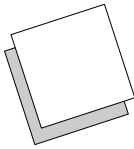


An executive summary for managers and executive readers can be found at the end of this issue



Measurement equivalence and applicability of core marketing concepts across Nigerian, Kenyan, Japanese and US firms

Kofi Q. Dadzie

Associate Professor of Marketing, Department of Marketing, J. Mack Robinson College of Business, Georgia State University, Atlanta, Georgia, USA

Wesley J. Johnston

Department of Marketing, J. Mack Robinson College of Business, Georgia State University, Atlanta, Georgia, USA

Boonghee Yoo

Department of International Business and Marketing, Hofstra University, New York, USA

Thomas G. Brashear

Department of Marketing, Isenberg School of Management, University of Massachusetts-Amherst, Amherst, Massachusetts, USA

Keywords Marketing concept, Measurement, Kenya, Nigeria, Japan, USA

Abstract Establishing the validity and measurement equivalence of core marketing concepts in the emerging market economies of Africa is a key step in assessing the transferability of modern marketing theory and managerial practice to these countries. However, measurement equivalence issues are rarely addressed in studies of marketing practices in Africa. Accordingly, this study examines the equivalence of core marketing concepts based on interviews of 459 marketing managers from Kenya, Nigeria, Japan and the USA. The results show that optimal scaling analysis of the managers' evaluations provide more valid and meaningful assessment than that of the raw data. The managers' evaluations of the concepts revealed amazingly similar or prototypical perceptions of marketing's core concepts and its applicability in their organizations, despite the profound country environmental differences. It appears that the concepts fall into two cross-national categories of applicability that permeate the industrialized and developing country categorization. Managerial and research implications are discussed.

Introduction

There is a long standing and unresolved debate in the marketing literature about marketing as a managerial practice applicable in all economies, especially the developing countries. Dating back to the early theoretical works of such scholars as Drucker (1958) and Emlen (1958) proponents espoused marketing as a more effective way to uplift the economies of developing nations. More recently, however, other scholars have continued the debate by examining marketing applications to social issues and concluded that marketing can be applied at the macro level and have suggested that such applications should be based on adaptation to local conditions (e.g. Akaah *et al.*, 1988; Hosley and Wee, 1988).

The authors acknowledge partial financial assistance of the Robinson College of Business at Georgia State University and the useful comments and suggestions of Edward Rigdon and Danny Bellenger on a previous draft of this manuscript.

Uplift economies of developing nations



The research register for this journal is available at
<http://www.emeraldinsight.com/researchregisters>

The current issue and full text archive of this journal is available at
<http://www.emeraldinsight.com/0885-8624.htm>



Issues yet to be addressed

To a large extent, however, this issue has not been addressed fully at the micro level: that is how practitioners implement such Western-origin concepts as core marketing concepts in local organizations in Africa. From a measurement perspective, the extant literature has overlooked the fact that practitioners in the developing countries of Africa may perceive and evaluate the applicability of marketing differently than would their counterparts from an industrialized market given the profound differences in field conditions. These differences can potentially confound the meaning and interpretation of the measures of marketing's usefulness (e.g. Aulakh and Kotabi, 1993; Bhalla and Lin, 1987; Brisling *et al.*, 1973). Yet, marketing scholars are yet to address such issues as: how valid are comparisons of marketing phenomena when made across developing countries of Africa and industrialized countries?

Core marketing concepts

The purpose of this study is to examine the measurement quality of core marketing concepts across Nigeria and Kenya, Japan and the USA and its implications for managerial practice of marketing in these countries. Specifically, this study seeks to:

- identify and separate the type of measurement equivalent problems associated with samples drawn from the different field conditions in Nigeria, Kenya, Japan, and the USA;
- examine the diagnostic value of major measurement equivalence diagnostic techniques, including optimal scaling and multi-group LISREL; and
- provide a substantive interpretation of such diagnostic assessment of the applicability of core marketing concepts.

Nigeria and Kenya were selected for this study because their economies reflect the type of field conditions that allow for comparison with the industrialized market economies of the West: marketing managers speak English, have some familiarity with marketing under conditions of low per capita incomes, intense state controls, political instability, and chronic incidences of a sellers' market environment. In addition, both countries have historical links with the global markets in Europe and the USA and use English as the official language of commerce. These conditions enable us to replicate and test the efficacy of standardizing cross-national marketing survey instruments in English, a common practice in international marketing research.

Four-country survey

We begin with an overview of the measurement equivalence issue in the LDC context. Next, we conduct a four-country survey specifically to address the measurement threats posed to an ordinal type marketing scale. Measurement equivalence has been deemed to be compounded with varying degrees of field conditions and measurement scales. By focusing on one type of qualitative scale, ordinal scale, the study can improve our understanding of the uniqueness of various types of measurement errors associated with the different field conditions in the emerging African market economies of Nigeria and Kenya.

Measurement equivalence issues in developing country context

Since the emergence of marketing as a managerial discipline, marketing scholars and practitioners have been keenly interested in comparing marketing practices and environments across field conditions in developing and the industrialized countries (e.g. Drucker, 1958; Currie, 1968; Emlen, 1958). Early efforts were motivated by practitioner and academic interest in formulating

appropriate marketing strategies for multinational corporations (Dixon, 1981). However, they soon were embroiled in a controversy about the relevance or applicability of marketing know-how given the environmental differences between developing and industrialized countries. Proponents long have argued that marketing should be directly applicable in the developing countries, regardless of environmental differences, because the fundamental concept of the discipline, need satisfaction, is universally relevant in all economies (e.g. Dixon, 1981; Drucker, 1958; Emlen, 1958; Kinsey, 1982; Rostow, 1960). Other scholars have argued that marketing will be relegated to a passive role in developing countries, based on the assumption that marketing and marketing institutions evolve after the economy has reached a sustained level of economic development (Arndt, 1972; Currie, 1968; Moyer, 1964). An emerging consensus suggests that Western marketing practices are applicable in developing countries if they are adapted to the unique local conditions (Akaah *et al.*, 1988; Hosley and Wee, 1988).

The question of whether marketing is less applicable in the different field conditions of developing countries is fundamental for measurement research because cross-national differences in field conditions can potentially threaten the validity of these competing perspectives (e.g. Douglas and Craig, 1983; Green and White, 1976; Mayer, 1978; Mullen, 1995; Singh, 1995).

Environmental differences and the applicability of marketing

Very low national incomes

The literature draws attention to several environmental differences between markets in developing countries and industrialized countries that could affect the way respondents evaluate marketing scales/measures. First, respondents from developing countries are likely to be less familiar with marketing measures because of the level of marketing practice. With very low national incomes (per capita incomes lower than \$400), most developing countries, especially in Africa, lack the foreign currency needed to purchase goods that must be imported due to a lack of domestic production capability. This situation often is associated with intense control by government agencies that seek to protect consumers from predatory marketing practices in conditions of economic shortages (e.g. Dadzie, 1989). As a consequence, a product-market environment emerges that corresponds to a sellers' market, in contrast to the traditional buyers' market that operates in most Western industrialized economies. As pointed out by several scholars, the extent of a sellers' market environment reduces marketing to demarketing (Dadzie, 1989).

High degree of state control over marketing decisions

A second reason why respondents from developing countries may be less familiar with marketing practices stems from the high degree of state control over marketing decisions in these countries. Unlike their counterparts in Western industrialized economies, marketing practitioners in developing countries often are not free to make decisions based on market conditions. Thus the traditional marketing-mix framework may be of little help to marketing managers because the government determines what distribution channels to use, what prices to charge, and the content of advertising messages (Akaah *et al.*, 1988). Although recent trends toward market reform suggest a decline in state control, the sustainability of these trends is uncertain. Therefore, marketing practitioners still must contend with a rather unpredictable government environment. In addition, the lack of political stability and consistency in policy orientations leaves considerable uncertainty for marketing managers, thus limiting the validity of the marketing framework.

Cross-national differences

Construct equivalence issues and developing country field conditions

Measurement scholars have drawn attention to several sources of measurement errors that frequently stem from cross-national differences (e.g. culture, language, individual). These sources represent serious threats to the validity of conclusions from comparative analysis (Douglas and Craig, 1983; Green and White, 1976; Mullen, 1995). These studies have emphasized that the most serious of such threats is the lack of construct equivalence (Bhalla and Lin, 1987; Douglas and Craig, 1983; Parameswaran and Yaprak, 1987). Construct equivalence refers to cross-national differences in the meaning of concepts or constructs being investigated and can lead to false interpretations and conclusions. Singh (1995) categorizes three conditions for enhancing construct equivalence. The first is functional and conceptual equivalence; that is, respondents from different countries must possess similar meaning and understanding of the construct under investigation. The second condition is instrument equivalence; that is, the scale items or response categories and other questionnaire stimuli (e.g. instructions) must be interpreted similarly in cross-national settings. The third condition is measurement equivalence; each scale item must measure the underlying construct similarly across nations. This condition requires that respondents from different countries/cultures with the same values on some variables must score similarly for that particular variable (i.e. there must be scaler equivalence) and use the same scoring format throughout the study (i.e. there must be scoring format reliability) (Bhalla and Lin, 1987; Douglas and Craig, 1983). In Figure 1, we present a summary of the key measurement issues discussed by Singh (1995).

Implicit in this discussion is the conclusion that the lack of measurement equivalence is a serious threat to the validity of conclusions about marketing's applicability in LDCs. However, the gravity of these issues will vary in different LDC contexts.

