Marketing Science Institute Special Report 09-211

An Improved Method for the Quantitative Assessment of Customer Priorities

V. Seenu Srinivasan and Gordon A. Wyner

AN IMPROVED METHOD FOR THE QUANTITATIVE ASSESSMENT OF CUSTOMER PRIORITIES

V. "SEENU" SRINIVASAN
Stanford University

GORDON A. WYNER
Millward Brown International

INTRODUCTION

A number of marketing contexts require the quantitative assessment of customer priorities:

> Consider the introduction of a newer version of a software product. A number of features can potentially be added to the product. In order to determine which subset of features should be included, two types of information are required: (a) inputs from customers regarding the features they consider to be the most valuable, and (b) the development cost and time for incorporating the features.
>
> Consider enhancing a service such as a hotel or an airline. We need (a) customer input as to which service improvements are most valuable, and (b) the incremental cost of providing those improvements.

The empirical study reported here was done in cooperation with the Marketing Science Institute (hereafter, MSI), an organization that links marketing academics with marketing practitioners. MSI surveys its trustees to assess the priorities they place on different marketing topics such as "managing brands," "channels and retailing," and "new media." The priorities are then used by MSI in deciding which academic research proposals to fund and to inform marketing academics as to the topics on which research would be most useful.

Our focus in this paper is on the measuring the importances (or values) customers attach to a large (ten or more) number of topics (or features, or attributes). It is related to, but different from, conjoint analysis in which the measurement of values takes place over attributes *and the levels of those attributes*. Our focus is on the measurement of topic importances at the individual respondent level, rather than merely at the aggregate level. Individual level measurement permits cross-classifying importances against respondent descriptors, e.g., demographics, psychographics, and purchase behavior. It also enables the determination of benefit segments, i.e., clusters of respondents who are similar in terms of their importances for the topics.

In this paper, we compare the constant sum method (hereafter, CSUM) with a new method called ASEMAP (pronounced Ace-map, Adaptive Self-Explication of Multi-Attribute Preferences, Netzer and Srinivasan 2009) for the quantitative assessment of customer priorities. After describing the methods, we detail the research set-up to examine the relative performance of the two methods in the context of MSI's assessment of research priorities for different marketing topics by managers from MSI member companies. We then present the empirical results and offer our conclusions.

METHODS FOR MEASURING IMPORTANCE

One of the simplest ways of measuring the importance of topics is through the use of rating scales. For instance, respondents are asked to rate different topics on a 5-point rating scale, varying from "not at all important" to "extremely important." The main problem with such a rating task is that respondents tend to say every topic is important, thereby minimizing the variability across items. Netzer and Srinivasan (2009) compare ASEMAP to preference measurement methods based on rating scales in the context of conjoint analysis and conclude that ASEMAP produces a substantial and statistically significant improvement in predictive validity. Chrzan and Golovashkina (2006) compare six different methods for measuring importance and conclude that MaxDiff, CSUM, and Q-Sort are the preferred methods. (The other three methods are ratings, unbounded ratings, and magnitude estimation.) Q-Sort asks the respondent to give 10% of the items a rating of 5 (most important), 20% of the items a rating of 4, 40% a rating of 3, 20% a rating of 4, and 10% a rating of 1 (least important). Forcing such a distribution on every respondent is theoretically unappealing. In the present paper, we compare CSUM against ASEMAP. For many years, MSI has been employing CSUM to measure research priorities. Thus MSI provided a natural setting for a head-to-head comparison of CSUM against ASEMAP. At the very end of this paper, we briefly summarize the results from a study by Srinivasan and Makarevich (2009) comparing CSUM, ASEMAP, and MaxDiff.

THE CONSTANT SUM APPROACH (CSUM)

The data collection format for CSUM in the context of the MSI study is shown in Figure 1. The order of presentation of the topics is randomized across respondents. Because it is a web-based survey, the computer provides feedback regarding the current total. An error message appears when the respondent tries to move to the next screen, but the total is not equal to 100. The main advantage of CSUM is that it avoids the tendency of respondents mentioned earlier to state that everything is important. Also it recognizes the ratio-scaled nature of importance by asking the respondent to give twice as many points to one topic compared to another, if the first topic is twice as important. The CSUM approach works well when the number of topics is small; however, if the number of topics is large (say, ten or larger), respondents have great difficulty in allocating points across a large number of topics; they resort to simplifying tactics such as placing large round numbers for a few topics and zero (or blank) for the remaining.

FIGURE 1: CONSTANT SUM APPROACH (CSUM)

Please distribute a total of 100 points across the 15 research topics to indicate how important each topic is to you. Please allocate twice as many points to one topic compared to a second topic if you feel it is twice as important.

Innovation and new products	☐
Engaging customers	☐
…	☐
Driving loyalty	☐
Total	☐

ADAPTIVE SELF-EXPLICATION (ASEMAP)

The ASEMAP procedure starts with the respondent ranking a randomized list of topics from the most important to the least important. In case the number of topics is large, the ranking is facilitated by first categorizing the topics into two (or three) categories in terms of importance. For instance, in the MSI context, there were fifteen topics that were categorized into more important topics (8) and less important topics (7). The categorization step in shown in Figure 2. The respondent then ranks the items within each category by a "drag and drop" as shown in Figure 3. The end result of the above two steps is a rank order of the full list of topics.

