The Determinants of Vendor Selection: The Evaluation Function Approach

YORAM WIND
PAUL E. GREEN
University of Pennsylvania

PATRICK J. ROBINSON
Marketing Science Institute

Why did vendor X and not vendor Y get a given order? What factors affected the buyer's vendor selection decision? What is the relative importance of each of these factors? These questions, and especially the latter, have been of major concern to industrial marketers.

Many students of industrial buying behavior have developed lists of factors that might affect the buyer's selection decision. Those who accept the assumptions of classical economic theory assume that industrial buyers' behavior is strongly motivated by rational economic considerations. The Industrial Marketing Committee Review Board of the American Marketing Association, for example, stated:

> Rational buying motives appear to predominate in the industrial field (as against the emotional motives in the consumer field) but their influence declines with the increase in product similarity.[1]

Other assertions of this type, as well as some supporting empirical evidence, have led to an increasing recognition that industrial buyers can be influenced by emotional (non-economic) motives in addition to rational economic considerations.[2]

Lists of possible determinants of industrial buyers' vendor selection, though helpful, are, by themselves, non-operational. Knowledge of the existence of a multiplicity of variables without a consideration

---

[1] Industrial Marketing Committee Review Board, "Fundamental Differences Between Industrial and Consumer Marketing," **Journal of Marketing**, Vol. 19 (October, 1954), p. 153.

[2] Hector Lazo, "Emotional Aspects of Industrial Buying," in Robert S. Hancock (ed.), **Dynamic Marketing for a Changing World**, Proceedings of the 43rd National Conference of the American Marketing Association (Chicago: American Marketing Association, 1960), pp. 258-65.

of their relative importance provides little direction for the industrial marketer's efforts, such as which selling strategy he should employ.

Despite the importance and relevance of this problem, only a few studies have tackled it, and these using primarily the technique of a multiple linear regression analysis.[3] Since the standard linear regression model ignores the interaction terms in the predictive equation (provides main effects only), there is some doubt as to its applicability to the evaluation problem. Hence, the primary purpose of this paper is to suggest a more refined model, one which will take interaction effects into account, for the problem of assessing the "importance weights" of the various factors affecting the buyer's vendor selection decision.

THE DETERMINANTS OF VENDOR SELECTION

Industrial buyers' decisions in general, and their vendor selection decisions in particular, are functions of a number of determinants. One conceptual scheme for the analysis of such behavior has classified the determinants of industrial buyers' decisions into five sets of variables.[4]

1. The buyer's own characteristics, especially his psychological mechanisms and behavioral characteristics, which serve as the major mediating processors between the inputs to which he is subject and his outputs (responses).
2. Interpersonal influences of other organizational members.
3. Organizational variables. The effect of these variables on the behavior of the organization members has been widely recognized by behavioral scientists but almost entirely neglected by marketing experts.
4. Inputs from the various sources of supply. These inputs are generally of two types: (a) those supporting source X and (b) those contradicting inputs which attempt to negate the influence of the supporting inputs for source X.
5. Environmental variables, which are of three types: (a) general variables affecting the value system of the people of a given society, (b) general business conditions, and (c) regular business constraints.

Of these variables, the fourth set is of greatest value to the marketing manager, since its variables are controllable by the vendor. Even though the buyer's perception of these inputs might not correspond with the vendor's actions (i.e., the perceived stimuli might differ from the actual or "true" stimuli), these variables still are subject to the vendor's control and manipulation more than any of the other determinants of the buyer's source selection decision. Furthermore, the "importance weight" which a buyer ascribes to each

---

[3] Yoram Wind, "Industrial Buying Behavior: Source Loyalty in the Purchase of Industrial Components," (Unpublished Ph.D. dissertation, Stanford University, 1966).

[4] Yoram Wind, "The Determinants of Industrial Buyers' Behavior," in Patrick J. Robinson and Charles W. Faris, **Industrial Buying and Creative Marketing** (Boston: Allyn & Bacon, Inc., 1967), p. 152. Reprinted by permission of publisher.

variable also might reflect some aspects of the other four sets of variables, suggesting that from an operational point of view it might suffice to consider only the performance characteristics of the various vendors and their respective weights.

