Journal of Official Statistics
Vol. 8. No. 1, 1992, pp. 77–86

The Golden Numerical Comparative Scale Format for Economical Multi-Object/Multi-Attribute Comparison Questionnaires

Linda L. Golden[1], Patrick L. Brockett[2], Gerald Albaum[3], and Juan Zatarain[4]

Abstract: This paper presents a scale format for designing self-administered questionnaires that offers significant cost and space economies over previously developed alternative formats for gathering perceptual information across multiple objects and image dimensions. Results of empirical investigations show that the Golden Numerical Comparative Scale obtains these economies with no loss of data quality or reliability. In so doing, the Golden Numerical Comparative Scale also reduces the potential for measurement and coding errors.

Key words: Self-administered questionnaire format; multiple object comparisons; image measurement; graphic positioning scale; semantic differential scales.

1. Introduction

Many surveys are designed for the purpose of contrasting the perceptions of several objects along multiple attribute dimensions. For example, competing retail stores might be compared across multiple image dimensions, countries might be compared across several important subjective dimensions, politicians may be compared across several issue dimensions, or several manufacturers (or brands) might be compared across multiple desirable attribute dimensions.

The physical implementation of the desired multi-object/multi-attribute comparison can be cumbersome, difficult to structure and relatively expensive to produce when using traditional questionnaire scale formats (such as the bi-polar adjective scales and their modified versions) for self-administered questionnaires. In addition, because of the need to repeat all the attribute scales for each object being compared, the traditional bi-polar adjective semantic differential formats can be space consuming, problematic (from the perspective of question order effects on perceptions of the objects being examined) and costly to produce and mail. Still, many researchers continue to use some version of the semantic differential or

[1] Department of Marketing Administration, University of Texas at Austin, Austin, TX 78712, U.S.A.
[2] Departments of Finance, Mathematics, Management Science and Information Systems, The Applied Research Laboratories, and IC2 Institute, University of Texas at Austin, TX, U.S.A.
[3] Department of Marketing, University of Oregon, Eugene, Oregon and IC2 Institute, University of Texas, Austin, TX, U.S.A.
[4] Banco Nacional de Mexico, Venustiano Carranza 63, 06000 Mexico DF, Mexico.

Acknowledgements: Sincere appreciation is expressed to Mary R. Zimmer of the University of Georgia for her contributions to the first study. The University of Texas Research Institute provided financial assistance for the graphic positioning scale coding in the first study and the New York University Institute of Retail Management provided funding for the first study's data collection.

horizontal bi-polar adjective rating scale because of the perceived lack of available space saving alternatives. In other instances, researchers may not collect all the desired data because of perceived cost or space constraints thereby redefining the research problem unnecessarily.

This paper summarizes major findings about the properties of a newly developed questionnaire scale format designed specifically for self-administered surveys whose objective is to simultaneously contrast multiple objects across multiple attributes. This new scale format (the Golden Numerical Comparative Scale (GNCS)) provides major space and data coding economies over the bi-polar adjective scales historically used for these types of questionnaires and, in addition, there appears to be no loss of data quality when the GNCS format is used.

2. Background on Multi-object/Multi-attribute Comparison Questionnaire Formats

The major versions of the semantic differential scale formats that have historically been used for multi-object, multi-attribute perception questionnaires are summarized below.

1. The *traditional semantic differential* (TSD), presented below, involves rating one object on all attribute dimensions before the next object is rated.

Object A

Attribute 1	low	____:____:____:____:____:____:____	high
Attribute 2	low	____:____:____:____:____:____:____	high
⋮			
Attribute *n*	low	____:____:____:____:____:____:____	high

Object B

Attribute 1	low	____:____:____:____:____:____:____	high
Attribute 2	low	____:____:____:____:____:____:____	high
⋮			
Attribute *n*	low	____:____:____:____:____:____:____	high

2. A *modified traditional semantic differential* (MTSD) is structured such that all objects are evaluated on a single attribute before another attribute is introduced.