FIGURE 2: ASEMAP CATEORIZATION STEP

From the list of 15 topics below, place a check mark on the eight topics that are most important to you:

☐	Emerging markets
☐	Engaging customers
☐	…
☐	Developing marketing competencies

FIGURE 3: ASEMAP – DRAG AND DROP FOR RANKING TOPICS

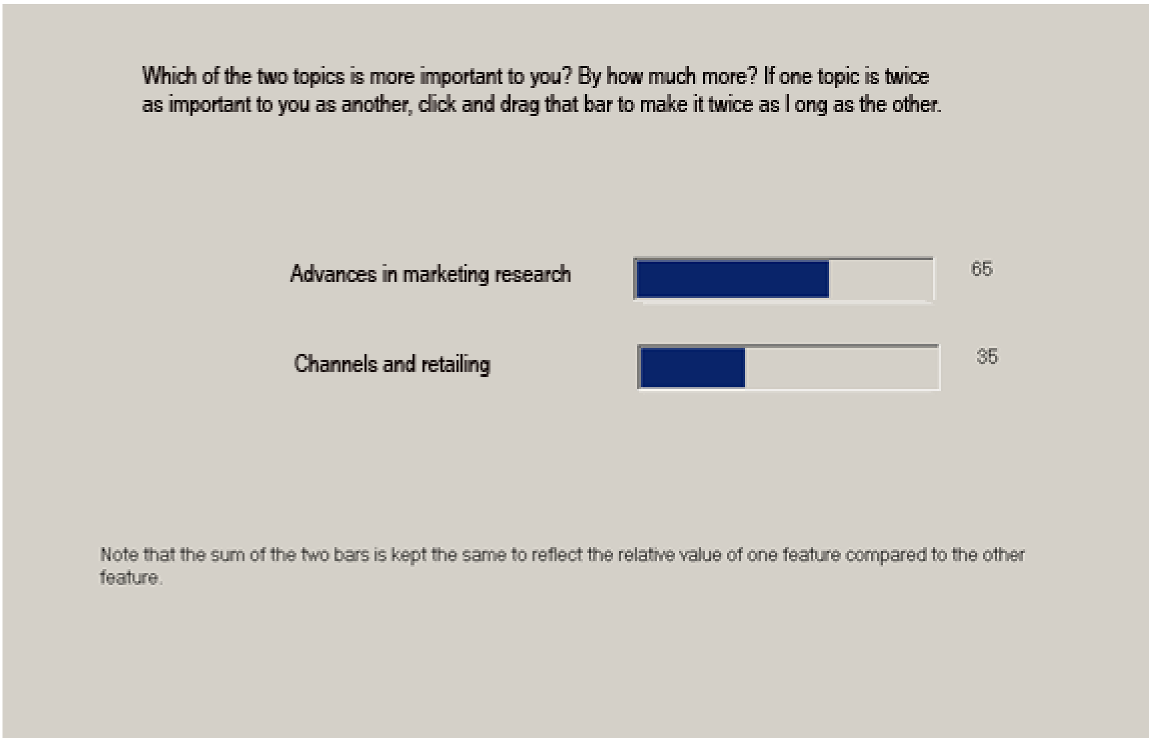
Please drag and drop the topics on this page so that they are ordered from the most important (on the top) to the least important (at the bottom).

Rank	Topic
1	Channels and retailing
2	Advances in marketing research techniques
…	
…	
8	Driving loyalty

The ranking task is insufficient from the point of view of assigning quantitative importances to the topics. ASEMAP determines the quantitative values by adaptively choosing a subset of topics (explained subsequently) and determining their importances by constant sum paired comparisons of two topics at a time. The quantitative vales for the remaining topics are determined by interpolation based on the rank order.

Figure 4 displays the ASEMAP screen for constant-sum paired comparisons. The two bars are initially equal in length (50:50 in terms of the numbers displayed on the right). As the respondent pulls one bar to the right (left), the other bar pulls itself to the left (right) so that the total length of the two bars remains the same, emphasizing the constant sum. The bars also provide visual cues in terms of how many times more important one topic is to another, thereby reinforcing the ratio-scaled nature of importance. For those respondents who are more quantitatively inclined, the numbers on the right provide the quantitative information (65:35 in Figure 4).

FIGURE 4: CONSTANT-SUM PAIRED COMPARISON MEASUREMENT OF TOPIC IMPORTANCE



Estimation of Weights

Suppose three among the full list of topics are A, B, and C, and the rank order among these three topics are A (most important), B, and C (least important). Suppose the respondent provides the following paired comparison data (ratio of bar lengths = ratio of the numbers on the right-side of the bars in Figure 4)

$W_A/W_B = 2$,

$W_B/W_C = 3$, and

$W_A/W_C = 5$.

Note that the redundancy among the three pairs provides information on the ratio-scaled consistency of the respondent's data. (The data would have been perfectly consistent had W_A/W_C been 6.) The ratio of the two weights would be undefined in the rare situation if the respondent had moved one bar all the way to the end (100:0). We recode such cases as (97.5: 2.5) because the allocations go in steps of 5, i.e., (85,15), (90,10), (95,5). Taking logarithms to the base 10, we obtain (taking the logarithms to any other base would not affect the obtained importances as long as they are normalized to sum to 100):

$\text{Log}(W_A) - \text{Log}(W_B) = \text{Log}(2) = 0.301$,

$\text{Log}(W_B) - \text{Log}(W_C) = \text{Log}(3) = 0.477$, and

$\text{Log}(W_A) - \text{Log}(W_C) = \text{Log}(5) = 0.699$.

To estimate the weights relative to the most important topic, we set $W_A = 100$ so that Log (W_A) = 2 and substituting this value in the previous three equations, we obtain

$- \text{Log}(W_B) = 0.301 - \text{Log}(W_A) = 0.301 - 2 = -1.699$

$\text{Log}(W_B) - \text{Log}(W_C) = 0.477$, and

$- \text{Log}(W_C) = 0.699 - \text{Log}(W_A) = 0.699\text{-}2 = -1.301$.