The Evaluation Function and Configural Scoring

In the application to follow we shall use the terms "evaluation function" and "configural scale." Whenever a buyer decides which vendor to select (or, for that matter, whenever one decides which restaurant to frequent, which business meeting to attend, which girl to date), he is faced with the problem of evaluating *multi-attribute* alternatives. One vendor may be highly reliable on delivery but poor with respect to marketing services, while another may be merely adequate on both, and so on. Somehow, in order to make decisions we must—and do—"collapse" partially-ordered alternatives (ordered only on a component-by-component basis) to a simple order of overall "worth." The resultant total worth sometimes may be measured numerically, such as pricing a used car; sometimes ordinally, such as ranking a group of job applicants in terms of their desirability for employment; sometimes only nominally, as when one's driving performance is evaluated on a pass or fail basis by the patrolman in charge of issuing operators' licenses.

We call the rule (explicit or implicit) by which such partially-ordered alternatives are transformed into simply-ordered alternatives the evaluation function.[5] One important subclass of such functions is known as *linear compensatory,* in which it is assumed that the overall worth of a multi-attribute alternative is equal to an additive combination of the part worths.[6] In this model it is assumed that the individual attributes of an alternative are scaled such that the more of each, the better. What often is not known explicitly are the importance weights to be assigned to each component of the multi-attribute alternative.

It may turn out, however, that a *simple* (main effects) additive combination is inadequate to account for some evaluations. For example, the fact that a used car has both an air conditioner and automatic transmission may result in its being priced higher than the sum of the separate part worths of the extras (perhaps a manifestation of the additional value of "elegance"). Configural scoring[7] repre-

---

[5] R. M. Dawes, "Social Selection Based on Multidimensional Criteria," **Journal of Abnormal and Social Psychology**, Vol. 68 (January, 1964), pp. 104-109.

[6] Roger N. Shepard, "On subjectively optimum selection among multiattribute alternatives," in Maynard W. Shelly, II and Glenn L. Bryan (editors), **Human Judgments and Optimality** (New York: John Wiley & Sons, 1964), pp. 257-281.

[7] H. G. Osborn and Ardie Lubin, "The Use of Configural Analysis for the Evaluation of Test Scoring Methods," **Psychometrika**, Vol. 22 (December, 1957), pp. 359-371.

sents an approach for dealing with compensatory models which may involve interaction terms.[8] The results of this approach then may be compared with the simpler main effects (regression) model in order to see if the additional sophistication is needed.

The configural scale involves a polynomial function of the component values:

$$V = d_0 + d_1 x_1 + d_2 x_2 + \ldots + d_t x_t + d_{12} x_1 x_2 + d_{13} x_1 x_3 + \ldots d_{123} x_1 x_2 x_3 + \ldots$$

where 2^t terms are involved and each component is a $+1$ or -1 coded variable. (Extensions can be made to deal with more than dichotomous-state variables.) As would be surmised, in contrast to regression models, all interaction terms are specified. More specifically, suppose one were asked to assign values V_1, V_2 and V_3 to characteristics 1, 2 and 3, separately. Then one might be asked to assign *total* values to various combinations of characteristics 1, 2 and 3. From these total values one could impute (through regression procedures) importance weights to each of the characteristics. Would the rank order of these weights, d_1, d_2 and d_3, agree with the rank order of V_1, V_2 and V_3?

Actually one can do more than this through configural scoring methods. One can find imputed interaction terms as well as main effects. If the interaction weights of variables 1 and 2 are significant, then

$$V \quad (\text{Char. 1 } \textit{and} \text{ Char. 2}) \neq d_1 + d_2.$$

Both of the above questions can be attacked by the use of configural scoring methods; as such, linear regression is viewed as a special case of the configural scale in which all interaction terms are assumed to be nonsignificant.

Research Questions

The study to be described here is an attempt to examine the applicability of the evaluation function approach to vendor selection in an experimental environment. More specifically, we were interested in the following questions:

1. Can the evaluation function approach be used to determine the implicit importance coefficients which purchasing agents assign to vendors' performance characteristics in making selection decisions?
2. Does the linear regression model (main effects only) represent a good approximation to the configural scale?