Attribute 1

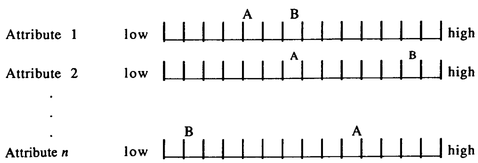
Object A	low	____:____:____:____:____:____:____	high
Object B	low	____:____:____:____:____:____:____	high
⋮			
Object K	low	____:____:____:____:____:____:____	high

Attribute 2

Object A	low	____:____:____:____:____:____:____	high
Object B	low	____:____:____:____:____:____:____	high
⋮			
Object K	low	____:____:____:____:____:____:____	high

While there are other variants of the TSD and MTSD (e.g., some researchers have used categorized graphic scales or have presented numerical values between the verbal anchors), it is readily apparent from the above illustration that when there are multiple attributes or dimensions to consider, and multiple objects to contrast along these dimensions, the TSD format and its variants can occupy considerable space potentially requiring multiple pages for expression. Accordingly, these scale formats can be relatively more expensive to produce in terms of printing costs, questionnaire length and mailing costs.

In order to overcome some of the space and cost deficiencies noted above for self-administered multi-attribute/multi-object comparisons, Narayana (1977) proposed the use of a graphic positioning scale (GPS) whereby all objects are simultaneously rated on the same scale line for each pair of bi-polar adjectives. This scale also allows the respondents to provide the desired multiple object cognizance contrasting each of the measured attributes of the rated objects at one time during the rating process (i.e., the relational properties of the values are developed in a simultaneous fashion by the respondent). With the GPS, the respondent evaluates all objects on the same scale via graphics (usually by placing letters representing all objects to be evaluated). This process represents the object's relative perceptual placement (as seen by the respondent) on the attribute scale between two bi-polar adjectives or phrases, as exemplified in the following example.



The graphic positioning scale was designed to provide a cost-effective alternative to the TSD and MTSD because of the space economy resulting from the measurement of perceptions of multiple objects on the same scale.

There are several potential disadvantages of the GPS format. For example, in order to perform statistical analysis using the results obtained with the GPS format, it is usually necessary to transform the non-numerical graphic responses into numerical values for subsequent data analysis. This additional task can be time-consuming (even when a template scale is specifically made for the purpose of conversion) and can provide a potential vehicle for introducing further ambiguities and uncertainties (and coding errors) into the actual numerical data obtained for analysis.

Illustrative of the ambiguities which arise and which might potentially increase the non-sampling survey error are the problems confronted by the coder in attempting to determine exactly where the letter in question is centered on a given graphical scale. Further ambiguities might arise due to the natural confusion on the part of both the respondent and coders as to how to handle situations in which different objects have very similar ratings resulting in several letters being crowded together at a single point on the scale. There is also an increased possibility of additional coding errors in the data records due to the fact that the questionnaires must have numbers *ascertained* as opposed to having given numbers already *recorded.* Still another disadvantage of the GPS is the fact that respondents tend to become confused and response forms appear cluttered when perceptions about a number of objects on a given attribute are close together.

In fact, Downs (1978) has compared the various semantic differential scale formats and found that respondents "preferred" the TSD scale and also found the TSD significantly "less difficult" to complete. This suggests potentially improved respondent reactions (e.g., response rate, omission rate, etc.) from a TSD type scale format. In spite of the relative ease of the TSD, however, it does not provide the space economies and agility of the GPS when it is desired to measure multiple image objects across multiple attribute dimensions. A scale combining the space economies of the GPS and non-error inducing ease of the TSD is desirable.

This paper examines a new *numerical comparative scale* first developed by Golden in a marketing research information gathering setting. As with Narayana's (1977) invention of the GPS, a motivating factor for the development was to obtain economical data gathering ability. It was also desirable to preserve the beneficial response and coding properties of the TSD scale format and to reduce (or avoid as much as possible) the time and involvement required of coders in imputing numerical responses for subsequent statistical analysis. The scale format which was developed (hereafter called the Golden Numerical Comparative Scale (GNCS)) is straightforward and exemplified below.

			Object A	Object B	Object K
Attribute 1:	low 1 2 3 4 5 6 7 high		___	___	___
Attribute 2:	low 1 2 3 4 5 6 7 high		___	___	___
.					
.					
.					
Attribute n:	low 1 2 3 4 5 6 7 high		___	___	___

In the GNCS format the respondents are asked to write that number which best describes their impression/attitude of the attribute in question for each of the objects being examined. For recording the response, a blank space is provided beside each of the attribute scales under a column heading bearing the name of each object. The GNCS combines the desirable numerical properties of horizontal bi-polar adjective phrase scales with the desirable space economy and cost-efficiency of the graphic positioning scale.

While data entry (i.e., file creation) is often taken for granted when examining questionnaire formats, this activity can, in fact, be a significant consideration in estimating total survey costs. This is one regard in which the GPS and GNCS differ (with the GNCS being much easier to code and with fewer input errors). This paper presents the empirical evidence concerning other attributes of the GNCS as compared to the GPS.