These three "observations" can be represented in the form of an OLS (ordinary least squares) multiple regression dataset:

$$\begin{array}{cc} \text{Log}(W_B) & \text{Log}(W_C) \end{array}$$
$$\begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & -1 \end{bmatrix} \qquad \begin{bmatrix} -1.699 \\ 0.477 \\ -1.301 \end{bmatrix}$$

The left hand side matrix shows three "observations" on two "independent dummy variables" whose regression coefficients are Log (W_B) and Log (W_C), respectively, and the right hand side column vector provides the values for the dependent variable. Performing an OLS multiple regression with no intercept, we obtain

$\text{Log}(W_B) = 1.725$ and $\text{Log}(W_C) = 1.275$.

Taking antilogs we obtain

$W_B = 10^{1.725} = 53.09$ and $W_C = 10^{1.275} = 18.84$.

It is possible to estimate the standard errors for the estimated weights by a Taylor series approximation (Netzer and Srinivasan 2009).

Suppose there actually were a total of six topics and the initial rank order provided by the respondent was (A, D, B, E, F, C). If we did not collect any paired comparisons in addition to the three paired comparisons stated earlier, we could infer the weights for D, E, and F by linear interpolation:

$W_D = (W_A+W_B)/2 = (100 + 53.09)/2 = 76.55$,

$W_E = W_B - (W_B\text{-}W_C)/3 = 53.09 - (53.09\text{-}18.84)/3 = 41.67$, and

$$W_F = W_B - 2\,(W_B\text{-}W_C)/3 = 53.09\text{-}2*(53.09\text{-}18.84)/3 = 30.26.$$

To normalize the weights so that they add to 100, we first compute

$$W_A + W_B + W_C + W_D + W_E + W_F = 320.41.$$

Multiplying all the weights by (100/320.41) we obtain the normalized weights

$$W'_A = 31.21,\ W'_B = 16.57,\ W'_C = 5.88,\ W'_D = 23.89,\ W'_E = 13.01, \text{ and } W'_F = 9.44,$$

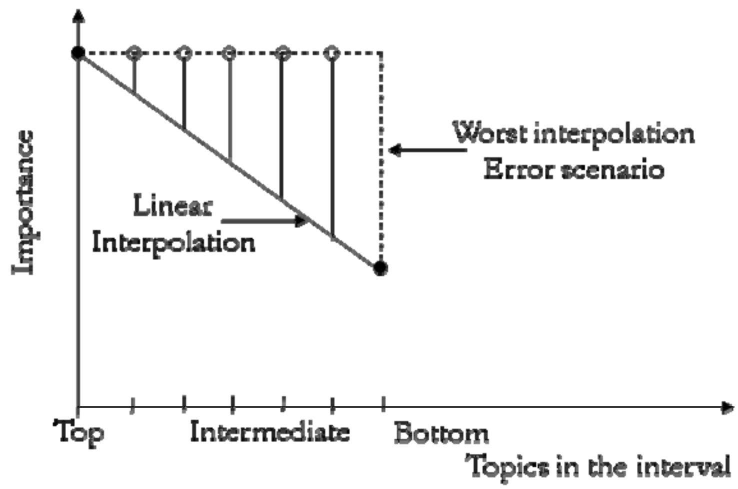
which add to 100. Note that $W'_A/W'_B = 1.88$, $W'_B/W'_C = 2.82$, and $W'_A/W'_C = 5.31$, which are close to the input data of 2, 3, and 5 for the paired comparisons.

Adaptive Measurement of Importances

In the example above, ASEMAP interpolates the importance weights for the intermediate topics not included in the paired comparisons. ASEMAP adaptively chooses the next topic to include in the paired comparisons so as to minimize the *maximum sum of interpolation errors* (hereafter denoted as MSIE). The "sum" in MSIE is computed over the interpolated topics. Because we do not know what the attribute weights will be for the intermediate topics had they been estimated, we take the "worst case scenario," i.e., the maximum sum of interpolation errors that can happen assuming that the rank order information is correct. The guiding notion is that ranking data are a lot more reliable than rating data (Krosnick 1999). (Empirical evidence indicates that only about 10% of the paired comparisons implied by the rank order get overturned in the paired comparison phase of data collection.)

We now illustrate the adaptive procedure for picking the next topic. In the MSI application there were 15 topics and we relabel the topics so that 1 denotes the most important topic, 8 denotes the topic with the middle rank in importance, and 15 denotes the least important topic. Because only interpolation is allowed and no extrapolation, we need to measure the importances of the most (#1) and least (#15) important topics. Figure 5 shows the general case in which we have a top attribute (at the start, it is topic #1), a bottom attribute (at the start, it is topic #15) for which the importances have been estimated, as shown by the solid dots in Figure 5. In the absence of any further paired comparison data, the importances of the intermediate topics will be obtained by linear interpolation (shown by the slanted line in Figure 5). (It can be shown that linear interpolation is no worse than any curvilinear interpolation in terms of MSIE.) The worst interpolation error scenario is shown by the open dots in Figure 5 in which the intermediate topics have importance equal to that of the top attribute. (The other worst case scenario in which the intermediate topics have importance equal to that of the bottom attribute will lead to the same value for MSIE.) It can be readily seen that the MSIE is approximated by the area of the triangle = difference in importance between the top and bottom topics times the number of intermediate topics/2. (This result can be shown to be an exact mathematical result, not merely a geometric approximation.)

