradit, and Stephanie Stahl. (2002). A simulated annealing heuristic for a bicriterion partitioning problem in market segmentation. *Journal of Marketing Research*, 34, pp. 99-109.

Cohen, Steven H. and Venkatram Ramaswamy. (1998). Latent segmentation models. *Marketing Research Magazine*, Summer, pp. 15-22.

DeSarbo, Wayne S., Venkatram Ramaswamy, and Steven H. Cohen. (1995). Market segmentation with choice-based conjoint analysis. *Marketing Letters*, 6, 2, 137-47.

DeSarbo, Wayne S., Kamel Jedidi, Karen Cool, and Dan Schendel. (1990). Simultaneous multidimensional unfolding and cluster analysis: An investigation of strategic groups. *Marketing Letters,* 3, pp.129-146.

Dillon, William R., Narendra Mulani, and Donald G., Frederick. (1989). On the use of component scores in the presence of group structure. *Journal of Consumer Research*, 16, pp. 106-112.

Finn, Adam and Jordan J. Louviere. (1992). Determining the appropriate response to evidence of public concern: The case of food safety. *Journal of Public Policy and Marketing*, 11:1, pp. 19-25.

Forsyth, John, Sunil Gupta, Sudeep Haldar, Anil Kaul, and Keith Kettle. (1999). A segmentation you can act on. *McKinsey Quarterly*. No. 3, pp. 7-15.

Green, Paul E. and Abba Krieger. (1995). Alternative approaches to cluster-based market segmentation. *Journal of the Market Research Society,* 37:3, pp. 231-239.

Haley, Russell I. (1985). *Developing effective communications strategy: A benefit segmentation approach*. New York: John Wiley & Sons.

Krieger, Abba M. and Paul E. Green. (1996). Modifying cluster based segments to enhance agreement with an exogenous response variable. *Journal of Marketing Research*, 33, pp. 351-363.

Louviere, Jordan J. (1991). Best-worst scaling: A model for the largest difference judgments. Working paper. University of Alberta.

Louviere, J.J. (1992). Maximum difference conjoint: Theory, methods and cross-task comparisons with ratings-based and yes/no full profile conjoint. Unpublished Paper, Department of Marketing, Eccles School of Business, University of Utah, Salt Lake City.

Louviere, Jordan J., Joffre Swait, and Donald Anderson. (1995). Best-worst Conjoint: A new preference elicitation method to simultaneously identify overall attribute importance and attribute level partworths. Working paper. University of Florida.

Lynch, John G., Jr. (1985). Uniqueness issues in the decompositional modeling of multiattribute overall evaluations: An information integration perspective. *Journal of Marketing Research*, 22, pp.1-19.

Myers, James H. (1996). *Segmentation and positioning for strategic marketing decisions*. Chicago: American Marketing Association.

Rossi, Peter E., Zvi Gilula, and Greg M. Allenby. Overcoming Scale Usage Heterogeneity: A Bayesian Hierarchical Approach. (1999) Working paper, University of Chicago School of Business.

Ter Hofstede, Frenkel, Jan-Benedict E.M. Steenkamp, and Michel Wedel (1999). International Market Segmentation Based on Consumer-Product Relations. *Journal of Marketing Research*, 36, 1-17.

Wedel, Michel and Wagner Kamakura. (1999). *Market Segmentation: Conceptual and Methodological Foundations*. Dordrecht: Kluwer Academic Publishers.

Wedel, Michel and Wayne S. DeSarbo. (2002). Market segment derivation and profiling via a finite mixture model framework. *Marketing Letters*, 13:1, pp. 17-25.

THE AUTHORS

Steve Cohen is Consultant in Marketing and Research, United States.

Paul Markowitz is Senior Director, Marketing Science, Knowledge Networks, United States.