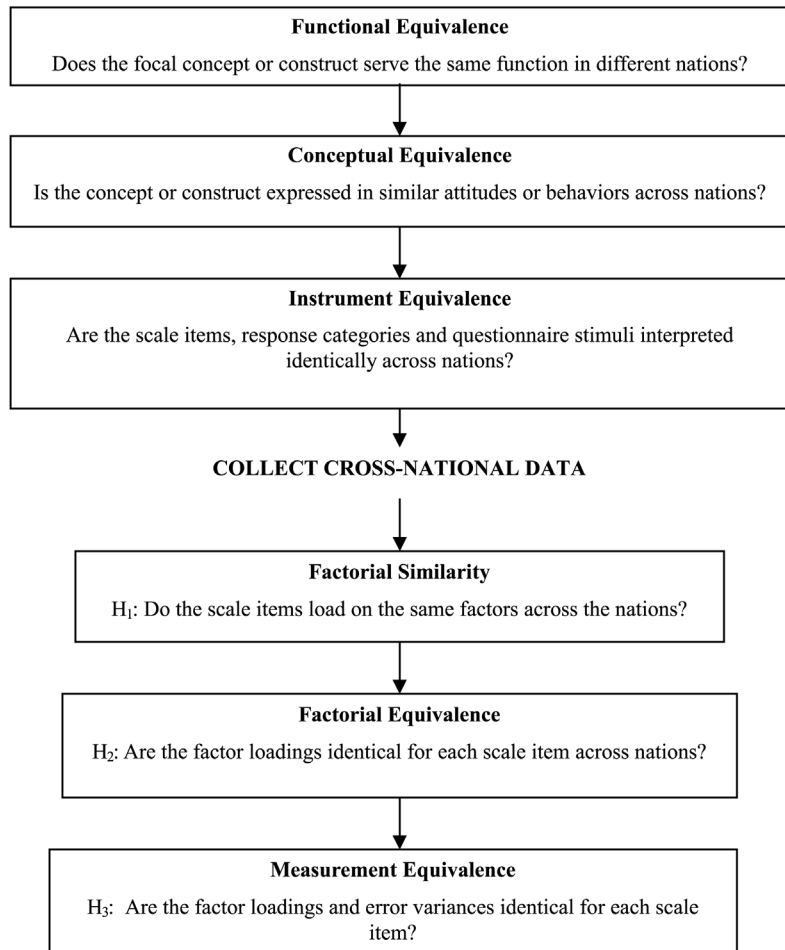
The meanings and functions of marketing

Functional equivalence issues. A potential threat to measurement equivalence in comparative analysis across LDCs and ICs is how respondents interpret the meaning and functions of marketing as a managerial philosophy and practice. The dramatically different product-market environments (sellers' vs buyers' markets) suggest corresponding differences in attitudes, familiarity, and perceived relevance of marketing and its related concepts (Akaah *et al.*, 1988). This difference should lead to a low high probability of functional equivalence.

In addition, evidence suggests that marketing activities are often among the first to be controlled by LDC governments (Akaah *et al.*, 1988). Hence, the LDC marketing manager is not free to make the type of decisions warranted by customer and market forces. In fact, government agencies often determine marketing decisions (Akaah *et al.*, 1988). In these conditions, traditional core concepts are viewed as lacking applicability. In contrast, US and Japanese managers enjoy a great degree of autonomy in making marketing decisions. These differences should also influence attitudes toward and perceptions of marketing in LDCs. As a result, functional equivalence will be a serious source of measurement error.

Effect of cultural differences

Scaler equivalence. Pronounced cultural, linguistic, and individual differences probably will manifest themselves in the way scales are interpreted and scored. The effect of cultural differences on construct equivalence is widely acknowledged, but an examination of the variation between LDCs and ICs is lacking.



Source: Jagdip Singh (1995) p. 597

Figure 1. Construct equivalence issues in cross-national research

The literature suggests that scalar equivalence will be affected by differences in US and Japanese respondents' cultural norms relating to humility, acquiescence, and conformity (Van de Vijier and Poortinga, 1982). Studies show that Japanese and other Asians tend to show a higher scale bias, in contrast to US citizens, who tend to show a lower scale bias (Van de Vijier and Poortinga, 1982). However, the effect of such cultural differences between US and African or between African and Asian respondents is not known. We believe that African respondents will show less upward scale bias than the Japanese and more upward scale bias than would US respondents because African societies value cultural norms regarding conformity and politeness. Hofstede (1991), for example, found that East Africans, West Africans and South Africans scored higher than US citizens on collectivism but lower on uncertainty avoidance. These views have been collaborated by other African scholars (Asante and Asante, 1985; Dia, 1991; Mazrui, 1990).

Differences in educational backgrounds

In addition, evidence suggests that individual differences can lead to upper and lower scale bias in scale interpretations (Wyser and Srull, 1989). These differences are probably more pronounced cross-nationally because of differences in educational backgrounds and familiarity with marketing between LDCs and ICs. Therefore, we expect differences in scale

Most optimal scaling studies have been conducted in domestic contexts

interpretations on the basis of cross-national patterns in individual differences. For example, US respondents should have less difficulty than Japanese and African respondents in scale interpretations.

Current diagnostic approaches

Several techniques are currently available for assessing the seriousness of measurement threats, including visual “examination of factor patterns for similarities” (Douglas and Craig, 1983) and testing for factor structure congruence (Douglas and Craig, 1983; Irvine and Carroll, 1980). However, the diagnostic value of these post-data collection procedures has been questioned because they do not reveal the underlying metrics and severity of measurement errors (Mullen, 1995; Singh, 1995). Hence, optimal scaling (Mullen, 1995) and multiple group LISREL have recently been proposed to provide more insight into the underlying metrics of cross-national data. These advocates recommend that LISREL provides the strongest overall test of measurement equivalence and should be used whenever possible (Mullen, 1995; Singh, 1995; Steenkamp and Baumgartner, 1998). However, because LISREL requires larger samples than may be feasible in many cross-national studies, its use may be limited in comparative marketing analysis where cell sizes can be under 50. Optimal scaling, on the other hand, can handle smaller samples than LISREL and may play a useful role in cross-national analysis. Yet, optimal scaling can not offer the rigorous hypothesis test that is possible in LISREL. Hence, most measurement scholars recommend that optimal scaling should be used with other techniques, preferably LISREL (Mullen, 1995). However, most optimal scaling studies have been conducted in domestic contexts (Didow *et al.*, 1985; Perreault and Young, 1980), but a few international studies have been conducted in Japan and the US (Mullen, 1995) and in the US and European countries (e.g. Davis *et al.*, 1981). Given the disparity between LDCs and ICs, the efficacy of optimal-scaling as a diagnostic tool should be a major concern for measurement scholars.

Some form of measurement error will always exist

Optimal scaling approaches to validation of measurement equivalence

Optimal scaling refers to a family of procedures for rescaling, validating, or improving non-metric data (i.e. data measured in categorical, nominal and ordinal scales) that are referred to under several names, including dual scaling and alternating least squares optimal scaling (Perreault and Young, 1980). The fundamental reason for rescaling field data is that some form of measurement error will always exist in survey data because field conditions are beyond the control of the researcher (Davis *et al.*, 1981; Perreault and Young, 1980). In addition, optimal scaling is desirable as a post-instrument development tool because users assign scale values to observed categories to maximize the relations between the observation and data analysis model, but respecting the measurement character of the data (Perreault and Young, 1980). The difference in field conditions between LDCs and ICs suggests that qualitative scales should be preferred over quantitative scales when comparing marketing concepts across LDCs; not only can respondents answer them more accurately, but they are also willing to provide the information requested in qualitative form (Perreault and Young, 1980). However, qualitative scales are also vulnerable to different interpretations. Therefore, optimal scaling of qualitative scales is recommended to reduce measurement error and to determine if measurement threats are serious enough to preclude cross-national comparisons (Mullen, 1995; Perreault and Young, 1980).

Procedure ideally suited

Recent improvements in optimal scaling using principal components of qualitative data (PRINQUAL) make the procedure ideally suited for a variety of field conditions. It can handle data from all levels of measurement (i.e. nominal, ordinal, interval, ratio, or mixed scales) and much smaller sample sizes (25-50) than does LISREL (Mullen, 1995). In addition, optimal scaling using PRINQUAL algorithms can display data along three dimensions, unlike the earlier procedures, using multidimensional unfolding (MDU) algorithms. This utility enhances cross-national comparison along two or three dimensions of cross-national data, a feature not available in previous studies (e.g. Mullen, 1995). Evidence suggests that PRINQUAL MDU solutions are more valid and are less likely to degenerate than other multidimensional scaling procedures such as correspondence analysis (Fisher *et al.*, 1987).

There are currently three types of optimal scaling algorithms in the PRINQUAL framework. They are:

- (1) the maximum total variance algorithm (MTV);
- (2) the maximum generalized variance (MGV); and
- (3) the maximum average correlation method (MAC).

For a detailed discussion of these algorithms, see Perreault and Young (1980).

MTV algorithm is appropriate

Maximum total variance algorithm

The MTV optimization algorithm (Kuhfeld *et al.*, 1985; Young, 1981; Young *et al.*, 1978) is based on the principal component model. This method attempts to maximize the total variance of the first r principal components whose number is much smaller than the number of variables. Under this method, Hotelling's (1933) ordinary principal components model is alternated iteratively with optimal scaling until convergence is obtained. The MTV algorithm is appropriate for studies in which maximizing the total variance is crucial. It is an improvement over the previous algorithm, labeled PRINCIPALS (Kuhfeld and Young, 1989), because it extends Hotelling's (1933) classical principal components analysis to a more general model that can deal with both quantitative and qualitative variables.

The ordinary principal component analysis handles an $m \times n$ matrix Z of m observations or subjects on n interval- or ratio-scaled variables where Z is composed of the following:

$$\hat{Z} = XF' \quad (1)$$

where X is a $m \times r$ matrix of m principal component scores on r principal components and F is an $n \times r$ matrix of n loading of the observed variables on the r principal components. To solve the identification problem, it is assumed that $X'X/m = I$ (identity) and $F'F = D$ (diagonal).

$$\theta = \text{tr}Z - \hat{Z}'(Z - \hat{Z}). \quad (2)$$

Conventionally, z is column standardized and then solved, such that equation (2) is minimized for a prescribed number of principle components where tr stands for trace, which is a sum of diagonal individual elements or variance terms. θ refers to the unexplained variance of Z , therefore, minimizing θ is equivalent to maximizing the total variance account for in Z . In this way, a scalar summary is generated on how well an original variance-covariance matrix is explained by its new predicted matrix.