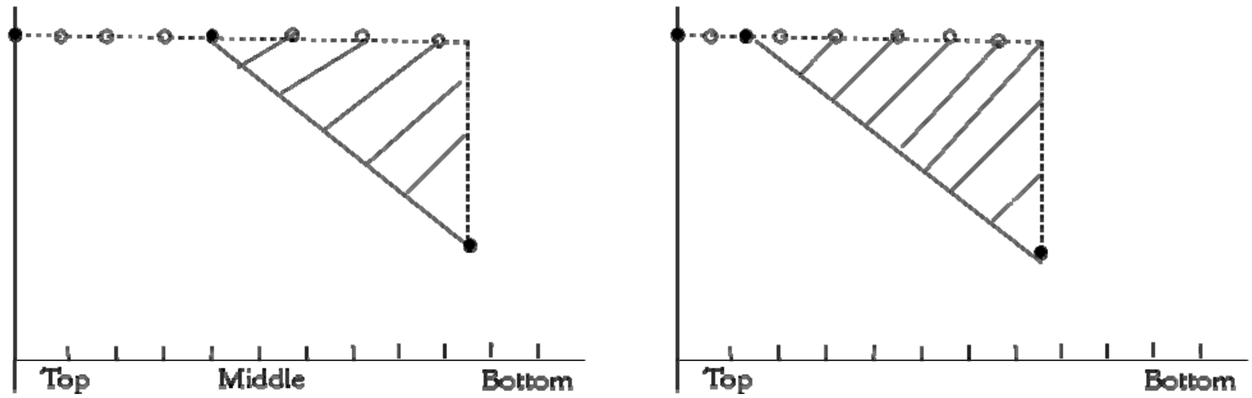
FIGURE 5: MAXIMUM SUM OF INTERPOLATION ERRORS (MSIE)



Sum of Interpolation Errors (MSIE) = [Difference in importance between the top and bottom topics x # of intermediate topics] ÷ 2

Where should we choose the next topic to estimate? A priori we do not know the importance of any intermediate topic except that it is bounded by the importance of the top and bottom topics, so we have to consider the worst possible scenario within these bounds. It can be shown that the choice should be the middle of the intermediate topics (at the start, it is topic #8). Figure 6(a) shows the worst case MSIE if the middle topic is chosen, and Figure 6(b) shows the worst case MSIE if some other intermediate topic is chosen. (The "worst case" refers to the scenario for the topic importance that would maximize MSIE. In case the number of intermediate topics is an even number, either of the middle topics can be chosen.) It is obvious that the MSIE in Figure 6(a) is smaller than that of Figure 6(b). It is also clear by comparing Figure 5 and Figure 6(a) that the MSIE in the interval is halved by estimating the middle topic. Consequently, in the general case (as will be illustrated later) in which there are multiple open intervals (i.e., intermediate topics bounded by top and bottom topics), we should chose (i) the interval with the largest value of the difference in importance between the top and bottom topic *times* the number of intermediate topics, and (ii) choose the middle topic in that interval as the next topic to estimate. It should be pointed out, however, this strategy only guarantees maximum reduction in MSIE *at each step*; i.e., it is not a globally optimal strategy considering the impact of any choice of topic on the future choice of topics. The latter would require the use of computationally time consuming dynamic programming, a luxury we cannot afford given our need to not make the respondent wait for computation to take its time.