3. Motivation for an Empirical Comparison of the GNCS and GPS

The self-administered mail questionnaire is a frequently used measurement vehicle in survey practice. It is well known and documented that the choice of a scale format can strongly influence the costs of data collection, initially through questionnaire length and subsequently through the ease of data coding and data analysis. The format may also influence the amount of respondent effort required and the time and labor involved in preparing the data for analysis (e.g., open-ended versus closed-ended formats). Within the context of closed-ended (categorical) questionnaires, the scale must be understandable and not so cumbersome, long, or involved as to increase unnecessarily the cost per respondent reached or the cost of data compilation (e.g., not adversely affect response rates or completion rates).

In order to examine empirically the relative merits of the GNCS vis-a-vis the other possible scale formats for self-administered multiple attribute/multiple object comparison survey questionnaires, this paper compares only the GPS and the GNCS. The empirical comparison is restricted to these two scale formats for several reasons. First, when the GPS has been compared with the traditional semantic differential, the GPS has been found to produce no loss of data quality (Altuner, Altuner and Chappell n.d.; Bunder, Vincent, and Ursic 1984; Narayana 1977). Secondly, there seems to be no evidence of reduced reliability with the GPS (Jaffe and Nebenzahl 1984; Stem and Noazin 1985). These reliability results are also supported by Churchill and Peter's (1984) meta-analytic results across four rating scale types, including the semantic differential. Stem and Noazin (1985) also investigated test-retest reliability for both five and six perceptual objects on three-, five-, seven-, and nine-position TSD scales and a GPS scale and concluded that the graphic positioning scale was just as reliable as the traditional bi-polar adjective format. Thus, given the desirable space considerations of the GPS format, and the absence of any evidence of changed analytical results emanating from the use of the GPS, it might be considered to be the scale format of choice (from among the GPS, TSD and MTSD formats) for multiple attribute/multiple object comparison questionnaires.

Both the GNCS and the GPS have an advantage over the TSD and MTSD in that perceptions of the multiple objects appear on the same scale line for each dimension.

This eliminates the need to reproduce the scale itself *n* times for *k* objects across *L* dimensions, resulting in reduced questionnaire length and the consequent savings on both production and mailing costs. This also allows the respondents to think of the desired multiple objects at one time for each of the measured attributes.

Because in previous publications the GPS has compared favorably with other formats for self-administered questionnaires, it is only necessary to compare "the previous winner" (i.e., the GPS) with the GNCS in order to provide more definitive insight into scale selection. To date the GPS has not been compared to other scale formats which offer similar economic space and postage advantages. Additionally, because the GPS has the disadvantage of requiring additional coding time and labor to convert graphic ratings to numerical ratings, and introduces another manipulation of the data opening up the possibility of introducing coder errors, if the GNCS can be shown to produce data which is indistinguishable or superior to that produced using the GPS format, then the GNCS should become the preferred format for self-administered multiple attribute/multiple object comparison scales.

Only Altuner, Altuner, and Chappell (n.d.) have compared alternative scale formats for mail survey data. However, according to Albaum and Peterson (1985), a number of dependent measures should be used to evaluate methodological issues regarding mail survey response: response rate, cost, response speed, data quality, and data quantity. Accordingly, the GNCS was compared to the GPS for mail survey response rate, data content, data quality, and cost considerations. Response speed was not addressed.

The null hypothesis to be examined is that there are no differences between the GPS and GNCS on the respondent behavior measures of response rate and item completions. With respect to response content the null hypothesis is that there are no differences in mean attribute ratings, variance of attribute ratings, or internal reliability of the two scale types.

4. Methodology and Results

In order to empirically compare the GPS and GNCS several surveys were conducted. In the first study, a sample of 1,600 adults was selected from a nationwide consumer mail panel to represent sex, region, population density, and demographic criteria proportionate to the population of the United States. Because the panel members overrepresented females, one-half of the cover letters instructed the panel member to fill out the questionnaire him/herself while the other set of cover letters instructed the panel member to ask his/her spouse to fill out the questionnaire. Subjects were randomly assigned to the two scale-type treatments (GPS and GNCS) with each treatment type being sent to 800 potential respondents.

