ABSTRACT

Comparability is a key issue in cross-cultural marketing research. In order to provide precise comparisons, marketing researchers must establish equivalence of constructs and measures across cultures. This article reviews the major types of equivalence and argues that no single measurement strategy can deal with all types of equivalence. A recommendation is made for a multistrategy approach that addresses most, if not all, types of equivalence.

Cross-Cultural Marketing Research: A Discussion of Equivalence Issues and Measurement Strategies

Gaurav Bhalla
Burke Institute

Lynn Y. S. Lin
Burke International Research, Inc.

Comparability of respondent data is critical issue confronting marketing researchers, regardless of whether the research is conducted in one country or in a number of countries simultaneously. To illustrate, Nestle may discover through research conducted in Belgium that

packaging, namely, yellow lids and red labels, plays a key role in creating an optimum brand image. Should Nestle generalize from its Belgian experience and standardize its packaging to create a unified global image? Alternatively, consider a major office equipment manufacturer, such as NCR or HP, conducting an image study in four of five major international markets with the objective of developing internal evaluation standards and training programs. Clearly, there is a need to compare the responses of the various countries to find out what works and what does not.

It is this need for comparability that gives rise to a host of methodological issues concerning research design and data collection. Traditional approaches in marketing research aiming to establish equivalence across cultures have typically devoted attention to just one aspect of it, namely, linguistic equivalence. Most tests on marketing and marketing research have numerous examples of mistakes that have been made when words and phrases have been used without appropriate translation. While it is true that words have different meanings in different cultures, cross-cultural equivalence is more than just linguistic equivalence.

Accordingly, this article has three major objectives: (1) to discuss the various aspects of the data collection process where equivalence must be established, (2) to present some empirical evidence demonstrating variation associated with cultural contexts, and (3) to suggest strategies that marketing researchers can employ to establish equivalence in the various phases of the data collection process.

TYPES OF EQUIVALENCE

A number of writers in social psychology and comparative sociology have written about the concept of cross-cultural equivalence (Anastasi, 1982; Berry, 1979; and Lonner, 1981). While a number of different labels and modifiers have been used, the following categories seem to overlap.

Construct Equivalence

Do psychological and marketing constructs, such as locus of control, innovative behavior, cognitive complexity, and brand loyalty have the same meaning across different cultures? For instance, in many countries product categories may not exhibit the same brand proliferation as in North America. Furthermore, in many countries the brands first introduced, or the dominant brands, have virtually become generic labels, and symbolize the entire product category,

Nestle's Milkmaid. Lever's Surf, and P & G's Pampers being good examples of this. Consequently, perspectives on brand loyalty that focus on the sequence of brand purchases in a multibrand context are likely to fall short. Construct equivalence has three distinct components: conceptual equivalence, functional equivalence, and category equivalence (Douglas and Craig, 1983).

Conceptual equivalence is concerned with the interpretation that people in different cultures place on marketing stimuli, product, and behavior. To illustrate, consumer promotion events involving markdowns and cents-off are a regular feature of the American marketing landscape. In other countries, however, especially in Asia and Africa, where markets tend to be sellers' markets rather than buyer's markets, price discounting behavior of manufacturers may be viewed with suspicion. Consumers may perceive the product to be of poor quality, defective, or a poor seller.

Functional equivalence refers to the similarity of goals of behavior across different cultures. In many European countries, as well as in a large number of developing countries, bicycles are predominantly a means of transportation rather than recreation. Similarly, whereas eating out is taken lightly and occurs highly frequently, to satisfy routine gastronomic and social needs in some cultures, it is a symbol of conspicuous consumption, and hence to be taken very seriously in others. The implications of this for marketing researchers are that definition of competitive sets of products, motives, and values — and needs satisfied by certain well-defined behaviors — cannot be blindly transplanted from one context to the other.

Category equivalence refers to the way that objects, stimuli, and behaviors are grouped. Socioeconomic and demographic variables typically used to classify individuals and their responses may not be equivalent across cultures. To illustrate, postgraduate refers to Masters and Ph.D. types of degrees in the commonwealth countries, whereas in the United States it refers to the work done after those degrees have been obtained. The category of principal shopper as either the Female Head of Household (FHH) or the Male Head of Household (MHH) may be inappropriate in countries where routine, everyday shopping is done by domestic servants. Moreover, the category "household" itself is subject to variation across cultures, involving as it does joint families or extended families in some countries.