FIGURE 6: CHOICE OF THE NEXT TOPIC TO INCLUDE IN ESTIMATION



a) Middle topic is chosen

b) Some other topic is chosen

MSIE = Maximum sum of interpolation errors is shown by the shaded area

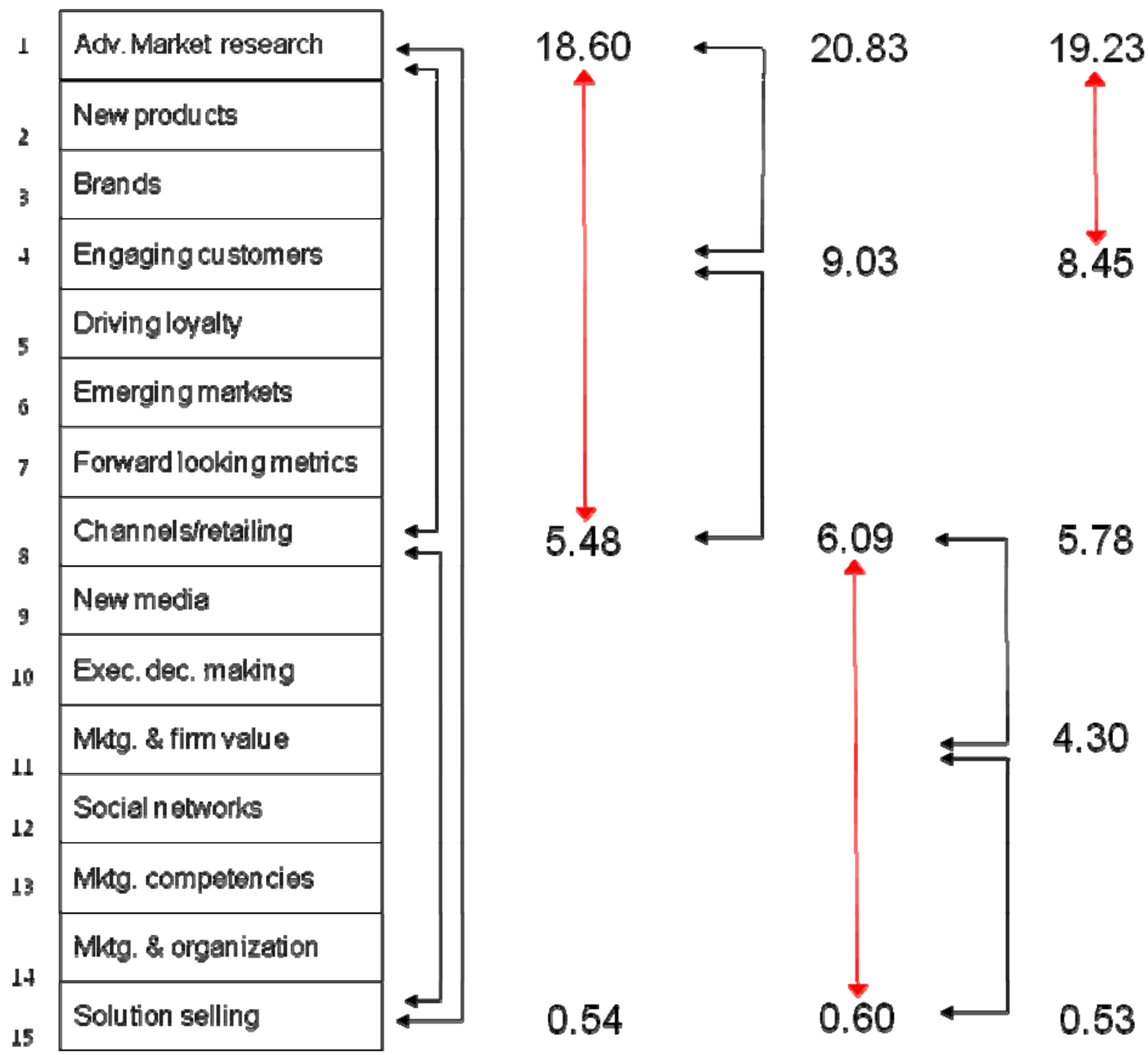
Having chosen an intermediate topic to estimate, the respondent evaluates two paired comparisons, the top topic compared with the middle, and the middle topic compared with the bottom. Although one of these paired comparisons is sufficient for the purpose of estimating the importance of the intermediate topic, we ask both questions so that the redundancy can inform us regarding the ratio-scaled consistency of the respondent, and also yield a more reliable estimate of standard errors. The total number of paired comparison (= number of observations in the multiple regression) is approximately twice the number of estimated parameters (importances for the topics). The ratio-scaled consistency of the respondent's paired comparisons is evaluated by the adjusted R^2 of the multiple regression. The procedure terminates when either a pre-specified number of paired comparison questions have already been asked, or if none of the differences in importance between the top and bottom topics of the intervals is statistically significant, as per a t-test (Netzer and Srinivasan 2009).

The ASEMAP approach is adaptive in that the paired comparison questions at each point in time are chosen based on the rank order information and the multiple regression-based importances estimated up to that point in time. The log-linear multiple regression computations are quick so that the respondent does not have to wait at the computer for the next paired comparison question to appear.

An Example

Figure 7 provides an example of the adaptive sequence of the paired comparison questions for one respondent. The rank order of the topics given by this respondent is shown on the left hand side of Figure 7.

FIGURE 7: EXAMPLE: ADAPTIVE SEQUENCE OF QUESTIONS FOR ONE RESPONDENT



1. We first ask the paired comparison comparing the top (most important) topic (#1) with the bottom topic (#15). As discussed in the previous section, the topic to be selected next is in the middle of the interval [1, 2, …, 15], i.e., #8, and we ask two questions comparing 1 with 8, and 8 with 15. These three paired comparisons are shown by the bracket-like double arrows. The log-liner multiple regression of the answers to these three questions yields the importances for #1 = 18.60, #8 = 5.48, and #15 = 0.54. These numbers are scaled in such a way that these numbers together with the interpolated values for all other topics sum to 100.

2. We now have two intervals [1, 2, …, 8] and [8, 9, …, 15]. The number of intermediate attributes in each of these two intervals is the same (=7) so that the

choice of the next topic is based only on the difference in importance between the top and bottom topics of the intervals. The larger differences corresponds to [1, 2, …, 8], so we open that interval (denoted by the vertical line with arrows at both ends) with it middle topic #4 (alternatively #5 could have been chosen). We ask two paired comparisons (1, 4) and (4, 8) again shown by the bracket-like double arrows. The answers to these two paired comparisons together with the previous three (a total of five) are analyzed by log-linear multiple regression to yield the results #1 = 20.83, #4= 9.03, #8 = 6.09, and #15 = 0.60. Note that the topic importance for 1, 8, and 15 have changed somewhat from the previous iteration. One reason is that the relative importance of 1 vs 8 is now determined by both that paired comparison and also by (1, 4) together with (4, 8).

3. We now have three intervals [1, 2, 3, 4], [4, 5, 6, 7, 8], and [8, 9,…,15]. We compute the quantity [difference in importance between the top and bottom topics times the number of intermediate topics] and find that we should choose [8, 9,…, 15]. We choose topic #11 as the next topic to evaluate and use two paired comparisons (8, 11) and (11, 15). The results of the log-linear regression of all seven paired comparisons are reported in the last column of Figure 7.

4. The procedure continues in this manner until a pre-specified number of paired comparisons (9 pairs in the MSI application) are asked.

ASEMAP Summary

To summarize, ASEMAP involves the following steps:

1. Divide the attributes into two (or three) sets.
2. Rank within each set.
3. The previous two steps results in an overall rank order.
4. Use the adaptive method to choose paired comparisons.
5. Estimate the importance of attributes included in the pairs by log-linear multiple regression.
6. The remaining attributes' importances are estimated by interpolation based on the rank order.
7. The importances are normalized to sum to 100.