No restriction on measurement characteristics

As an extension of this traditional principal components analysis, PRINQUAL handles variables measured at a variety of scale levels for both continuous and discrete variables. In addition, this algorithm can handle missing data or data that are either symmetric or asymmetric. In other words, there is no restriction on measurement characteristics. Therefore, Hotelling's (1933) model is a special case of PRINQUAL.

When some of the variables are qualitative, equation (2) which optimizes θ to solve Z , is: generalized to where Z^* is an $m \times n$ matrix of m observations or subjects on n optimally scaled variables:

$$\theta^* = tr(Z^* - \hat{Z})'(Z - \hat{Z}). \quad (3)$$

Here, Z^* is column-wise centered and normalized such that:

$$Z^{*'} l_m = 0_n \text{ and } diag \frac{[Z^{*'} Z^*]}{m} = I_n \quad (4)$$

where l_m and 0_n are vectors of ones and zeros and their subscripts refer to their orders.

PRINQUAL MTV optimizes θ^* in equation (3) under the normalization assumption of Z in equation (4) so that θ^* is optimized in terms of Z^* . The steps are iteratively alternated until convergence is obtained at the prescribed or default level.

Ideal for nominally scaled variables

One of the attractive features of PRINQUAL for rescaling cross-national data is its variety of fit functions for achieving data transformation. The first of these six functions is the optimal scoring function. It is ideal for nominally scaled variables (Fisher, 1938) with a category scoring function. The second function, monotonic B-spline (i.e. a smooth line in contrast to a straight line), is appropriate when the scale is at least ordinal and the variable is continuous (de Boor, 1978; de Leeuw, 1986). The third transformation, B-spline (de Boor, 1978), is appropriate under the same conditions as the monotonic B-spline. The fourth and fifth functions are monotonic with or without ties. Either function is appropriate when the scale is at least ordinal and the variable is discrete (Kruskal, 1964). The sixth and final fit function is the optimal linear method, which is appropriate for predicting missing values of interval variables.

The use of a B-spline transformation requires specifying the number of break points for the spline, or knots. This is possible using the SAS procedure (SAS Institute, 1992), in which a quadratic curve is used for a spline transformation of any variable x by a quadratic regression function (i.e. transformed $x = b_0 + b_{1x} = b_{2x}^2$). The number of knots n indicates a discrete variable with n different polynomial functions and n times decreased smoothness at the knots.

Empirical study of construct equivalence across Nigeria, Kenya, USA and Japan

Methodology

Cross-national study conducted

To determine if LDC respondents (Nigeria and Kenya) evaluate marketing scales differently from their counterparts in the industrialized countries (Japan and the USA), a cross-national study of applicability of marketing concepts was conducted.

Sample

A stratified random sampling plan, with country as a stratum, was used to select 1,450 firms from the Dun and Bradstreet *World Marketing Directory* (Dun and Bradstreet International, 1991). Firms were selected from each country in proportion to the number of firms in the respective country. The target sample surveyed includes 600 firms from the USA, 400 in Japan, 250 in Nigeria, and 200 in Kenya. A six-page, self-administered questionnaire was mailed to the senior manager in charge of marketing at each firm because previous evidence suggests that they are well informed about marketing in LDC firms (Akaah *et al.*, 1988). Another copy of the questionnaire was mailed to non-respondents six weeks later.

Cross-national comparison of marketing concepts

Nigeria and Kenya were selected for this study because field conditions in these countries are typical of most LDCs: low per capita income (less than \$500), chronic shortages of goods and services as a result of policy changes, prevalent sellers' market conditions, and strong state control. The countries Kenya and Nigeria provide some representation of eastern and western Africa respectively. These countries also have fledgling market economies and history of links to the global market economies of Europe and the USA. Hofstede (1991) found very few cultural differences among eastern and western Africans with respect to collectivism, power distance and uncertainty avoidance dimensions. These conditions were expected to contrast well with field conditions in Japan and the USA while still allowing for cross-national comparison of marketing concepts.

Questionnaire pretested in all four countries

The instrument was developed and administered in accordance with guidelines for designing an effective international marketing instrument (Brisling *et al.*, 1973; Douglas and Craig, 1983; Singh, 1995). To enhance functional and conceptual equivalence, the questionnaire was standardized in English for the US, Nigerian, and Kenyan samples, English is the official language for commerce and trade in these countries. Next, a professional Japanese translation firm translated a Japanese version of the questionnaire. It was pretested in the USA on a sample of bilingual Japanese students and later pretested in Japan. Then, in another attempt to increase functional and conceptual equivalence, it was back translated into English to ensure it corresponded with the English version (Brisling *et al.*, 1973). Finally, the questionnaire was pretested in all four countries. The pretest data were tested for instrument equivalence and found to be adequate for ordinal scale properties. Changes in the final survey incorporated the pretest information, including reducing the length of the instrument and more clearly defining major study variables and scale items.

Response rates slightly higher for USA

Of the 1,450 questionnaires mailed, a total of 459 were returned, of which 422 were usable, providing an overall net response rate of 29.1 percent. As expected, response rates were slightly higher for US (30.8 percent or 185/600), Japanese (29.5 percent or 118/400) than for Nigerian (26.8 percent or 67/250) and Kenyan (26.0 percent or 52/200) respondents. Individuals who were not senior marketing managers completed the few non-usable questionnaires. The sample characteristics are summarized in Table I. Most of the respondents held positions as senior marketing managers, held at least one college degree, and had direct marketing-related experience. Examination of respondent characteristics along selected variables suggests enough cross-national variation in field conditions to warrant measurement equivalence concerns. For example, the incidence of sellers' market conditions ($\chi^2 = 7.5$; $p < 0.01$) and intensity of state control of marketing

Characteristic	Percentages			
	USA (n = 185)	Japan (n = 118)	Nigeria (n = 67)	Kenya (n = 52)
<i>Industry category</i>				
Manufacturing	55.3	68.7	45.4	48.1
Transportation	6.1	1.7	3.0	11.5
Service	29.9	18.3	36.3	19.2
Distributive trade	8.9	11.3	15.1	21.1
<i>Existence of marketing department</i>				
Yes	91.1	71.1	78.5	80.4
No	8.9	28.9	21.5	19.6
<i>Thrust of marketing activity</i>				
Consumer marketing	24.4	24.6	28.6	25.5
Industrial marketing	48.9	57.0	25.4	43.1
Consumer and industrial marketing	26.7	18.4	46.0	31.4
<i>Size of organization (number of employees)</i>				
Less than 49	17.3	3.6	15.2	24.0
50 to 99	6.4	6.3	12.1	16.0
100 to 249	15.0	7.2	19.7	12.0
250 to 499	12.7	21.6	21.2	16.0
500 to 999	9.3	21.7	13.6	12.0
1,000 to 2,999	12.1	24.3	15.2	10.0
3,000 to 9,999	12.2	10.8	3.0	4.0
10,000 or more	15.0	4.5	0.0	0.0
<i>Organizational title/rank</i>				
CEO/president/owner/partner/principal	0.0	2.0	6.2	1.9
Senior VP/VP/general manager	25.7	6.9	13.8	19.2
Senior manager/project director	30.4	14.9	13.8	34.6
Functional manager	39.8	57.4	55.4	32.7
Other	4.1	18.8	10.8	11.5
<i>Gender</i>				
Male	72.3	99.1	98.5	90.4
Female	27.7	0.9	1.5	9.6
<i>Age</i>				
20-29	10.1	2.6	10.6	17.6
30-39	30.2	23.7	34.8	43.1
40-49	41.3	38.6	42.4	25.5
50-59	15.1	28.1	10.6	13.7
60 or older	3.4	7.0	1.5	0.0
<i>Marital status</i>				
Single	9.2	7.9	14.1	19.6
Married	78.8	90.4	85.9	78.4
Separated/divorced/widowed	12.0	1.8	0.0	2.0
<i>Higher education level</i>				
Some college or less	11.3	8.8	12.5	21.6
Bachelor's degree	25.9	85.0	29.7	29.4
Some post-bachelor work	27.0	3.5	18.8	15.7
Master's degree	31.4	2.7	35.9	29.4
Some post-master work	3.2	0.0	1.6	2.0
Doctoral degree	1.1	0.0	1.6	2.0
<i>Major field of study (if college or higher)</i>				
General business	11.9	19.8	6.3	12.0
Business – marketing	39.8	6.3	36.5	30.0
Business – accounting	1.7	1.0	6.3	6.0
Business – management	9.1	9.4	17.5	8.0
Business – statistics	0.6	2.1	0.0	0.0
Business – finance	2.3	1.0	3.2	6.0
Engineering	9.1	13.5	12.7	18.0
Other technical (e.g. physics)	2.8	6.3	1.6	6.0
Humanities	12.5	20.8	11.1	4.0
Other (e.g. education)	10.2	19.8	4.8	10.0

Table I. Characteristics of sample

decisions and activities ($\chi^2 = 11.5$; $p < 0.01$) were significantly higher among LDC respondents than among IC respondents ($\chi^2 = 7.5$; $p = 0.01$). Similarly, only a small proportion of the African (5 percent) and Japanese (0.9 percent) respondents were women, compared with nearly one-third (28 percent) of the US sample, a probable indication of cultural differences in the role of women in these countries. Also, fewer Japanese managers had business degrees than did respondents from the other countries. These cross-national differences among respondents' cultural and educational backgrounds underscore the need to validate the measurement equivalence of the present study as previously noted.