[8] Other classes of evaluation functions, such as the disjunctive and its dual, the conjunctive, are possible but are not elaborated on in this paper. See, for example, R. M. Dawes, "Toward a General Framework for Evaluation," MMPP 64-7, Michigan Mathematical Psychology Program, University of Michigan, September, 1964.

3. What is the relative importance of each performance characteristic, and how does this change as the multi-attribute alternatives under evaluation become more complex?
4. Are the explicit "importance weights" which purchasing agents assign to various performance characteristics of vendors consistent with those derived (implicitly) from overall worth ratings assigned to multi-attribute alternatives?

METHODOLOGY

The first question dealt with in this study was: What are the relevant vendor's performance characteristics? A preliminary set of structured and unstructured interviews with various buyers of two large U. S. and Canadian companies indicated that ten characteristics were thought to be salient in the choice of suppliers. These were the vendor's:

1. Delivery reliability
2. Quality/price ratio of his product
3. General reputation
4. Geographical location
5. Importance as a client (reciprocity)
6. Supply of information and market services
7. Extent of "personal benefits" supplied to the buyer
8. Extent of previous (satisfactory) contact with the buyer
9. Technical ability and knowledge
10. Technical innovativeness

The exhaustiveness of this list of attributes was tested explicitly on two groups of industrial buyers, and it seems reasonable to assume that it includes virtually all of the relevant attributes of alternative vendors.

The specific respondents of this study consisted of 20 purchasing agents employed by a major manufacturing firm. Since the purpose of this study was to determine the applicability of the evaluation function approach, they were confronted with a set of "stimuli" which they were asked to evaluate in terms of overall worth. The "stimuli" were descriptions of hypothetical vendors whose performance characteristics were multi-attribute. The ten performance characteristics listed earlier were dichotomized as "ideal" or "poor," "yes" or "no," whichever was appropriate, and were presented to the purchasing agents in a three-phased study.

In *phase 1* of the study, four combinations of three items each (stimuli numbers agree with the previous numbering of vendor characteristics) were prepared:

1. Stimuli 1-5-9
2. Stimuli 2-5-8
3. Stimuli 2-6-7
4. Stimuli 3-4-10

Each respondent received all the above combinations of three items in the form of slips of paper on which a particular level of each was printed. For a given replication, eight such levels were possible, ranging from the case where all three characteristics were at their "high" level to the case where all three characteristics were at their "low" level. The two extreme cases were "scored" 100 and 0, respectively, in advance. For the remaining six cases the respondent was asked to write down a number between 0 and 100 which reflected the overall "worth" of the given combination of characteristics to him in his role of purchasing agent. Combinations, such as the one illustrated in Figure 1, were assigned in random order to each purchasing agent.

FIGURE 1

EXAMPLE OF A VENDOR DESCRIPTION CARD

The vendor's delivery reliability is	This vendor is an important client of ours	The vendor's technical ability and knowledge is	OVERALL SCORE
POOR	NO	IDEAL	

In *phase 2* of the interview the purchasing agent was given similar slips of paper on which *six* characteristics were printed. Three sets of characteristics, comprised of the following combinations, were chosen:

1. Stimuli 2-3-5-7-9-10
2. Stimuli 2-4-5-6-7-8
3. Stimuli 1-2-4-5-8-9

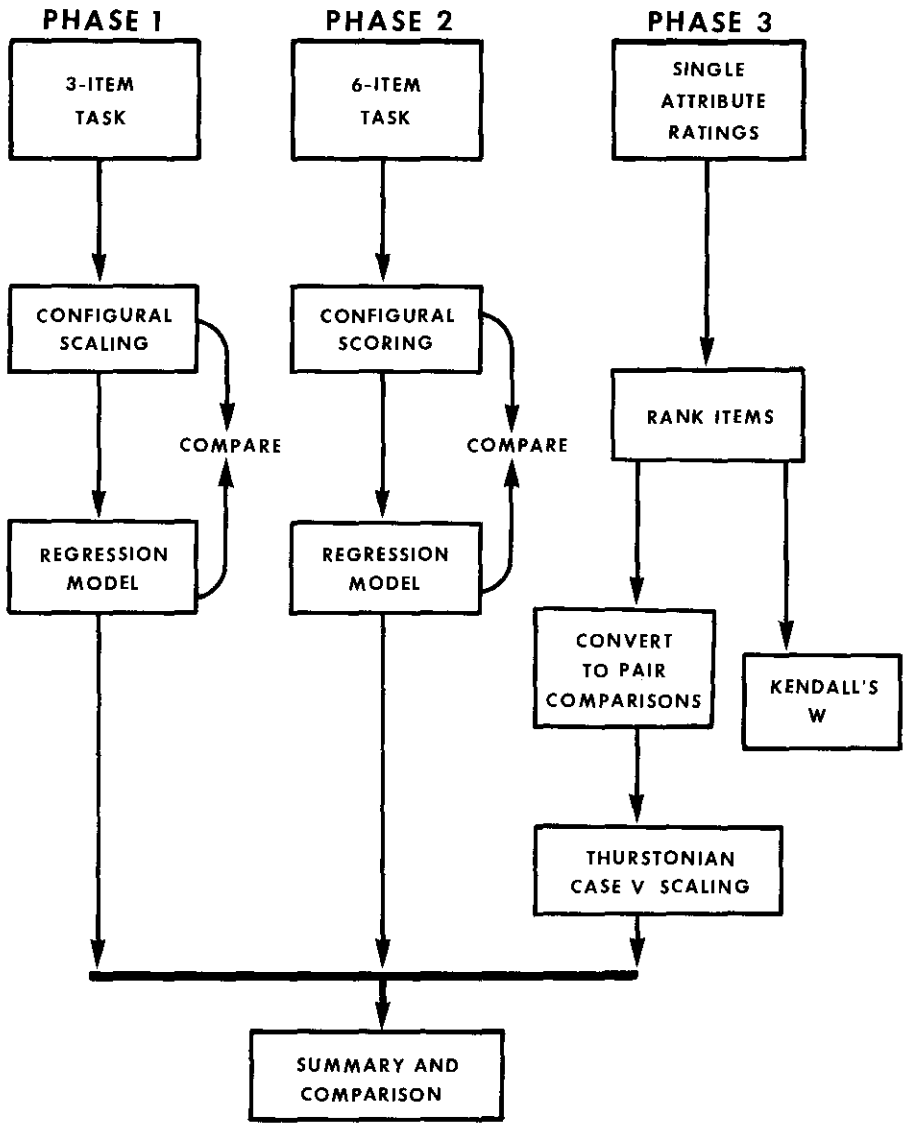
Since each set (two levels per factor) leads to 64 combinations, each respondent received only one-fourth of the total for each replication; thus each combination yielded five complete observations. The task of the respondent was similar to that described under the case involving three-stimulus combinations.

The last task, *phase 3,* involved the buyer's *direct* assignment of "part worth" to each of the ten stimuli. The respondent was asked to choose first the characteristic most important to him and assign it a worth of 100. Next he was asked to pick out the least important characteristic and assign it a worth of zero. Worths then were assigned to the remaining eight characteristics which reflected their importance relative to the anchor-point stimuli.

An overview of the analysis appears in Figure 2. As noted from the chart, the results of phases 1 and 2 were analyzed by the con-

FIGURE 2

OVERVIEW OF ANALYSIS — VENDOR RATING DATA



figural scoring and regression programs,[9] while phase 3 was analyzed by Thurstonian scaling techniques.[10] In addition, Kendall's coefficient

[9] F. J. Carmone, P. E. Green, and P. J. Robinson, "COFIS — an IBM 360/65 Program for Configural Scoring and Regression Analysis," University of Pennsylvania Computer Center, Philadelphia, Pa., November, 1967.

[10] J. M. W. Hogan, "STAT-PREF — An IBM 360/65 Program for Unidimensional Analysis of Preference Data," University of Pennsylvania Computer Center, Philadelphia, Pa., October, 1967.

of interjudge agreement was computed.[11] In all cases only the stimuli were scaled, the stringent assumption being that the respondent group was homogeneous.