EMPIRICAL STUDY

Selection of Topics

MSI had conducted focus group-like qualitative discussions with its trustees to identify research topics of significant interest to them. Based on the qualitative study, MSI staff assembled the list of fifteen major research topics in marketing, shown in Table 1. To elaborate on what each topic meant, MSI listed several examples of research projects subsumed by the topic label. To illustrate, the topic of "advances in marketing research techniques" had the following examples:

Internet-based marketing research

RFID-based research opportunities

Internet auctions

Text analysis of blogs

Cognitive science applications for marketing research

Permission-based marketing research

Use of virtual worlds in marketing research

Respondents

The respondents for the present empirical study were non-trustee managers from MSI member companies who had attended past MSI conference(s) and/or had requested MSI publications. E-mails were sent to managers requesting their participation with the MSI survey of research priorities; they were reminded up two times in case they did not respond to the earlier e-mail(s). The e-mail stated that MSI plans to use the results to better serve individuals in member companies. No other incentive was provided. The response rate was 17.2%. There were no statistically significant differences between respondents and non-respondents in terms of the proportions of companies (B to C, B to B, Services) represented.

The managers were e-mailed either the CSUM or ASEMAP web-based questionnaires (random assignment). In the ASEMAP survey the number of paired comparisons for each respondent was limited to nine pairs. Previous research (Netzer and Srinivasan 2009) has shown that ASEMAP's predictive validity does not increase appreciably as the number of paired comparisons is increased past nine pairs. The final sample sizes were n_{CSUM} = 161 and n_{ASEMAP} = 159. No statistically significant differences were found across the two samples in terms of the proportions of companies (B to C, B to B, or Services) represented.

TABLE 1: FIFTEEN MAJOR RESEARCH TOPICS IN MARKETING

Emerging markets

Engaging customers

Driving loyalty

Social networks and word-of mouth

Marketing and the organization

Developing marketing competencies

Improving executive decision making

Advances in marketing research techniques

Innovation and new products

Managing brands

Solutions selling

Channels and retailing

New media

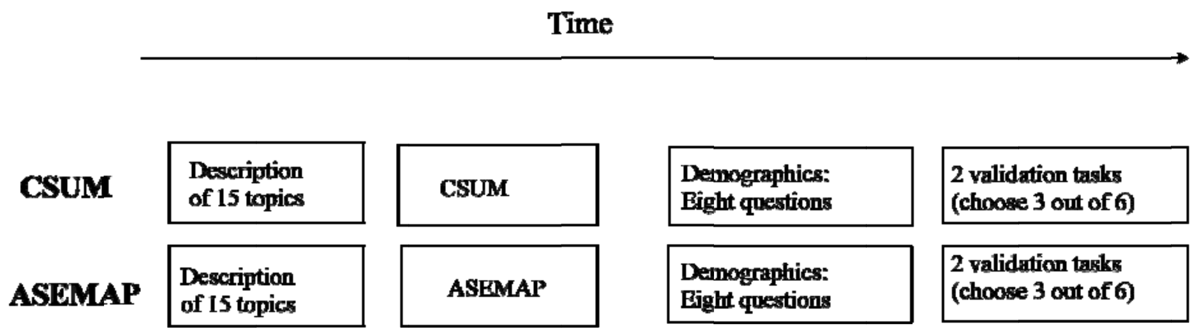
Forward-looking metrics

Marketing and firm value

Research Design

Figure 8 displays the research design employed in the study. After a description of the fifteen topics (in terms of examples of research projects for each topic), respondents prioritized the fifteen topics by CSUM or ASEMAP. This was followed by a set of eight demographic questions. The demographic questions also helped minimize short term memory effects from the measurement of importances by CSUM or ASEMAP to the "validation task," described next.

FIGURE 8: EMPIRICAL RESEARCH DESIGN



Validation

MSI funds a limited set of research proposals submitted by academics. We used that context to set up a validation task that would be meaningful to MSI. For each respondent, we chose randomly six out of the fifteen topics and asked the respondent to choose three out of the six topics, as shown in Figure 9. A second validation task with another random set of six topics followed.

FIGURE 9: VALIDATION TASK

Suppose MSI has received six proposals for research on the following topics. Suppose, however, that MSI can fund only three out of the six topics. Place a check (√) next to the three topics you want MSI to fund.

☐ Marketing and firm value

☐ Channels and retailing

☐ Driving loyalty

☐ Developing marketing competencies

☐ Managing brands

☐ New Media

We computed the extent to which the quantitative importances measured by CSUM or ASEMAP predicted the paired comparisons implied by the validation data. This served as a measure of predictive validity of the method. We illustrate below the computation of predictive validity measure. In the validation task the respondent chooses three out of the six topics. Suppose we denote by (A, B, C) the three topics chosen and (D, E, F) the three topics not chosen in a validation task. From the choices we can infer that topics A, B, and C should each have a higher importance than each of the topics D, E, and F. (Because the validation task is a simple task of choosing three out of six, we assume that the respondent will break ties, if any, in favor of the topic that is slightly more important.) From the validation task, we cannot infer anything about the relative importances of (A, B, C) among themselves, or among (D, E, F).

We use the percent of pairs correctly predicted as the measure of validity. Table 2 illustrates the calculations. Consider for instance, the pair comparing A and D. The validation data, i.e., A was chosen, but D was not chosen, implies A>D. The example preference data (measured by CSUM or ASEMAP) listed in Table 2 are A = 10 and D = 10. Because the preferences are tied, they do not predict A>D so it gets counted as incorrect (=0) as far as strict definition is concerned, but 0.5 if the weak definition is used (the prediction is neither correct, nor false). The percent correct is computed for both validation tasks (of choosing 3 topics out of 6), and the result averaged across the two tasks to yield the measure of predictive validity for the method (CSUM or ASEMAP) for that particular respondent.