The contextual setting of this study concerned the scaled image perceptions of three large nationwide retail store chains (Sears, K-Mart, and Montgomery Wards) and the respondents' shopping frequency in each store. The results are discussed in more detail in Golden, Albaum, and Zimmer (1987). For the purposes of questionnaire scale format comparisons discussed in this paper, however, only that part of the data dealing with the multiple image perceptions of the multiple store chains will be used. These results will then be compared across the GPS and GNCS scale format types. The basis of scaled image perceptions was 19 store characteristics (such as price, cleanliness and employee friendliness) which were selected based upon a review of the literature on retail store image. Both of the two scale versions (GPS and GNCS) presented the attributes in the same order, and each was scaled in seven "categories," as appropriate for the format.

Respondents receiving the numerical comparative questionnaire scale treatment were asked to write the number from the scale that best described their perception of the particular store chain in the blank provided for each store chain. Those receiving the graphic positioning questionnaire scale treatment were told to write the first letter of each store chain (S, K or W) above the point on the scale which best described their impression of the store chain. Questionnaire returns were gathered for six weeks after the original mail-out of the questionnaire, and the final sample obtained consisted of 894 usable questionnaires. More GNCS questionnaires were returned (453) than GPS questionnaires (441). However, this difference was not statistically significant (using a binomial model for return probability).

The first phase of the analysis investigated sub-sample comparability. There were no statistically significant demographic (sex, age, education or income) or shopping frequency (e.g., behavioral familiarity for each of the three stores) differences between the GNCS and GPS. In addition, the samples represented a geographical cross-section of the United States.

Next, item completion rates were calculated at three distinct levels. At the most aggregate level, the proportion of respondents who completed *all* items for *each* store was computed. While the completion rates in Table 1 illustrate that the GNCS was always consistently higher than the GPS, the only situation for which there was a significant difference ($p < .01$) between formats occurred for the most familiar chain (Sears). Also presented in Table 1 is the second level at which item completion rates were analyzed, the average *omission* percentage *per respondent* for each store.

Results of paired t-tests for the respondent average omission rates indicated that all within-scale pairs were significantly different ($p < .01$). More people completed the GNCS format than did the GPS format. In addition, t-tests for independent group comparisons yielded a significant difference in mean omission percentages between scale types for both Sears and K-Mart ($p < .05$).

A third and final level of analysis for item omission rates was performed for *each* of the 19 individual scale items and concerned the percentage of respondents who omitted that particular item. With one exception the GNCS produced consistently lower item omission rates across scale items for both Sears and K-Mart, while for Wards (which was the least familiar object being evaluated) the reverse ordering was exhibited. Thus, in general, the GNCS dominated the GPS in terms of respondent completion rate characteristics.

In order to assess whether or not the scale format had any overall effect upon the measured image evaluation of any of the three store chains, the mean and variance of the responses to each of the 19 image attributes was computed for each scale and each store chain. For the GNCS data these numerical responses could be entered directly from the questionnaires; however for the GPS the data had to first be scored numerically (converted from the graphical to a numerical form) for subsequent analyses. The GPS responses were coded in increments of 0.25 and transformed to the same range as the GNCS for these analyses in order to obtain comparability of the results.

The data preparation was considerably more time and labor-intensive (and, hence, much more expensive to implement) for the GPS than for the GNCS. Because of the

Table 1. Item completion rates and omission rates

	All item completion rates		Average omission rates per respondent	
	GNCS	GPS	GNCS	GPS
Sears	81.9%	67.1%	4.3%	8.9%
K-mart	65.1%	59.4%	8.6%	12.8%
Ward's	53.0%	49.2%	32.5%	30.1%

need to convert respondents' GPS perceptions to numerical codes, the coding error possibilities for the GPS are very high. For this study, the initial data were coded numerically by a group of coders and then checked by another set of coders. The second coders, the coding quality assessors, found a sufficiently large number of GPS coding errors made by the initial coders such that another set of coders were trained to provide yet a third data coding quality check for all questionnaires. Part of the high coding error rate is due to the fatigue factor in coding a large number of GPS questionnaire responses. After this process of multiple data coding and entry checks, the GPS coding was considered essentially "error-free."

Fifty-seven one-way analysis of variance tests were run on the mean image ratings obtained using the two scale formats, and ten of these were statistically significant (more than one would expect to find simply due to chance). This is, of course, consistent with previous research indicating that different scale formats can result in different mean ratings (Jaffe and Nebenzahl 1984). It must be emphasized, however, that, since the "true" mean value for the population on these image variables is unknown for any given scale, it is impossible to say from the data whether one or the other scale is a more "accurate" depiction of the respondents' views. In this regard, to determine the relative accuracy of the different scale formats, one must turn to classical reliability theory and to an examination of item response variances.