Measurement of core marketing concepts

Following Akaah *et al.* (1988), we measured the perceived applicability of core marketing concept using multiple items for the following variables:

- the marketing concept;
- optimization of the marketing mix;
- product differentiation;
- market segmentation;
- the building of brand loyalty;
- product positioning; and
- test marketing.

Each item scaled

Consistent with our focus on qualitative scale, the perceived usefulness of each item was recorded on a 1 (not at all useful) to 4 (extremely useful) Likert-type ordinal scale. Each item was scaled with the following phrases: 1 = not at all useful, 2 = not useful, 3 = quite useful, and 4 = extremely useful. The definition of each concept was provided as a further means of minimizing differences in translation and conceptual equivalence.

Diagnostic approach

Our five-phase diagnostic process followed recommended guidelines (Mullen, 1995; Perreault and Young, 1980; Singh, 1995). Accordingly, the first phase examined the original data matrix for validity and treated each country as a mutually exclusive sample. The test criterion was functional and translation equivalence of the study instrument, that is, the extent to which the scale items, response categories, and interpretation of the items are similar and meaningful across nations (Singh, 1995). In the second phase, we optimally transformed the original data and conducted validity checks for scalar equivalence and scoring format consistency across countries. These tests were conducted on individual study variables only. In the third phase, we repeated the second phase on all variables to determine measurement equivalence of the overall data structure.

Invariance test typically followed a hierarchical pattern

The fourth phase of the diagnostic process checked the validity of the optimally scaled data, first by examining the data for factorial similarity and factorial equivalence (Douglas and Craig, 1983), and second by using multi-group structural equation modeling using LISREL 8.30. The LISREL test is stronger than preceding tests of measurement validity, equivalence, and invariance (Bollen, 1989; Marsh and Hocevar, 1985). The invariance test typically follows a hierarchical pattern. The recommended pattern starts with the assumption that the number of latent constructs is the same for all groups.

Transformation performed separately for each country

The second test examines multigroup invariance of the factor loading, lambdas, of the latent construct. The next step tests the invariance of the errors terms of the manifest variables across groups. Finally, the correlations between the latent constructs were also tested. Because only one latent construct, marketing concepts, is examined in this study, the final step is not applicable.

In the fifth and final diagnostic check, we examined the entire data structure for cross-national comparability, using the multidimensional unfolding (MDU) algorithm in PRINQUAL. MDU is ideally suited to detect overall similarity/dissimilarity of data structure because it is less likely to yield a degenerated solution than other procedures (Fisher *et al.*, 1987). We performed our optimal scaling PRINQUAL analysis using the monotonic fit function (SAS Institute, 1992), which is recommended for ordinal data (Perreault and Young, 1980). The transformation was performed separately for each country. Convergence was achieved in less than 30 iterations with significant improvement (14 percent) in total explained variance over the original data. This indicates increased fit between our cross-national data and the Likert ordinal scale after optimal scaling. Whether such improvement aids measurement equivalence is investigated later in the study.

Cross-national comparison

Initial validation check for functional and construct equivalence

Our initial task was to determine if the data structure of original data were comparable cross-nationally. In Table II we present the mean score for the matrix of 422 respondents for the original data. The values in Table II are the percentage of respondents for each response category. For example, the first column in the US sample for item V1 (optimization of the marketing

Country	<i>n</i>	Raw ^a	V1 ^b	V2	V3	V4	V5	V6	V7
USA	185	(1)	5.5	4.4	7.7	6.0	4.4	6.0	30.2
		2	31.3	11.5	20.9	14.3	15.4	12.1	41.8
		3	29.7	37.9	31.3	33.0	32.4	31.9	19.2
		4	33.0	46.2	40.1	46.7	47.8	50.0	82.0
Kenya	52	(1)	3.0	1.5	23.9	3.0	11.9	11.9	20.9
		(2)	14.9	16.4	19.4	9.0	10.4	10.4	22.4
		(3)	43.3	47.8	32.8	32.8	23.9	23.9	28.4
		(4)	38.8	34.3	23.9	55.2	52.2	52.2	28.4
Nigeria	67	(1)	5.2	7.0	20.9	6.1	10.4	10.4	29.6
		(2)	37.4	30.4	37.4	21.7	33.0	33.0	40.9
		(3)	39.1	40.0	25.2	38.3	31.3	31.3	23.8
		(4)	17.4	22.6	14.8	36.5	25.2	25.2	6.1
Japan	18	(1)	3.8	1.9	21.2	1.9	1.9	1.9	32.7
		(2)	11.5	13.5	23.1	13.5	13.5	13.5	25.0
		(3)	28.8	26.9	30.8	26.9	26.9	26.9	30.8
		(4)	53.8	55.8	51.9	57.7	57.7	57.7	11.5

Notes: ^aRaw values on the “original scale” with 1 = not at all useful, 2 = slightly useful, 3 = quite useful, and 4 = extremely useful; ^bthe seven marketing concepts are V1 = optimization of the marketing mix (OPTMIX), V2 = product positioning (PROPOS), V3 = marketing segmentation (MARSEG), V4 = the marketing concept (MARCON), V5 = product differentiation (PRODIF), V6 = the building of brand loyalty (BRALOY), and V7 = testing marketing (TESMAR); the numbers represent the percentage of the frequency of each cell. For example, 5.5 in the first cell means that 5.5 percent of the US respondents (5.5 percent times 185 respondents = 10 respondents) chose the response category (1) for V1 = OPTMIX

Table II. A contingency table of original response categories: distribution of respondents by country (sample size weighted)

mix) represents the 5.5 percent of US respondents who evaluated the first item as belonging to the “not at all useful” scale category. The rest of the table values may be interpreted in a similar fashion.

An initial examination of the responses in Table II clearly suggests that LDC respondents rated marketing concepts as more useful than did US respondents. For example, the first marketing concept, optimization of the marketing mix, was evaluated as “not at all useful”/“slightly useful” by approximately 36.8 percent of the US respondents compared with only 17.9 percent of the Kenyan and 15.3 percent of the Japanese respondents. Only Nigerian respondents reported a higher average score for both response categories.

Similarly, for item V4, “the marketing concept,” a substantially larger proportion of US respondents (20.3 percent) evaluated it as “not at all useful”/“slightly useful”, compared with only 12.0 percent of the Kenyan and 15.4 percent of the Japanese respondents. Only Nigerian respondents had a higher “slightly useful”/“not at all useful” average score (27.8 percent). These findings seem counterintuitive because they suggest that marketing concepts are less applicable to US firms than they are to Kenyan and Japanese firms. We suspect that this interpretation results from differences in the interpretations of marketing concepts or lack of functional and construct equivalence. An optimal scaling diagnosis is thus warranted to provide an item-by-item diagnosis of the source and gravity of measurement inequivalence.

Optimal transformation analysis and results

The initial optimal scaling results required to determine metric equivalence across the four countries are summarized in Table III and Figures 2 and 3, which also show several notable patterns of inequivalence. First, Table III shows that linguistic differences tended to be associated strongly with scale interpretation across all four-scale

Findings seem counterintuitive

Several notable patterns of inequivalence

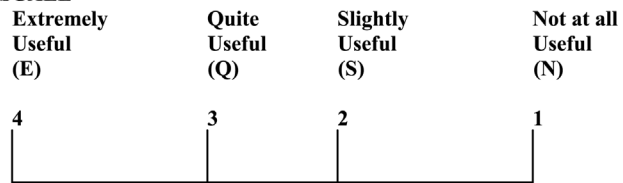
Country	n	Raw ^a	V1 ^b	V2	V3	V4	V5	V6	V7
USA	185	(1)	-0.188	0.002	0.670	0.226	0.634	0.533	0.759
		(2)	2.547	2.760	2.067	2.406	2.406	2.411	2.503
		(3)	3.306	3.262	3.286	3.588	3.154	3.095	2.524
		(4)	3.483	3.725	3.819	3.588	3.914	3.912	3.512
Kenya	52	(1)	0.032	-0.894	0.835	0.065	-2.168	-2.178	1.163
		(2)	2.744	3.377	2.398	2.566	3.151	3.209	1.907
		(3)	3.258	3.435	2.854	2.984	3.480	3.562	2.727
		(4)	3.806	3.588	3.961	3.932	3.480	3.562	4.468
Nigeria	67	(1)	-0.272	-1.206	0.811	-0.522	0.948	1.483	0.635
		(2)	2.456	2.293	2.358	3.112	2.649	2.253	2.488
		(3)	3.049	3.119	3.051	3.112	2.810	2.306	3.345
		(4)	3.867	3.790	3.827	3.836	4.038	4.142	3.538
Japan	118	(1)	1.460	1.460	1.602	1.518	2.101	1.898	1.294
		(2)	2.049	1.954	2.018	1.976	2.101	1.898	1.712
		(3)	2.676	2.865	2.056	2.789	2.745	2.580	3.037
		(4)	4.271	4.159	4.643	4.141	4.126	4.285	4.363

Notes: ^aRaw values on the “original scale” with 1 = not at all useful, 2 = slightly useful, 3 = quite useful, and 4 = extremely useful; ^bthe seven marketing concepts are V1 = optimization of the marketing mix (OPTMIX), V2 = product positioning (PROPOS), V3 = marketing segmentation (MARSEG), V4 = the marketing concept (MARCON), V5 = product differentiation (PRODIF), V6 = the building of brand loyalty (BRALOY), and V7 = testing marketing (TESMAR)

Table III. PRINQUAL – transformed values

“The” Marketing Concept

ORIGINAL SCALE

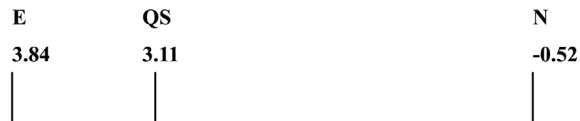


PRINQUAL SCALE (TRANSFORMED)

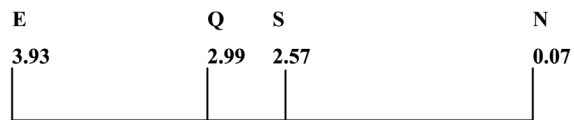
U.S. Managers



Nigerian Managers



Kenyan Managers



Japanese Managers

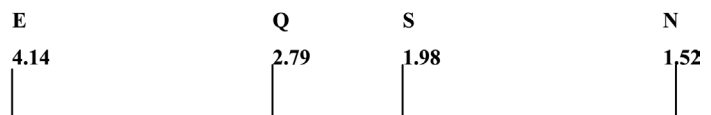


Figure 2. Test of linearity of the measurement scale: the usefulness of “the” marketing concept

categories. This is suggested by the fact that all the English-speaking respondents were not able to distinguish between the third and fourth categories, whereas the Japanese respondents were unable to distinguish between the second and third categories. Moreover, among the English-speaking respondents, the fourth category was not preserved, in contrast to the third category which was not preserved among the Japanese respondents. Second, the average scores among Japanese respondents were much higher than 4 for the higher scale category (i.e. upper scale bias) while the average scores among the US respondents on the higher scale category were less than 4 (i.e. lower scale bias). Interestingly, the African respondents averaged between 3 and 4, suggesting less upper scale bias than the Japanese but more than the US respondents.