RESULTS

Predictive Validity

Table 3 compares the predictive validity results for CSUM with ASEMAP. The percent of pairs correctly predicted was computed as described earlier for each of the respondents in the CSUM sample (n=161) and likewise for the respondents in the ASEMAP sample (n = 159). A two samples comparison test reveals that ASEMAP'S predictive validity is significantly larger than that of CSUM (p <.01) under both the strict and weak definitions. Furthermore, the differences are substantial. Under the strict definition ASEMAP is 36.9% better; even under the weak definition ASEMAP is 8.7% better. It is possible to improve ASEMAP'S predictive validity even higher by hierarchical Bayes methods (Netzer and Srinivasan 2009). (The hierarchical Bayes method is not applicable for the CSUM approach). We believe that the strict definition is more appropriate in applied marketing contexts because one reason for assessing the importance is to guide decision making in terms of which topic is more important. A naïve method that says all topics are equally important will obtain a 50% predictive validity under the weak definition, but will obtain only a 0% predictive validity under the strict definition. The constant sum method produces a large percentage of ties, 37% compared to 2.5% for ASEMAP. As expected, CSUM is much faster than ASEMAP. The CSUM questionnaire took on average, only 6.10 minutes compared to 8.97 minutes for the ASEMAP questionnaire (p<.01).

TABLE 2: CALCUATION OF THE RECENT OF PAIRS CORRECTLY PREDICTED

Let (A, B, C) be the topics chosen over (D, E, F) in the validation task. To illustrate the calculation of the percent by pairs correctly predicted, suppose

A = 10, B = 20, C = 0, D = 10, E = 15, and F = 0

are the importances measured by a method (CSUM or ASEMAP).

	Validation Pair	Do the Measured Preference Predict the Pairs? (1 = Yes, 0 = No, 0.5 =Tie)	
		Strict Defn.	Weak Defn.
	A>D	0	0.5
	A>E	0	0
	A>F	1	1
	B>D	1	1
	B>E	1	1
	B>F	1	1
	C>D	0	0
	C>E	0	0
	C>F	0	0.5
Total	9	4	5
Percent Correct		44.4%	55.5%

TABLE 3: AVERAGE PERCENT OF PAIRS CORRECTLY PREDICTED

	CSUM	ASEMAP	Z-stat
Strict defn.	59.6	81.6	9.23**
Weak defn.	75.9	82.5	3.9**

** The differences are substantial and highly statistically significant (p <.01)

Consistency of ASEMAP Data

The average adjusted R^2 of the ASEMAP log-linear regression was 0.93 indicating that the constant-sum paired comparison data were very good in terms of ratio-scaled consistency. Furthermore, the average rank order correlation coefficient between the initial rank order of the topics and final rank order of the topics (based on ASEMAP importances) was 0.90 showing that the paired comparisons were mostly consistent with the initial rank order. Unlike the above two statistics, there is no internal consistency measure for CSUM data.

Comparison of Average Topic Importances Across Methods

Table 4 reports the average topic importances for ASEMAP and CSUM. The topic importances are self-explanatory in terms of which topics were perceived to be more important and to what extent. On three of the fifteen topics the differences in means across the two methods are statistically significant ($p < .05$). The overall correlation between the two sets of means is 0.74.

TABLE 4: AVERAGE TOPIC IMPORTANCES FOR ASEMAP AND CSUM

Topics rearranged in decreasing order of ASEMAP-based average importance

Topic	ASEMAP Means	CSUM Means	
Innovation and new products	9.93	10.40	
Forward-looking metrics	**9.17**	**7.01**	($p < .05$)
Engaging customers	8.44	7.59	
Driving loyalty	8.06	7.14	
Adv. Marketing research techniques	**7.61**	**10.54**	($p < .05$)
Managing brands	7.02	6.94	
Improving executive decision making	**6.77**	**4.93**	($p < .01$)
Developing marketing competencies	5.87	5.78	
Channels and retailing	5.80	5.53	
Emerging markets	5.78	7.13	
Social networks and word-of-mouth	5.54	6.17	
Solutions selling	5.27	6.52	
New media	5.12	5.35	
Marketing and firm value	4.87	4.61	
Marketing and the organization	4.76	4.36	
Total	100.00	100.00	

Note: Statistically significant differences are highlighted.

Benefit Segments

We determined benefit segments based on the more valid ASEMAP importances. A benefit segment is a cluster of respondents who are close to each other in terms of topic importances. We used the k-means methods for cluster analysis and chose a three-segment solution based on the pseudo-F criterion. Table 5 provides the results.

Segment 1 (76.1% of the respondents) is a "customer-focused" segment that places more importance on "engaging customers" and "driving loyalty." Segment 2 (12.6% of the respondents) is an "innovation and new products" oriented segment. Segment 3 (11.3% of the respondents) emphasizes "advances in market research techniques" and "forward looking metrics." The demographic variables such as industry type and managerial position in the company do not, in general, discriminate significantly across the three segments except that segment 3 has a disproportionately large percentage of market researchers (78% compared to 44% for segments 1 and 2 combined, $p < .01$).