Response variance analysis, and results of Bartlett's test for homogeneity of variance, indicated that there were 31 scales (out of a total of 57) where differences in variance between the two scale formats were statistically significant at $p < .05$. As the following data show, in 30 of the 31 differences the variance of GPS was greater than that of GNCS:

	Number of significant scales	Number with Var(GPS) > Var(GNCS)
Sears	6	5
K-mart	14	14
Ward's	11	11

A hypothesis postulating no significant difference in variances between scale types clearly cannot be supported. This is important because classical psychometric reliability theory posits that the observed score Y is related to the "true" underlying unobserved score T via the equation $Y = T + \varepsilon$ where ε is the error term. (A further discussion of this model can be found in Nunnally 1967, pp. 174–175.) Presumably the variance of the true score T is the same for the two scale formats (since both formats ask the respondent the same question about the same object and attributes in the same order). Thus, because $\text{Var}(Y) =$

$\mathrm{Var}(T) + \mathrm{Var}(\varepsilon)$ for each scale format, a smaller observed variance for Y using one scale format implies a smaller error variance (and hence higher reliability) for this scale format. By this measure, the GNCS format appears, in general, to be more reliable (subject to less error variance) than is the GPS format.

To confirm the general tendencies observed in the scale format comparison from the first survey discussed above, a second survey was conducted using these same two questionnaire scale formats (GNCS and GPS) but in a different context. This time, a study was developed to measure perceptions of products made in various countries. The countries considered (United States, Japan, Israel, East Germany, and Great Britain) were simultaneously compared using 13 bi-polar attribute dimensions of the products. The countries and attributes used in this study were the same as those used by Jaffe and Nebenzahl (1984), and the questionnaires were administered to 114 undergraduate students with respondents assigned to scale format treatments using a double change-over experimental design. A more detailed discussion of the results of this study can be found in Albaum and Golden (1991) and Zatarain, Golden, Albaum, and Brockett (1986).

Once again the data collected from the respondents receiving the GNCS treatment could be readily entered into a computer directly from the survey instruments whereas numerical values had to be physically measured and assigned (coded) for the GPS data prior to any statistical analysis. For comparability with the GNCS (a one to seven range), the left-most mark on the GPS scale was scored a 1 and the right-most marked was scored a 7 and 28 intermediate sub-intervals were represented as equally spaced values between 1 and 7. (This is the most frequently used form for coding of the GPS.) An individual respondent's score was then coded by determining to the best of the coder's ability, the sub-interval into which the graphic scale mark fell. The transformation to numerical values was coded on the questionnaires by hand and then entered into the computer data file following the same process as was implemented for the first study described.

Clearly, the physical coding processes which are needed in order to make the GPS operationally analyzable increase the costs of data preparation and potentially has data quality costs as well since it is more likely for a coder to incorrectly transpose from the graphic form to numerical equivalent than it is to key the numerical value given by the respondent. Even mechanical scanners do not put the coding of the GPS on an equal footing with the GNCS in terms of coding costs.

In this second study there were five countries whose product attributes were simultaneously examined along thirteen different attribute dimensions, for a total of 65 comparisons which were possible between the GPS and the GNCS scale formats. Looking first at the means, there were only four scales, out of a total of 65, where the mean difference between the two scale formats was statistically significant at a level of 0.05 or less. This number is well within the levels which chance alone would assign based on the binomial distribution. In addition, there was no discernable pattern with respect to either the incidence of significant difference or to the relative directionality of the non-significant means. Thus, from the mean value perspective, the results obtained by the two formats can be considered to be statistically indistinguishable.

Turning next to the comparison of variances produced by the two scale formats, (and by our previous discussions of the classical reliability models, to the corresponding implications for reliability of the two different scales), we note that with 65 comparisons there are expected to be 3.25 statistically significant differences at the 0.05 level, and that as many as 6.77 significant differences are required before one can conclude that there is a statistically significant difference in reliability between the two scales. The results of this analysis showed that there were eight statistically significant differences; however, in this study neither scale dominated the other statistically in the sense that directionality went in one way consistently. In fact, the GPS had statistically significant smaller error variances for five of the eight comparisons which proved to be statistically significant, and GNCS had statistically significant smaller error variances for three of the eight significant comparisons. Thus, although there were measurement error differences which were found, neither the GPS nor the GNCS emerged as statistically "better" or "worse" in this second study of the analytical properties of the two format types.