Suggested lack of measurement equivalence

The optimally scaled results shown in Table III clearly suggest that lack of measurement equivalence may be a problem by virtue of differences in some of the values of the transformed data. In essence, the four-point Likert-type scale was transformed into different scale values across the four countries. For example, US respondents used the first three scales with the fourth (highest) scale category collapsed into the third. Japanese respondents also

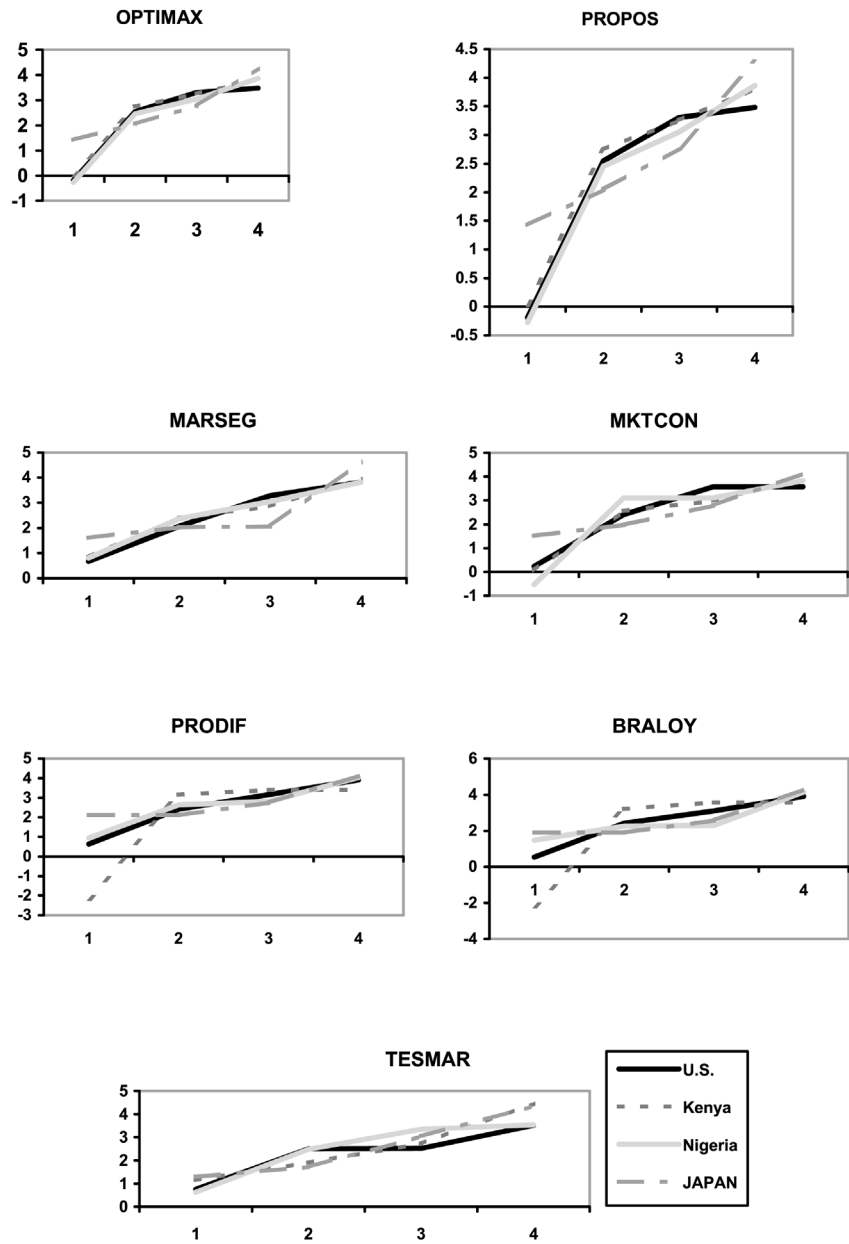


Figure 3. A plot of optimal transformed values against original scale values or individual study variables

used three scale categories, but with third scale category collapsed into the second. Nigerian and Kenyan respondents were similar to US respondents, in that most, but not all, of the fourth scale categories were collapsed into the third scale category. These findings suggest that the English-speaking respondents did not consider the fourth category to be important, whereas the Japanese respondents considered the third category to be unimportant. They also suggest that the Japanese respondents interpreted the scale with an upward scale bias (i.e. several items of the transformed values had greater than 4, whereas the English-speaking respondents interpreted the scale with a lower scale bias (i.e. several of the items had transformed values of less than 1, and ten had values of 4). Clearly, the scale lacks scalar equivalence across linguistic groupings but not LDC and IC groupings.

LDC respondents more inconsistent

Diagnosis for scoring format consistency

The results for the second diagnostic dealing with cross-national consistency in the use of the scoring format are shown in Table III. It appears that, for all four-scale categories, the LDC respondents were more inconsistent than their counterparts from the USA and Japan.

First, US respondents consistently interpreted the first scale category as less than 1 for all seven concepts. The Japanese respondents interpreted it as 1 for all concepts, except for V2. In contrast, the Kenyan respondents used three different scoring formats for the first category for all seven concepts: items V1, V2, V3 and V4 received scores below 1, items V5 and V6 were scored as 2s, and item V7 was scored as 1. Similarly, the Nigerian respondents used two scoring formats (1 for V6 and less than 1 for the remaining items).

Second, US respondents scored all second scale category items as 2, whereas Japanese respondents used two scoring formats: 2 for items V1, V3, and V5, and less than 1 for items V2, V4, V6, and V7. Nigerian respondents also used two scoring formats: 2 for all items except V4, which was scored as 3. Kenyan respondents used three scoring formats: 2 for items V1 and V4, 2 for items V2 and V5, and 1 for item V7.

Third, in the third scale category, all respondents used two scoring formats. The US respondents used 3 for all items except item V7, which was scored as 2; the Japanese respondents used two for all items except item seven, which was rated a 3; and the Nigerian and Kenyan respondents used 3 for all items except V5 and V6 among Nigerians and V3, V4, and V7 among Kenyans.

Fourth, similar patterns of inconsistent scoring are evident for the fourth scale category. US respondents again consistently scored all seven items as 3, whereas the Japanese consistently scored all items as 4. However, Nigerian respondents used two scoring formats: 3 for items V1, V2, V3, and V4, and 4 for items V5, V6, and V7. Similarly, Kenyans scored all items as 3 except for V7. This scoring pattern further suggests that the African respondents tended to be less capable of using this fairly simple ordinal scale consistently than did their IC counterparts.

Some cross-national variations exist

Diagnosis for evidence of improvement in the order of scale categories

The third measurement equivalence check involved ascertaining the degree of preservation of the ordering of the scale categories or parallel spacing after optimal scaling. If this were an interval scale, the orders of the scale categories should be equidistant. Parallel spacing indicates a monotonic relationship and, therefore, attainment of scalar equivalence (Perreault and Young, 1980). Figure 2 shows the relationship between the optimal scale categories plotted against the original scale categories for one item, V4, the marketing concept. It shows one piece of evidence of improvement in measurement equivalence: the order of the scale categories is preserved in all countries after transformation, as would be expected of an ordinal scale. However, some cross-national variations exist in the ordering of the scale, which suggests the lack of equidistance or interval properties. In addition, the unequal spacing of scale categories seems to be unrelated to LDC and IC groupings. The scores of Japanese respondents again reflect upward scale bias, whereas those of the US respondents reflect a lower scale bias. The scores of respondents from the two African countries are generally in between those of US and Japanese respondents, as was found in the scoring format consistency test.

Data exhibit some modest degree of structural similarity

Diagnosis for evidence of monotonic relationship in overall data structure
The fourth and most compelling motivation for the optimal transformation is to improve measurement equivalence in the overall data. While individual scale items may lack measurement equivalence, the existence of measurement equivalence in the data structure means that cross-national comparison can still be performed. This is indicated by the existence of essential features of a monotonic linear relationship among scale categories across countries after optimal transformation. To check for improvement in monotonic relationships in these data, the optimally transformed response categories were plotted against the original response categories for all study variables, as shown in Figure 3. The relationships among the seven marketing concepts all point in the same general direction, an indication of essentially linear relationships. In addition, prior observations about the cross-national nature of scale bias are reflected in Figure 3. Except for PRODIF (product differentiation), Japanese respondents have the highest score on the high scale category, whereas US respondents' scores on the high scale category are among the lowest. The African respondents' scores generally fall in between those of Japanese and US respondents.

Overall, however, the relationships among all seven study variables are essentially monotonic cross-nationally. Therefore, the four-country data exhibit some modest degree of structural similarity after the optimal transformation.