Operationalization Equivalence

Operationalization specifies the transition from theory to measurement. Marketing researchers often collect psychographic data in segmentation, image, and product usage studies to arrive at a more

vivid consumer profile. While "leisure" may be operationalized via sailing, watching Saturday afternoon baseball, or lying in the sun in the United States and Canada, it may not be possible to do so in countries that do not play baseball, or where sailing is beyond the reach of the average consumer. Furthermore, not all countries enjoy 24-hour T.V. transmission and lying in the sun may represent erratic behavior in countries where people are born with a permanent brown hue.

Closely connected with operationalization equivalence is *item equivalence* (Hui and Triandis, 1985). Item equivalence is a more concrete and microlevel perspective, and presupposes both construct and operationalization equivalence. For item equivalence to be established the construct should be measured by the same instrument. To illustrate, a marketing researcher interested in public policy issues related to drinking and driving may employ the constructs of Alienation and Responsibility to explain social drinking. However, if the constructs are measured using separate instruments, one for each culture, the direct comparison of test scores is misleading and illegitimate (Hui and Triandis, 1985).

Scalar Equivalence

This type of equivalence is established if the other, more abstract, types of equivalence have been attained, and if it can be demonstrated that two individuals from separate cultures with the same value on some hypothesized variable, say cognitive complexity, will score on the same level on the same test. While this type of equivalence is ideal for quantitative cross-cultural comparisons, it is the most difficult to achieve for two reasons (Hui and Triandis, 1985). First, not all countries express equal familiarity with all scales and scaling procedures. In the United States a five- or seven-point scale may suffice, while in other countries a 10- or 20-point scale may be more appropriate (Douglas and LeMaire, 1974). Second, the issue of score equivalence is even more complex and difficult (Van de Vijier and Poortinga, 1982). Cultures differ in their response-set characteristics, such as social desirability, acquiescence, and evasiveness, which influence response scores.

Take for instance the future-purchase-intention scale, which is an integral part of most new product forecasting models, including BASES. Does a similar top box or top two boxes score on an intentions-to-purchase scale reflect a similar likelihood of purchase. The need for adjusting the intentions-to-purchase scores is well reflected in the marketing literature (Lin et al., 1980; Mullet and Karson, 1985). The BASES experience in Europe has been that different

TABLE 1
Impact on Responses of Different Scaling Procedures

	N = 434		N = 412
Definitely Would Buy	15%	Definitely Would Buy	12%
Probably Would Buy	33%	Very Probably Would Buy	22%
Might or Might Not Buy	23%	Probably Would Buy	20%
Probably Would Not Buy	16%	Might or Might Not Buy	21%
Definitely Would Not Buy	13%	Probably Would Not Buy	11%
		Very Probably Would Not Buy	6%
		Definitely Would Not Buy	8%

TABLE 2
Purchase Intent—Concept Test
(Household Cleaning Product)

	Italy	France	West Germany
Definitely Would Buy	36%	22%	15%
Probably Would Buy	49%	42%	48%
Might or Might Not Buy	5%	21%	13%
Probably Would Not Buy	3%	9%	14%
Definitely Would Not Buy	7%	6%	10%

countries require different conversion rate coefficients for identical products and product categories (Lin, 1984). Among European countries it has been observed that West Germans are least likely to overstate their intentions, whereas the Spanish and the Italian are most likely to overstate them. Countries such as the U.K., Belgium, and Switzerland lie somewhere in between these two extremes. Tables 1 and 2 present some data from the BASES data bank on the facets of item equivalence just discussed. Table 1 presents an example of the impact of different scaling procedures on responses. Table 2 presents cross-country data for France, Germany, and Italy for a new product test done for a commonly used household cleaner. It may be added that the product is doing equally well in all three countries on key measures such as brand development index (BDI), net continuation, and market share. Table 3 presents samples of purchase intent conversion rates for a product category among three countries and regions within each country. These are expressed in form of index to show the relative differences among and within these countries.