5. Conclusions

This paper summarized certain major findings regarding the GNCS as compared to other scale formats which might be used for self-administered (e.g., mailed out) multiple object/multiple attribute questionnaires. Because previous studies have shown that the graphic positioning scale (GPS) questionnaire format provides economic advantages over both the TSD and MTSD scale formats, the GPS was used as a basis for comparison with the newly presented Golden Numerical Comparative Scale (GNCS). Like the GPS format, the GNCS provides certain major space and data coding economies over the TSD and MTSD formats for self-administered questionnaires. In addition, there appears to be no loss of data quality when the Golden Numerical Comparative Scale is used as compared to the Graphic Positioning Scale. It was also found that the effort required to code GNCS responses was much less than that required with the GPS, while the reliability of the results obtained was no worse, and in some situations significantly better, than the reliability obtained using the GPS (smaller error variance).

The GNCS also was found to compare favorably with the GPS on response rate, data content, and data quantity measures such as lowering the item omission rates. This may stem from the fact that the GNCS is more closely related to the TSD scale format, and it has been shown that respondents prefer the TSD format.

There are a number of differences between the GPS and the GNCS and, thus, the exact (single) reason for the observed empirical differences in results is not clear. On the one hand, the GNCS is discrete and the GPS is continuous in response form. In addition, the GPS provides for within-scale respondent perceptions whereas the respondents write their answers immediately to the right of the scale for the GNCS.

It would be reasonable to speculate whether or not the removal of response locations from within the GNCS has an effect on response patterns for that scale. However, previous research has indicated that remote scale effects are distance related (Stem, Lamb, and MacLachlan 1978) and, since the GNCS is a version of the true semantic differential and not a distance remote scale (responses are immediately adjacent to each scale), there should not be any response bias induced.

While conclusions herein are restricted to the specific characteristics of these validation studies, they do appear fairly consistent across both studies conducted. For self-administered mail questionnaires involving the contrasting of multiple objects across multiple attribute dimensions, the GNCS is superior to the TSD and MTSD scale formats (as is the GPS), and, in addition, the GNCS can provide a substantial additional economic advantage over the GPS format by lowering data coding costs with no loss of response rates, potentially higher completion rates, potentially lower item omissions per respondent, and lower measurement error (variance).

6. References

Albaum, J. and Golden, L.L. (1991). Alternative Measurement Formats for Multiple Comparisons Across Multiple Image Objects. Journal of Global Marketing, 4, 89–121.

Albaum, G. and Peterson, R.A. (1985). A Paradigm for Methodological Research on Mail Survey Response. Paper presented at the Annual Educators' Conference of the American Marketing Association, August.

Altuner, H.J., Altuner, D., and Chappell, V.G., Jr. (n.d.). The Effect of Traditional versus Graphic Positioning Scales on Response Rate and Accuracy in Mail Surveys. Unpublished paper.

Bunder, P., Vincent, F., and Ursic, M. (1984). Graphic versus Semantic Differential Scales: An Empirical Comparison. Paper presented at the Southeast American Institute for Decision Sciences Annual Conference, March.

Churchill, G.G. and Peter, J.P. (1984). Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis. Journal of Marketing Research, 21, 360–375.

Downs, P.E. (1978). Testing the Upgraded Semantic Differential. Journal of the Market Research Society, 20, 99–103.

Jaffe, F.D. and Nebenzahl, I.D. (1984). Alternative Questionnaire Formats for Country Image Studies. Journal of Marketing Research, 21, 463–471.

Golden, L.L., Albaum, G., and Zimmer, M.R. (1987). The Numerical Comparative Scale: An Economic Format for Retail Image Measurement. Journal of Retailing, 63, 393–410.

Narayana, C.L. (1977). Graphic Positioning Scale: An Economical Instrument for Surveys. Journal of Marketing Research, 14, 118–122.

Nunnally, J.C. (1967). Psychometric Theory. New York: McGraw-Hill.

Stem, D.E., Jr., Lamb, C.W., Jr., and MacLachlan, D.L. (1978). Remote versus Adjacent Scale Questionnaire Designs. Journal of the Market Research Society, 20, 3–13.

Stem, D.E., Jr. and Noazin, S. (1985). The Effects of Number of Objects and Scale Positions on Graphic Position Scale Reliability. Paper presented at the Annual Educators' Conference of the American Marketing Association, August.

Zatarain, J., Golden, L.L., Albaum, G., and Brockett, P.L. (1986). The Numerical Comparative Scale in Country Image Studies. Proceedings of the National Institute for Decision Sciences Conference, Vol. 1, 515–517.

Received March 1991
Revised October 1991