Improvement in overall data structure based on factorial structural similarity

To ascertain the degree of improvement of the validity and reliability of the data following optimal transformation, a principal component analysis was performed on the optimally scaled data. The results of these analyses are summarized in Table IV. A visual inspection showed only one poor factor loading (V4) in the optimally transformed values, implying that the optimally transformed values provide discriminate validity (Didow *et al.*, 1983). However, the factor loading varies slightly across countries

Table IV also shows that the reliability scores on the PRINQUAL transformed values are more than acceptable ($\alpha = 0.94$). The improved clarity of the optimally scaled factor structure is an indication of construct validity following optimal transformation of the marketing concept scale (Didow *et al.*, 1985).

Multi-group measurement invariance test

Multiple-group LISREL test for measurement equivalence

The final and strongest diagnostic test for improvement in construct measurement equivalence after optimal scaling is a multigroup measurement invariance test (Mullen, 1995; Singh, 1995; Steenkamp and Baumgartner, 1998). The multi-group invariance test is a series of hierarchical models beginning with a "configurable invariance model" (Steenkamp and Baumgartner, 1998), or a finding that the measurement items all load on the same factors across each group. The next stage tests the configurable model with an additional constraint, equivalence of the factor loadings (Singh, 1995). The final stage includes the configurable invariance, factor loading invariance, and the invariance of the error variances. The equivalence across the four countries is tested using multiple group structural equation modeling with LISREL 8.30 (Joreskog and Sorbom, 1996). One construct, core-marketing concepts, was tested and discussed using the results of a three stage hierarchy as summarized in Table V. As in the previous validity

Country	n	Optimal scale values			Comm ^a
		Factor 1	Factor 2	Comm ^a	
USA	185	V1	0.109	0.794	0.642
		V2	0.420	0.659	0.610
		V3	0.462	0.509	0.472
		V4	0.093	0.771	0.604
		V5	0.861	0.082	0.748
		V6	0.731	0.184	0.569
		V7	0.621	0.205	0.428
Percentage of total variance			29.6	28.6	58.2
Eigenvalues			2.07	2.00	
Kenya	52	V1	0.541	0.582	0.631
		V2	0.824	0.194	0.718
		V3	0.263	0.682	0.535
		V4	0.181	0.812	0.693
		V5	0.967	0.075	0.940
		V6	0.971	0.037	0.945
		V7	0.121	0.590	0.363
Percentage of total variance			42.4	26.7	69.1
Eigenvalues			2.97	1.86	
Nigeria	67	V1	0.752	0.303	0.657
		V2	0.812	0.319	0.761
		V3	0.332	0.760	0.688
		V4	0.840	-0.165	0.734
		V5	0.657	0.456	0.640
		V6	0.598	0.295	0.444
		V7	0.047	0.854	0.732
Percentage of total variance			40.4	26.0	66.4
Eigenvalues		2.83	1.82		
Japan	118	V1	0.686	0.411	0.640
		V2	0.625	0.465	0.608
		V3	0.863	-0.078	0.751
		V4	0.631	0.273	0.472
		V5	0.433	0.666	0.632
		V6	0.014	0.858	0.737
		V7	0.255	0.758	0.639
Percentage of total variance			32.3	31.7	64.0
Eigenvalue			2.26	2.22	
Cronbach alpha					0.94

Note: ^aCommunality estimates

Table IV. Principal component analysis of optimally transformed scale values

Hypotheses	χ^2	df	$\chi^2_{diff.}$	df	GFI	CFI	TLI
<i>Form</i>							
Raw data	138.78	56			0.94	0.88	0.82
Transformed data	146.08	56			0.94	0.94	0.92
<i>Equal loadings</i>							
Raw data	154.88	77	16.10	19	0.93	0.89	0.88
Transformed data	157.69	77	11.61	19	0.93	0.95	0.94
<i>Equal loadings and errors</i>							
Raw data	176.48	98	22.60	22	0.92	0.89	0.90
Transformed data	240.17	98	82.48	22	0.93	0.91	0.92

Table V. Multigroup measurement models fit statistics

Transformed data had better fit indices than raw data

checks in this study, the emphasis is placed on the comparison between the original and the optimally scaled data, with separate discussions of each data in the hierarchy.

The first step, a test of the equal factor pattern, shows that the baseline model for the raw data has an adequate fit according to the goodness of fit index (GFI), comparative fit index (CFI), and the Tucker Lewis index (TLI). The results are presented in Table V. Although the overall χ^2 for the multi-group test was significant, this test can be affected by sample size and has limitations (Bollen, 1989). The transformed data had better fit indices than the raw data. All fit indices, GFI, CFI, and TLI were above the recommended threshold of 0.90, signifying a good fit.

The second step of the multi-group invariance test is to test the pattern and equality of loadings on the latent construct. The raw data again had an adequate fit. The χ^2 difference tested between this model and the baseline model was insignificant, with $\chi^2(19) = 16.10$. The fit for the transformed data was extremely good, with all indices above 0.90. The χ^2 difference between this model and the baseline was insignificant, with $\chi^2(19) = 11.61$. Again, the fit for the transformed data was better than for the raw data.

The final step of the invariance test examines the previous constraints plus one additional constraint, the invariance of the error variances of the manifest measures. This is achieved by comparing models for no constraints on the error variances with a model that contains the variances as equal. The raw data had an adequate fit according to the fit indices. As in the previous steps, the results in Table V show that the transformed data had a better fit.

Optimally scaled data are preferable

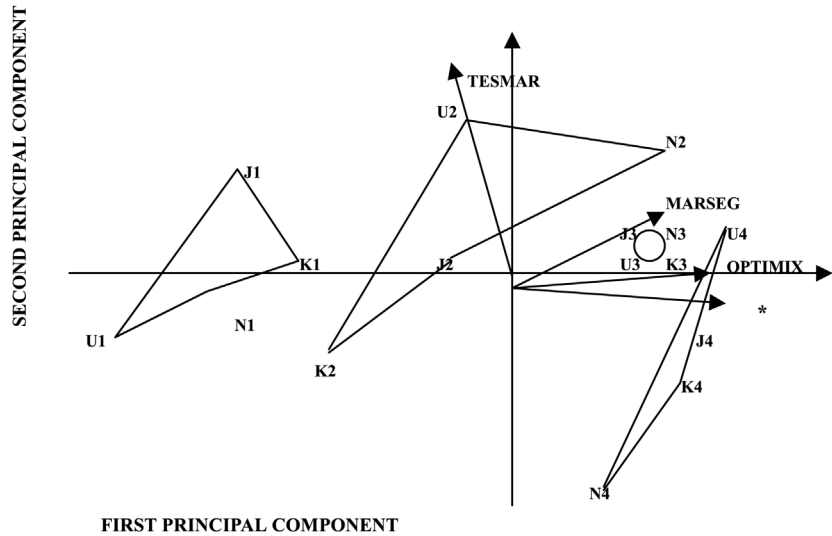
Although the hierarchy tests seem to show both data sets have adequate fit, the transformed data fit the model much better than the raw data. These findings are critical to this study in two ways. First, they are evidence that the original data set contained minimum fit requirements for measurement equivalence and could be used for cross-national comparison. Second, the optimally scaled data are preferable to the original data because they consistently perform much better on all of the measurement invariance tests, thereby allowing cross-cultural comparison.

Multidimensional comparison of applicability of marketing concepts

Having established the improved measurement equivalence of the overall data (i.e. high measurement invariance and a more monotonic relationship than the original data), it is now justifiable to use the PRINQUAL data for cross-national comparison. This can be done along two or three dimensions of the data using the PRINQUAL MDU routine. The main criterion is the extent to which the study's variables are separated cross-nationally. Comparing the relationships between the original response values and the transformed response values captures this measurement property.

Multidimensional analysis is considered useful

In Figure 4, we present the MDU results. This type of multidimensional analysis is considered useful for cross-national comparison of constructs because of its rich geometric properties. Therefore, we briefly explain these properties before interpreting the MDU results. In the spatial map for the PRINQUAL results (Figure 4), multidimensional relationships are represented by both points and vectors in an informative geometric map that represents a relationship only as points. The values in the figure represent the first letter of the country name (e.g. U for the USA, N for Nigeria, and so forth), the perceived usefulness score for the marketing concepts (e.g. 1, 2, and so forth). Thus, country responses are represented as points and scores



*PROPOS, MARCON, PRODIF, and BRALOY are overlapped

Notes: Countries are identified by the first letter of their names:
 U= United States, J = Japan, N = Nigeria, and K = Kenya

Figure 4. PRINQUAL map for the perceived usefulness of marketing concepts

for the marketing concepts are represented as vectors. Relationships among the groups therefore can be captured geometrically. This implies that the closeness of the points and the direction of the vectors reflect the cross-national usefulness of marketing concepts. Vectors pointing toward a specific marketing variable reveal the importance of that marketing concept across countries.

The first two dimensions account for 98.6 percent of the explained variance. Most of the concepts and techniques that received high applicability ratings are represented on the high end of the first dimension. They include the marketing concept (MARCON), brand loyalty (BRALOY), product differentiation (PRODIF), and product positioning (PROPOS). In addition, the first dimension accounts for nearly 82 percent of the explained variance. Thus, it seems to represent the concept of “primary applicability”. The second dimension also improves interpretation of the spatial relationship among response categories and study variables by separating the concepts that were evaluated as less applicable than the rest. These concepts include optimization of the marketing mix (OPTMIX), market segmentation (MARSEG), and test marketing (TESMAR) or those of a “secondary applicability”.

Pattern of applicability emerges

In addition, a notable pattern of applicability emerges in Figure 4 that suggests identical perceptions about the usefulness of the seven marketing concepts across the four countries by virtue of four relationships among the variables. First, all countries are represented in the four clusters. Second, when corresponding points of all four countries are linked, none of the clusters overlap. Third, the country response vectors point toward the same marketing concept measure. For example, with the exception of N4, every country response for the “extremely useful” (4) and “quite useful” (3) response categories is pointed toward BRALOY, PRODIF, MARCON and PROPOS. These concepts are highly loaded on the first dimension as well.