TABLE 3
Examples of Approximate Purchase Intention Conversion Rate
By Country and Regions—Product Category X
(Expressed in Terms of Index)

Country	Belgium		U.S.	Italy			
Region	A	B		A	B	C	D
Conversion Rate Index	100	60	65	50	60	30	15

Linguistic Equivalence

Linguistic equivalence refers to both the spoken and the written language forms for cooperative use in questionnaires, interviewing, classifying respondents' doubts and probing open-ended responses. Of all types of equivalence, the problems associated with linguistic equivalence have received the most consistent attention (Werner and Campbell, 1970; Sechrest et al., 1972; Brislin et al., 1973; and Brislin, 1976). Cooper et al. (1982), after extensive pretesting, decided to use differently worded dependent-variable questions in a study on public drinking in Canada and the U.S. The U.S. version read: "In the past 12 months, how often have you been to a bar or a tavern?" The Canadian version read: "In the past 12 months, how often, if at all, have you gone to a club or public place where drinks are served?" The Canadian version also used clearing, which defined a club or public place as taverns, bars, night clubs, lounges, beverage rooms, or private clubs. No clearing was needed or used for the U.S. version. The authors are correct in stating that whereas the usage of "bar and tavern" in the Canadian questionnaire would have given the appearance of comparability, in cross-cultural research it is the understanding that is crucial. The goal of linguistic equivalence is this commonality in understanding.

CROSS-CULTURAL MEASUREMENT STRATEGIES

Several strategies that attempt to demonstrate or improve different types of equivalence are considered in this section.

Nomological Validation

This method borrows heavily from Cronbach and Meehl's (1955) construct validation approach, the idea being that if a construct has

the same meaning across different cultures then it must also exhibit the same network of relationships with other constructs—antecedent and dependent. Evidence of similar networks then suggests that instruments used in the validation process are cross-culturally applicable and equivalent (Hui, 1982). The similarity of networks can be statistically evaluated using structural equation modeling (Bagozzi, 1980; and Joreskog and Sorbom, 1984).

As this is a global approach, it is insensitive to issues concerning item and scalar equivalence. Also, as long as similarity in networks is exhibited; construct equivalence and, hence, conceptual, functional, and categorical equivalence can be assumed. However, when causal modeling approaches fail to demonstrate similarity, it is not easy to say which of the constructs are nonequivalent across cultures.

Item Structure Congruence

These methods aim to address the issue of operationalization equivalence. They presuppose construct equivalence and confine their testing to the components of the construct, with the objective of demonstrating the same relationships between components across cultures. The two statistical procedures usually employed most often to demonstrate internal structure congruence are factor analysis, including confirmatory factor analysis, and multidimensional scaling (MDS).

Factor analysis approaches have been used cross-culturally by Mayberry (1984) to develop equivalent attitude scales across American and Japanese cultures, and by Kumar and Waisanen (1979) to explain the effects of education and individual modernity on the adoption of innovations.

MDS applications have been reported by Hui and Triandis (1983). Hispanic and non-Hispanic Navy recruits were tested on an abbreviated locus of control scale. Data were analyzed using a variation of the INDSCAL model (Caroll and Chang, 1970). The analysis identified five dimensions, three of which were shared by the two ethnic groups. The other two were declared to be cross-culturally nonequivalent.

Item-Response Methods

Item-response theory (IRT) (Lord, 1977, 1980) is an internally generated test to determine whether similar responses originating in different cultures reflect the same intensity of a latent trait, such as self-esteem, occupational and recreational attitudes, risk aversion, and thrift. According to IRT, for items to be cross-culturally equi-

valent they must exhibit the same item characteristic curve (ICC), which represents the probabilities of responding to an item in a specified manner at different levels of the latent trait to be measured. ICCs obtained from different cultures are tested statistically for differences. Similar ICCs demonstrate item and scalar equivalence. This method is especially useful in Attitude, Trial, and Usage (ATU) type of studies, in which the marketing researcher is interested in consumers' agree/disagree responses to a number of items concerning their attitudes, values, and behaviors.

Translation Methods

As indicated earlier in the article, at a minimum, the research instrument must be translated effectively from one language to another to ensure proper meaning. Brislin (1976, 1980) has proposed a number of methods such as back-translation, bilingual and committee approach, decentering, and pretests. Effective translation can also help overcome lack of item equivalence. In all fairness, though, a very accurate translation is not always foolproof (Werner and Campbell, 1970). Practical problems associated with data gathering, such as interviewer training, motivation of respondents, and other nonsampling errors, can and do pose problems. For instance, in a number of Middle-East countries, due to religious and social reasons, women respondents can be interviewed only by women investigators. In addition, respondent recruitment may often be the product of personal affiliations. Such uncontrollable factors introduce biases and reduce comparability on otherwise-precise research instruments.