RESULTS OF THE ANALYSIS

In describing the results of the analysis, phase 3, which involved individual stimulus rating, is discussed first. Each subject's rating of each of the ten stimuli was first transformed to a simple rank order, ranging from "1" (highest rating) to "10" (lowest rating). The 20 x 10 matrix of rank order then was tested for interjudge agreement by computing Kendall's concordance measure.

The concordance coefficient was 0.65, with an associated F-ratio of 34.68, which was highly significant statistically. It may be concluded that substantial interjudge agreement exists with regard to the importance ranking of the ten vendor performance characteristics.

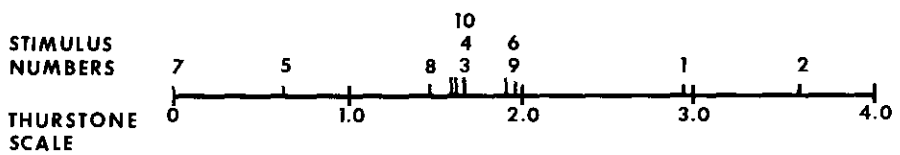
The rank data concerning ten performance characteristics then were converted into pair comparisons and submitted to the Thurstonian scaling program. Results of this analysis appear in Table 1.

As can be seen from Table 1, performance characteristic 2 (quality/price ratio) received an extremely large scale value, followed by characteristic 1. Characteristics 9, 6, 3, 4, 10 and 8 tended to cluster

TABLE 1

RESULTS OF THURSTONIAN CASE V SCALING

	Performance Characteristic	Scaling Value	First Differences
2	Quality/price ratio	3.61	
1	Delivery reliability	2.94	0.67
9	Technical ability and knowledge	1.95	0.99
6	Information and market services	1.86	0.09
3	General reputation	1.65	0.21
4	Geographical location	1.63	0.02
10	Technical innovativeness	1.61	0.02
8	Extent of previous contact with buyer	1.44	0.17
5	Importance as client (reciprocity)	0.61	0.83
7	Extent of "personal benefits" supplied to buyer	.0	0.61



[11] Maurice G. Kendall, **Rank Correlation Methods** (New York: Hafner Publishing Co., 1962).

in terms of intermediate scale values, while characteristics 5 and 7 received very low importance ratings. For this group of purchasing agents it is of interest to observe the possible relationship between responses and social stereotypy. Such socially unaccepted characteristics as reciprocity and "personal benefits" received extremely low scale values, while "economic-man" characteristics (quality/price ratio and delivery reliability) received quite high scale values.

Turning now to the question of whether "importance weights," assigned to characteristics singly, remain stable when presented in combinations, the three-attribute case (phase 1 of the study) next was analyzed. The results of this analysis are shown in Table 2.

TABLE 2

RESULTS OF THREE-VARIABLE CONFIGURAL SCORING ANALYSIS

Stimuli	d_0	d_1	d_2	d_3	d_{12}	d_{13}	d_{23}	d_{123}	eta^2	R^2	F
1-5-9	47.2	29.1	5.6	16.3	—0.2	2.7	0.3	—0.9	0.81	0.80	1.26
2-5-8	47.9	31.6	4.7	14.3	—0.6	3.6	—0.9	—0.6	0.84	0.83	2.54
2-6-7	46.6	33.4	15.5	1.8	2.8	—0.1	0.6	—0.7	0.88	0.87	1.96
3-4-10	42.8	21.0	12.8	18.1	3.0	2.5	1.6	—1.9	0.79	0.77	3.59

The coefficients, d_0, d_1, d_2, etc., represent the importance weights obtained from the configural scoring method.

We first can check to see if the interaction terms from the configural scoring procedure add anything beyond a simpler (main effects) model. This is done by comparing the eta-squared value[12] with the R^2 value obtained from assuming the simpler model which ignores interaction terms. In this case stimulus sets 2-5-8 and 3-4-10 are statistically significant at the 0.05 level, indicating that the configural model is superior to the simpler regression model. For example, in the 2-5-8 stimulus case it is seen that if both the quality/price ratio *and* the extent of previous (satisfactory) contact are high, the effect on total worth exceeds the sum of their separate effects.