Two-stage hierarchy of perceived usefulness

Fourth, preferences for the first response category, “not at all useful” (1), seem to be orthogonal to preferences for the “extremely useful” (4), and “quite useful” (3) categories, as indicated by the scores pointing toward the first dimension and away from scores for the 3s and 4s.

These findings suggest a two-stage hierarchy of perceived usefulness of marketing concepts that transcends national and IC versus LDC boundaries. These geometric relationships do not correspond to LDC and IC groupings; thus, there is support for marketing scholars who argue that marketing knowledge is equally relevant in all countries regardless of the level of economic development (Drucker, 1958; Emlen, 1958; Kinsey, 1982).

Discussion

Summary of objectives and results of the diagnosis of measurement properties

The general objective of this research was to examine how measurement equivalence issues affect comparative marketing analysis involving LDC samples. We note that most of the concerns about measure quality are pronounced in analysis involving LDCs. Yet discussions of marketing measures in international marketing not even began to consider field conditions in Africa. We have argued that these environmental differences can lead to different types of measurement errors across LDCs and ICs

Item-by item diagnosis

We also argue that the effect of LDC field conditions, such as small samples and the desirability of qualitative measures, augur well for the use of optimal scaling. In addition, optimal scaling facilitates an item-by item diagnosis of the sources of measurement equivalence and can potentially determine the uniqueness of various types of measurement errors across countries.

In our comparative analysis of the applicability of core marketing concepts and techniques, we first examined the limitations of existing guidelines for designing cross-national marketing instruments by conducting a four-nation survey of the perceived applicability of core marketing concepts and techniques across Japanese, US, Nigerian, and Kenyan firms. Then, we examined the raw data and found evidence suggesting a lack of construct equivalence in the study measures between a developing country such as Kenya, an industrialized country such as Japan, and the USA on the other. On optimally scaling the data and examining the applicability of the items we found that:

- LDC respondents from the two African countries tended to use the scoring format of the scale inconsistently; and
- respondents from the two ICs, Japan and the USA, tended to interpret the scale with an upper and a lower scale bias, respectively.

Subsequent tests consistently confirmed this last finding.

Optimal scaling did improve measurement equivalence

The LISREL invariance tests make it clear that optimal scaling did improve measurement equivalence, as did the following three results of the optimal scaling:

- (1) evidence of essential linear properties (i.e. scale categories pointed in the same general direction);
- (2) none of the circles linking each scale category overlapped; and
- (3) the measurement items were orthogonal.

Two-stage hierarchy

A comparative analysis based on the PRINQUAL transformed data revealed a two-stage hierarchy of applicability of marketing concepts and techniques that transcends LDC and IC groupings. These findings underscore the importance of optimal scaling in diagnosing and improving the measurement properties of qualitative, cross-national data involving small samples.

Implications for LDC marketing and cross-national marketing research

The major contribution of this study lies in the specific argument and finding that optimal scaling with PRINQUAL can handle measurement equivalence threats that are particularly common in LDC samples: small sample and qualitative measures. Overall, the study provides evidence that suggests that scoring format consistency may be a major problem for LDC respondents, whereas upper and lower scale bias may be a more serious threat for US and Japanese respondents. But optimal scaling improved the overall measurement properties and enhanced more accurate comparison and conclusions than would be possible with the original data. Although the link between these different types of measurement errors and the environmental differences was not examined directly, the different types of errors associated with the LDC and IC samples confirms our argument that some types of measurement errors may be unique to LDCs and ICs and that greater attention should be paid to these differences when comparing marketing phenomena. More careful anchoring of the scales should be used to reduce differences in interpretation of constructs in both the pre-test and post data collection phases (Zhang and Dadzie, 1991). Given the limited sample sizes of most pretest, optimal scaling should be used to validate these measures in the pre-test stage as well as the post data collection stages. However, LISREL should be used to test overall measurement equivalence after optimal transformation of the data, as has been demonstrated in the present study.

Hierarchy that is cross-national in nature

In addition, this study suggests that variations in the applicability of core marketing concepts and techniques do not correspond to LDC and IC groupings, but rather, they reflect a hierarchy that is cross-national in nature. This conclusion suggests the need for greater attention to the internal or organizational environment rather than organization's country environment when addressing the applicability issue. Overall, the universal applicability claim long made by scholars such as Drucker (1958) and Emlen (1958) is supported when the measure are optimally scaled.

Significant implications for practitioners

Implications for marketing practitioners

The study has some significant implications for practitioners interested in standardizing their global marketing programs. The results point to the cross-national applicability of marketing's core concepts when dealing with LDC field market conditions. Despite the differences in market conditions, attention should be paid to the level of applicability within the organization rather than the level of economic development in a given country. It appears that the more fundamental concepts such as the marketing concept, product differentiation, brand loyalty, and product positioning, can be applied readily and in both LDC and IC market conditions within an organization before much less commonly applied concepts and techniques such as optimization of the marketing mix and test marketing.

Limitations and research implications

Although the results in this study support optimal transformation as a useful procedure for diagnosing cross-national data, given its ability to handle

Scales with finer shades or categories pose greater scoring problems

Caution should be exercised

qualitative data and small sample sizes, several limitations of the procedure and analysis should be considered in its application. Many of the limitations may also serve as directions for further research.

First, optimal scaling is an exploratory technique that cannot be used to test hypotheses about relationships. Thus, the amount of error in many estimates in the study cannot be determined. Hence performing invariance test with LISREL is recommended. Our study demonstrated the benefit of such a combined application.

Second, the ordinal scale is only one of several measurement scales marketing scholars use in comparative marketing studies. This scale's measurement equivalence may be less vulnerable to measurement errors because of fewer metric assumptions than other scales have. It is thus conceivable that our four-point Likert-type, ordinal scale may have posed fewer measurement problems than would an interval scale. Thus, the optimal scaling results in this study may not be generalized to other study results based on other measurement scales, such as interval scales. Therefore, it should be of interest to know if one type of scale is more or less susceptible to measurement errors in samples drawn from LDCs and ICs. Such additional research should consider the influence of individual background characteristics on scoring reliability. Although our four-point, ordinal scale may be described as simple, the LDC respondents from the two African countries were consistently inconsistent in scoring the study variables. It is thus likely that scales with finer shades or categories pose greater scoring consistency problems for LDC respondents with little experience in marketing. This issue should be investigated further.

Third, some caution should be exercised in generalizing this study's conclusion regarding the comparative usefulness of core marketing concepts and techniques. Although our findings lead us to conclude that marketing concepts are no more or less useful to firms in LDCs than they are to firms in ICs, more research is needed. In particular, research is needed on organizational conditions that enhance implementation of the marketing philosophy and its underlying techniques in LDCs.

References

- Akaah, I.P., Dadzie, K.Q. and Riordan, E.R. (1988), "Applicability of marketing concepts and management activities in the Third World: an empirical investigation", *Journal of Business Research*, Vol. 16 No. 2, pp. 133-48.
- Arndt, J. (1972), "Temporal lags in comparative retailing", *Journal of Marketing*, Vol. 34, October, pp. 40-5.
- Asante, M.K. and Asante, K.W. (1985) (Eds), *African Culture*, Greenwood, Westport, CT.
- Bhalla, G. and Lin, L.Y.S. (1987), "Cross-cultural marketing research: a discussion of equivalence issues and measurement strategies", *Psychology and Marketing*, Vol. 4 No. 4, pp. 275-85.
- Bollen, K.A. (1989), *Structural Equation Modeling with Latent Variables*, John Wiley & Sons, New York, NY.
- Brisling, R.W., Lonner, W.J. and Robert, M. (1973), *Cross-cultural Research Methods*, John Wiley & Sons, New York, NY.
- Currie, L. (1968), "Marketing organization for underdeveloped countries", in Reed Moyer and Hollander, S.C. (Eds), *Markets and Marketing in Developing Economies*, R.D Irwin, Homewood, IL.
- Dadzie, K.Q. (1989), "Demarketing strategies in shortage market environments", *Journal of the Academy of Marketing Science*, Vol. 17, Summer, pp. 157-65.
- Davis, H.L., Douglas, S.P. and Silk, A.J. (1981), "Measure unreliability: a hidden threat to cross-national marketing research?", *Journal of Marketing*, Vol. 45 No. 2, pp. 98-109.