TABLE 5: MEAN TOPIC IMPORTANCES FOR THREE BENEFIT SEGMENTS

Segment #	**1**	**2**	**3**
% of respondents	76.1%	12.6%	11.3%
	Mean	**Mean**	**Mean**
Emerging markets	5.45	**8.93**	4.49
Engaging customers	**9.62**	6.08	3.21
Driving loyalty	**8.87**	5.19	5.80
Social networks and word-of-mouth	5.47	4.69	6.90
Marketing and the organziation	5.04	4.41	3.19
Developing marketing competencies	6.15	4.85	5.10
Improving executive decsion making	7.36	5.44	4.27
Adv. Markting research techniques	5.47	5.46	**24.38**
Innovation and new products	7.20	**30.00**	5.96
Managing brands	7.63	5.42	4.70
Solutions selling	5.70	4.45	3.29
Channels and retailing	6.61	2.93	3.60
New media	5.68	3.00	3.66
Forward-looking metrics	8.28	5.89	**18.76**
Marketing and firm value	5.47	3.26	2.69
Total	100.00	100.00	100.00

SUMMARY AND CONCLUSIONS

In this paper we compared the well known constant-sum method (CSUM) to a new method for measuring importances called ASEMAP, pronounced Ace-Map (Adaptive Self-Explication of Multi-Attribute Preferences). The ASEMAP method is particularly suitable when the number of topics for which importances need to be measured is large (≥ 10). The method involves the following steps for each respondent: (a) divide the topics into two or three categories (e.g., more important or less important), (b) drag and drop to rank the topics in each of the categories from the most-important to the least important; steps (a)-(b) together result in a total rank-order of all the topics; (c) adaptively choose constant-sum paired comparisons of a subset of topics and estimate their importances by log-linear multiple regression, and (d) estimate the importances of topics that were not included in the paired comparisons by interpolating their values based on the importances of the topics that were included in the paired comparisons and the initial rank order. At each iteration, the adaptive method chooses the next topic to include in the paired comparisons so as to minimize the maximum sum of interpolation errors.

The Marketing Science Institute, for a number of years has been determining its research priorities among marketing topics by polling their trustees using the constant sum method. The empirical study in this paper compared two random subsets of marketing managers from MSI member companies who provided information regarding importances of fifteen marketing topics, one group by the CSUM method and the other group by the ASEMAP method. The methods were compared on the basis of their ability to predict which three of a random subset of six topics the managers would choose as more important. ASEMAP produced a substantial and statistically significant improvement in validity, with the percent of correctly predicted pairs increasing from 60% for CSUM to 82% for ASEMAP. Even after giving tied predictions half-credit each (rather than zero credit), the percentage of correctly predicted pairs was higher for ASEMAP by 7 percentage points. Both these improvements are statistically significant at the .01 level. CSUM produced 37% tied pairs of importances compared to 2.5% for ASEMAP. The ASEMAP survey, however, took three minutes more, on average, compared to CSUM. The average importances across methods differed statistically significantly in three of the fifteen topics.

A Replication

The empirical results of this study were replicated in a second study on assessing priorities for the U.S. president, as seen by voters in July 2008 (Srinivasan and Makarevich 2009). A set of seventeen topics were included in the study with a random third of the respondents providing their priorities by (a) CSUM, (b) ASEMAP, and (c) MaxDiff (Sawtooth Software 2007). After the data collection regarding priorities and an intervening data collection on demographics and past political participation, the validation task involved a constant-sum task on two random subsets of four topics each. (The idea is that with a small subset of topics such as four, respondents would be able to provide importance information more accurately in CSUM.) The methods were compared on the basis of mean absolute error between the validation responses and rescaled importances from the original data collection by the three methods. As seen from Table 6, ASEMAP significantly ($p < .01$) and substantially out-performed CSUM and

MaxDiff in terms of both mean absolute error and its standard deviation (computed across respondents). CSUM took significantly shorter time than ASEMAP and MaxDiff. The difference in average times between MaxDiff and ASEMAP was not statistically significant.

TABLE 6: COMPARISON OF MEAN PREDICTION ERRORS AND TIMES ACROSS METHODS IN THE REPLICATION STUDY

Method	Mean absolute error	Improvement in mean absolute error over CSUM	Std. dev. in error	Average time (minutes)
CSUM	12.62	----	7.36	6.11
MAXDIFF	11.31	10.4%	4.91	8.63
ASEMAP	8.73	30.8%	4.29	9.38

Overall our research indicates that ASEMAP produces better predictive validity than CSUM and MaxDiff when the number of topics is large (in our studies the number of topics varied from 15 to 17).

Epilog

The Marketing Science Institute chose to use ASEMAP as the research method (rather than CSUM) for the subsequent research priorities study conducted with its trustees in 2008.

ACKNOWLEDGMENTS

Our thanks to Mike Hanssens, Russ Winer, Earl Taylor, Ross Rizley, Marni Clippinger, and Michelle Rainforth of the Marketing Science Institute, Andrew Latzman and Ronit Aviv of Dynamic Logic Inc., Linda Bethel, Ravi Pillai, and Jeannine Williams of Stanford University Graduate School of Business, Taylan Yildiz of Google, Inc., and Oded Netzer of Columbia University for their help with this study.

REFERENCES

Chrzan, Keith and Natalia Golovashkina (2006), "An Empirical Test of Six Stated Importance Measures" International of Market Research, 4:6, pp. 717-740.

Krosnick, Jon A. (1999), "Maximizing Questionnaire Quality," in John P. Robinson, Philip R. Shaver, and Lois S. Wrightsman (Eds.), *Measures of Political Attitudes*, New York: Academic Press.

Netzer, O. and V. Srinivasan (2009), "Adaptive Self-Explication of Multi-Attribute Preferences," Working paper, Stanford, CA: Graduate School of Business, Stanford University.

Sawtooth Software Inc. (2007), "The MaxDiff/Web System Technical Paper," Sequim, WA: Sawtooth Software, Inc.

Srinivasan, V. and Alex Makarevich (2009), "Assessing Presidential Priorities: A Comparison of Three Methods," Working paper (under preparation), Stanford, CA: Graduate School of Business, Stanford University.