In the special case of an equal number of observations obtained for each combination, the regression weights are identical to the main effect weights (d_1, d_2, d_3 of the configural model).

It is of interest to note, however, that the *rank order* of main effects weights is perfectly correlated in each case (separately) with the rank order obtained from the Thurstonian scaling (Table 1). This suggests that the respondents of this study displayed high consistency between single characteristic ratings and values imputed from their

[12] H. G. Osborn and Ardie Lubin, op. cit., p. 362.

responses to items in combination. In two of the four cases, however, interaction terms were significant, thus suggesting the inadequacy of a main effects (and, hence, single characteristics) model in predicting total worths.

In phase 2, six (rather than three) attributes were used as a stimulus. Each respondent received only 16 of the 64 combinations; with 20 subjects this resulted in five observations for each combination. The results of the configural scoring model (main effects and two-variable interactions only) are shown in Table 3.

The F-ratios are 1.28, 1.42 and 0.57 for the three cases, none of which is statistically significant at the 0.05 level. Thus it is of interest to note that the simple main effects (regression) model is adequate to account for the responses in all cases.

If one examines the rank order of the d-coefficients (main effects only) for each case separately, and compares it to the Thurstonian scaling results, the following results emerge:

Case 1 Stimulus Rank	2-9-10- 3-5-7
(Thurstonian)	2-9- 3-10-5-7
Case 2 Stimulus Rank	2-8- 6- 4-5-7
(Thurstonian)	2-6- 4- 8-5-7
Case 3 Stimulus Rank	2-1- 8- 9-4-5
(Thurstonian)	2-1- 9- 4-8-5

TABLE 3

RESULTS OF SIX-VARIABLE CONFIGURAL SCORING ANALYSIS
(Main Effects and Two-Factor Interactions)

Coefficients	Stimuli 2-3-5-7-9-10	Stimuli 2-4-5-6-7-8	Stimuli 1-2-4-5-8-9
d_0	42.9	45.7	43.5
d_1	19.8	21.4	10.3
d_2	8.5	6.7	21.5
d_3	3.3	3.8	4.7
d_4	2.9	8.9	3.7
d_5	9.7	1.1	5.4
d_6	8.9	9.4	4.8
d_{12}	3.4	2.7	2.1
d_{13}	0.5	—0.9	0.1
d_{14}	—0.2	2.6	—1.0
d_{15}	3.0	—0.3	0.8

TABLE 3 (continued)

Coefficients	Stimuli 2-3-5-7-9-10	Stimuli 2-4-5-6-7-8	Stimuli 1-2-4-5-8-9
d_{16}	—0.4	2.2	—0.8
d_{23}	0.3	—0.4	0.9
d_{24}	1.2	—0.6	1.2
d_{25}	0.8	0.2	0.9
d_{26}	2.2	—0.5	0.6
d_{34}	0.7	—0.8	0.6
d_{35}	2.3	—0.4	0.6
d_{36}	—2.3	—1.3	—1.3
d_{45}	1.1	0.3	—0.5
d_{46}	0.4	1.9	—0.1
d_{56}	—0.8	0.7	1.0
eta^2	0.74	0.81	0.87
R^2	0.66	0.75	0.86
F	1.28	1.42	0.57

While the results are not as good as phase 1, this comparison still indicates rather close agreement. In case 1 we note only one reversal in rank—stimuli 10 and 3. Cases 2 and 3 show more reversals, but in all instances the reversals occur in the "middle group" of Thurstonian scale values, namely those associated with stimuli 9, 6, 3, 4, 10 and 8. The "extreme" stimuli (2, 1, 5 and 7) are clearly differentiated even when as many as six stimuli are used in combination.

SUMMARY AND IMPLICATIONS

This study has five major findings which require further test and validation:

1. For this group of purchasing agents, the importance weights associated with single factor rating agree in rank order with those obtained from configural scoring methods used in phase 1, suggesting rather high consistency in their evaluation.
2. The Thurstonian scale values indicate two highly important characteristics, quality/price ratio and delivery reliability; a cluster of middle-scale values; and two low rated characteristics, reciprocity and "personal benefits."
3. The phase 1 configural scoring procedure indicated that in two of the four cases overall evaluations reflected significant interactions, suggesting the danger of using main effects models exclusively, at least when a small number of attributes are involved. In addition, phase 1 showed perfect rank order correlation with scale values obtained by the Thurstonian procedure.
4. The phase 2 configural scoring procedure also yielded rather consistent results with the Thurstonian scaling technique, at least insofar as rank order of importance weights is concerned.