- de Boor, C. (1978), *A Practical Guide to Splines*, Springer Verlag, New York, NY.
- de Leeuw, J. (1986), *Regression with Optimal Scaling of the Dependent Variable*, Department of Data Theory, The University of Leiden, Leiden.
- Dia, M. (1991), "Development and cultural values in sub-Saharan Africa", *Finance and Development*, Vol. 28 No. 4, December.
- Didow, N.M. Jr, Perreault, W.D. Jr and Williamson, N.C. (1983), "A cross-sectional optimal scaling analysis of the index of consumer sentiment", *Journal of Consumer Research*, Vol. 10, December, pp. 339-47.
- Didow, N.M., Jr, Keller, K.L., Barksdale, H.M. Jr and Franke, G.R. (1985), "Improving measure quality by alternating least squares optimal scaling", *Journal of Marketing Research*, Vol. 22, February, pp. 30-40.
- Dixon, D.F. (1981), "The role of marketing in early theories of economic development", *Journal of Macromarketing*, Vol. 1 Fall, pp. 19-26.
- Douglas, S.P. and Craig, C.S. (1983), *International Marketing Research*, Prentice-Hall, Englewood Cliffs, NJ.
- Drucker, P. (1958), "Marketing and economic development", *Journal of Marketing*, Vol. 22, January, pp. 252-9.
- Dun and Bradstreet International (1990), *Principal International Business: The World Marketing Directory*, Dun and Bradstreet International, New York, NY.
- Emlen, W.J. (1958), "Lets export marketing know-how", *Harvard Business Review*, Vol. 36, November-December, pp. 70-6.
- Fisher, D.G., MacKinnon, D.P., Anglin, M.D. and Thompson, J.P. (1987), "Parental influences on substance use: gender differences and stage theory", *Journal of Drug Education*, Vol. 17 No. 1, pp. 69-86.
- Fisher, R.A. (1938), *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh, pp. 285-98.
- Green, R.T. and White, P.D. (1976), "Methodological considerations in cross-national consumer research", *Journal of International Business Studies*, Vol. 21 No. 4, pp. 597-622.
- Hofstede, G. (1991), *Cultures and Organizations: Software of the Mind*, McGraw-Hill, New York, NY.
- Hosley, S. and Wee, C.H. (1988), "Marketing and economic development: focusing on the less developed countries", *Journal of Macromarketing*, Vol. 4, Spring, pp. 43-53.
- Hotelling, H. (1933), "Analysis of complex statistical variables into principal components", *Journal of Educational Psychology*, Vol. 24, pp. 417-41.
- Irvine, S.H. and Carroll, W.K. (1980), "Testing and assessment across cultures: issues in methodology and theory", in Triandis, H.C. and Lonner, W. (Eds), *Handbook of Cross-Cultural Psychology: Methodology*, Allyn & Bacon, Boston, MA, pp. 181-244.
- Joreskog, K.G. and Sorbom, D. (1996), *LISREL 8: User's Reference Guide*, Scientific Software International, Chicago, IL.
- Kinsey, J. (1982), "The role of marketing in economic development", *European Journal of Marketing*, Vol. 16 No. 6, pp. 64-77.
- Kruskal, J.B. (1964), "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis", *Psychometrika*, Vol. 29, pp. 1-29.
- Kuhfeld, W.F. and Young, F.W. (1989), "PRINCIPALS versus OSMOD: a comment on Saito and Otsu", *Psychometrika*, Vol. 54, December, pp. 755-6.
- Kuhfeld, W.F., Sarle, W.S. and Young, F.W. (1985), "Methods of generating model estimates in the PRINQUAL macro", *SAS Users Group International Conference Proceedings*, SAS Institute, Inc., Cary, NC.
- Marsh, H.W. and Hocevar, D. (1985), "Applications of confirmatory factor analysis to the study of self-concept: first- and higher order factor models and their invariance across groups", *Psychological Bulletin*, Vol. 97, pp. 562-82.
- Mazrui, A.A. (1990), *Cultural Forces in World Politics*, James Curry, London.
- Moyer, R. (1964), "The structure of markets in developing economies", *MSU Business Topics*, Vol. 12, Autumn, pp. 43-60.
- Mullen, M. (1995), "Diagnosing measurement equivalence in cross-national research", *Journal of International Business Studies*, Vol. 26 No. 3, pp. 573-95.

- Parameswaran, R. and Yaprak, A. (1987), "A cross-national comparison of consumer research measures", *Journal of International Business Studies*, Vol. 8 No. 1, pp. 35-49.
- Perreault, W.D. Jr and Young, F.W. (1980), "Alternating least squares optimal scaling: analysis of nonmetric data in marketing research", *Journal of Marketing Research*, Vol. 17, February, pp. 1-13.
- Rostow, W.W. (1960), *The Stages of Economic Growth*, Cambridge University Press, Cambridge.
- Singh, J. (1995), "Measurement issues in cross-national research", *Journal of International Business Studies*, Vol. 26 No. 3, pp. 573-96.
- Steenkamp, J.E.M. and Baumgartner, H. (1998), "Assessing measurement invariance in cross-national consumer research", *Journal of Consumer Research*, Vol. 25, June, pp. 78-90.
- Van de Vijier, and Poortinga, Y.N. (1982), "Cross-cultural generalizations and universality", *Journal of Cross-Cultural Psychology*, Vol. 13, pp. 387-408.
- Wyser, R.S. and Srull, T.K. (1989), *Memory and Cognition in its Social Context*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 317-49.
- Young, F.W. (1981), "Quantitative analysis of qualitative data", *Psychometrika*, Vol. 46, December, pp. 357-88.
- Young, F.W., Takane, Y. and de Leeuw, J. (1978), "The principal components of mixed measurement level multivariate data: an alternating least squares method with optimal scaling features", *Psychometrika*, Vol. 43, June, pp. 279-81.
- Zhang, L. and Dadzie, K.Q. (1991), "Developing more meaningful and accurate measurement models in global marketing research: an empirical illustration", in Gilly, M. (Ed.), *Enhancing Knowledge Development in Marketing*, American Marketing Association Educators Proceedings, pp. 610-11.

Further reading

- Alder, N.J. (1983), "A typology of management studies involving culture", *Journal of International Business Studies*, Vol. 14, Fall, pp. 29-47.
- Aulakh, P.S. and Kotabi, M. (1993), "Assessment of theoretical and methodological developments in international marketing", *Journal of International Business Studies*, Vol. 1 No. 2, pp. 5-28.
- Ayal, I. and Hornik, J. (1986), "Foreign source effects on response behavior in cross-national mail surveys", *International Journal of Research in Marketing* (Netherlands), Vol. 3 No. 3, pp. 157-67.
- Bartels, R. (1983), "Is marketing defaulting its responsibilities?", *Journal of Marketing*, Vol. 47, Fall pp. 32-5.
- Betchel, G.G. and Ofir, C. (1988), "Aggregate item response analysis", *Psychometrika*, Vol. 53, March, pp. 93-107.
- Carroll, J.D. (1972), "Individual differences and multidimensional scaling", in Shepard, R.N., Romney, A.K. and Nerlove, S.B. (Eds), *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, Vol. 1, Seminar Press, New York, NY.
- Cavusgil, T.S. and Yavas, U. (1984), "Transfer of management know-how to developing countries: an empirical investigation", *Journal of Business Research*, Vol. 12, pp. 35-50.
- Cunningham, W., Cunningham, I.C. and Green, R.T. (1977), "The ipsative process to reduce response set bias", *Public Opinion Quarterly*, Vol. 4, Fall, pp. 379-94.
- Dawson, S. and Dickson, D. (1988), "Conducting international mail surveys: the effects of incentives on response rates with an industry population", *Journal of International Business Studies*, Vol. 19, Fall, pp. 491-6.
- Dholakia, N. and Dhokhia, R.R. (1982), "Marketing in the emerging world order", *Journal of Macromarketing*, Vol. 2, pp. 47-56.
- Ghiselli, E.E., Campbell, J.P. and Zedeck, S. (1981), *Measurement Theory for the Behavioral Sciences*, W.H. Freeman and Company, New York, NY.
- Gifi, A. (1990), *Nonlinear Multivariate Analysis*, John Wiley & Sons, New York, NY.
- Guttman, L. (1941), "The quantification of a class of attributes: a theory and method of construction", in Horst, P. (Ed.), *The Prediction of Personal Adjustment*, Social Science Research Council, New York, NY.

- Hoffman, D.L. and Perreault, W.D. Jr (1987), "Consumer preference and perception", in Young, F.W. and Hamer, R.M. (Eds), *Multidimensional Scaling: History, Theory, and Applications*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 199-218.
- Hughes, M.A. and Garrett, D.E. (1990), "Intercoder reliability estimation approaches in marketing: a generalizability theory framework for quantitative data", *Journal of Marketing Research*, Vol. 27, May, pp. 185-95.
- Kent, D.P. and Young, F.W. (1986), "New developments in alternating least-square SAS procedures", in Kuhfeld, W.F. and Thomas, C. (Eds), *SAS Users Group International Conference Proceedings*, SAS Institute, Inc., Cary, NC.
- Ross, C.A. (1984), "The marketing task in the Third World", *Journal of Marketing Education*, Vol. 6, Spring, pp. 20-7.
- Rust, R.T. and Cooil, B. (1994), "Reliability measures for qualitative data theory and implications", *Journal of Marketing Research*, Vol. 31, February, pp. 1-14.
- Saito, T. and Otsu, T. (1988), "A method of optimal scaling for multivariate ordinal data and its extensions", *Psychometrika*, Vol. 53, March, pp. 5-25.
- SAS Institute Inc. (1992), "The PRINQUAL procedure", *SAS/STAT User's Guide*, Version 6, 4th ed., SAS Institute, Inc., Cary, NC, pp. 1265-323.
- Scibrowsky, J.A. and Collins, R.H. (1990), "The brand manager's statistical package: SPSS's categories module", *Journal of Personal Selling & Sales Management*, Vol. 10, Summer, pp. 97-100.
- Slater, C.C. (1968), "Marketing processes in developing Latin American society", *Journal of Marketing*, Vol. 32, July, pp. 50-5.
- SPSS Inc. (1990), *SPSS Categories*, SPSS Inc., Chicago, IL.
- Srinivasan, V. and Basu, A.K. (1989), "The metric quality of ordered categorical data", *Marketing Science*, Vol. 8 No. 3, pp. 205-30.
- Takane, Y., Young, F.W. and de Leeuw, J. (1977), "Nonmetric individual differences in multidimensional scaling: an alternating least squares method with optimal scaling features", *Psychometrika*, Vol. 42, March, pp. 7-67.
- Tenenhaus, M. and Vachette, J.L. (1977), "PRINQUAL: un programme d'analyse en composantes principales d'un ensemble de variables nominales ou numeriques", *Les Cahiers de Recherche*, No. 68, CESA, Jouy-en-Josas, France.
- Tenenhaus, M. and Young, F.W. (1985), "An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and other methods for quantifying categorical multivariate data", *Psychometrika*, Vol. 50 March, pp. 91-119.
- Van de Geer, J.P. (1993), *Multi-variate Analysis of Categorical Data: Theory (Series Vol. 2) and Applications, Vol. 3*, Sage Publications, Newbury Park, CA.