5. In the six-variable case of phase 2 *no* significant interactions were noted, suggesting that as the number of attributes increases, the main effects model (without interaction terms) may be a good approximation to how respondents "combine" importance weights for individual attributes.[13]

Since the study was conducted on a small sample of purchasing agents of one company, few generalizations can be drawn with any degree of confidence. Yet, the findings suggest some interesting implications, at least for the given sample. These implications are of three major types: implications for industrial marketers, for industrial buying organizations, and for future applied research.

The most interesting findings of this study, which suggest a number of possible implications, appear to be the general applicability of the concept of the compensatory evaluation function and the determination of imputed importance weights. As was noted in the analysis of the results of phase 1 of this study, significant interactions may occur which would not be revealed in the usual method of single-factor rating. Given this finding, it seems desirable that industrial marketers try the evaluation function approach in attempting to study buyers' preferences for alternative suppliers. Such an approach appears feasible and the data required can be generated fairly easily.

Having conducted such a study, and assuming that the buyer's responses reflect the "real" importance weights that they ascribe to the various vendor characteristics, industrial marketers could gain a better understanding of the relative importance of the various determinants of the buyer's vendor selection decision. This, in turn, could guide them in the design of their marketing strategies so as to better fit the buyer's expressed preferences.

There is no assurance, however, that the buyer *actually* would assign the same weights in real-life situations. It is possible that the buyer in a test situation might tend to give socially acceptable answers, whereas in daily operations he might place a much higher weight on less-socially-acceptable variables, such as reciprocity and "personal benefits." Although such a situation is conceivable, our previous studies in this area indicate quite clearly that an acceptable price/quality ratio and reliable delivery are prerequisite for the consideration of any vendor. Hence, the direction of the findings of this study are consistent with other existing data.

Further implications for industrial marketers might involve incorporation of the findings from an evaluation function approach into a generalized micro-simulation of industrial markets.

---

[13] D. B. Yntema and W. S. Torgerson, "Man-Computer Cooperation in Decisions Requiring Common Sense," **IRE Transactions on Human Factors in Electronics**, Vol. HFE-2 (March, 1961), pp. 20-26.

The approach suggested in this paper also might be of some value to the buying organization, particularly in attempting to determine the importance of the variables that the buyers consider in their source selection decisions and to assess whether they are consistent with the company's objectives. In our specific case, for example, it is evident that reciprocity has a very small weight in the overall source selection decision. If a company should decide that it would like to "promote reciprocity," explicit instructions to this effect would have to be issued to the buyers. The effectiveness of such a policy then might be tested by using the evaluation function at a later time and comparing the relative weight of reciprocity to its weight in the base period. The validity of all of these implications depends, of course, on the validity of the following assumptions:

1. The verbal responses reflect the actual weight that would be assigned to each attribute in an actual behavioral situation.
2. The general weights that were derived are valid for all products. This assumption does not represent a serious limitation, since it is possible to estimate an evaluation function for each product class.
3. The buying population is homogeneous with respect to their evaluation of the various vendors.

Finally, the evaluation function procedure readily might be combined with perceptual mapping[14] in order to portray not only the evaluative weights but also the appropriate dimensions which are viewed as salient in making overall ratings of a set of multi-attribute stimuli. Such extensions would permit more careful study of the *particular* evaluative model being used, e.g., the "vector" or compensatory model (assumed in this study) or some other type of model, e.g., "ideal point" or even lexicographic.[15] We intend to make these extensions in subsequent research.

---

[14] P. E. Green, M. H. Halbert, and P. J. Robinson, "Perception and Preference Mapping in the Analysis of Marketing Behavior," paper presented at the Attitude Research on the Rocks Conference, Puerto Rico, October 26-31, 1967.

[15] C. H. Coombs, **Theory of Data** (New York: John Wiley & Sons, 1964).