Direct Comparisons and Statistical Testing

An intuitive and popular approach is to administer the same set of questions across a variety of cultures and to use simple t-tests or MANOVA to test differences between cultures. Such an approach assumes that the construct being measured exists in both cultures and is equally operationalized. Furthermore, the use of statistical tests of difference assumes that scalar equivalence also exists. Because of the number of questionable assumptions that such a method makes, an approach that is often recommended is to generate questionnaires and scales with twice or thrice the amount of items or questions actually required. Through extensive pretesting the only questions retained are those that have the same meaning and operationalization in different cultures. Very often, the use of native and foreign experts is made to generate this shorter list of questions.

Miller (1972) has demonstrated that this method has both convergent validity, between native and foreign exports, and discriminant validity, between countries that are different, for example, England and Peru.

DISCUSSION

Comparability is a prerequisite for valid inferences in cross-cultural research. This can be attained by adopting universals from pertinent disciplines or by demonstrating the equivalence of marketing concepts and data across different cultural groups. This article has identified some of the more important types of equivalence that must be established in cross-cultural research. Clearly, it is not easy to make a watertight distinction between the various types of equivalence mentioned in the article. Some, such as construct equivalence, are more abstract and others, such as scalar equivalence, are more concrete. Currently available research methods do not permit a marketing researcher to address simultaneously all types of equivalence on this abstract-concrete continuum. Different methods for assessing cross-cultural equivalence have different emphases and are not mutually replaceable. In addition, different methods make different assumptions about the various types of equivalence. A sampling of some of the major methods used in cross-cultural assessment has also been provided. By no means is it an exhaustive listing. Rather than search for the one best method, or assume the naive view, as indicated by the direct-comparison approach, a multistrategy approach is preferable. Some methods can address the more abstract issues of equivalence, while others can tackle the more concrete types. A good example of a multistrategy approach is provided by Hui and Triandis (1982), who used multidimensional scaling, factor analysis, and nomological validation in demonstrating that aspects of the locus of control construct are generalizable among American mainstream and Hispanic respondents. While it may not be possible to demonstrate all types of equivalence, it is only through attempts that tackle multiple equivalence that precise quantitative measurements can be made across cultures. Finally, even though the article has used the words "country" and "culture" interchangeably, the same country can exhibit dramatically different cultures, India, Canada, Belgium, and Italy being good examples. More precisely, the prescriptions made above are equally valid even within the context of a single-country study when subjects are drawn from different cultural pools. To illustrate, the BASES new-product-forecasting model uses different coefficients of adjustment on the future purchase intention scale, de-

pending on the area of Italy in which the study is conducted — northwest, northeast, central, or southern.

REFERENCES

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.

Bagozzi, R. (1980). *Causal models in marketing.* New York: Wiley.

Berry, J. W. (1979). On cross-cultural comparability. *International Journal of Psychology, 4,* 119–28.

Brislin, R. (1976). *Translation: Application and research.* New York: Wiley/Halsted.

Brislin, R. (1980). Translation and content analysis of oral and written materials. In H. C. Triandis and J. W. Berry (eds.), *Handbook of cross-cultural psychology: Methodology (Vol. 2).* Boston: Allyn & Bacon.

Brislin, R., Lonner, W. J., and Thorndike, R. M. (1973). *Cross-cultural research methods.* New York: Wiley.

Carroll, J. D., and Chang, J. (1970). Analysis of individual difference in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika, 35,* 283–319.

Cooper, R., Neuman, B., Okraku, I. O. (1982). Public drinking in Canada and U.S.A. *International Journal of Comparative Sociology, XXIII,* Nos. 3-4, September-December, 204–215.

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Douglas, S., and LeMaire, P. (1974). Improving the quality and efficiency of lifestyle research. XXV ESOMAR Congress, Budapest, 1974.

Douglas, S. P., and Craig, C. S. (1983). *International marketing research.* New Jersey: Prentice-Hall.

Hui, C. H. (1982). Measurement in cross-cultural psychology: A review and comparison of strategies for empirical research. Unpublished manuscript, University of Illinois.

Hui, C. H., and Triandis, H. C. (1983). Multistrategy approach to cross-cultural research: The case of locus of control. *Journal of Cross-Cultural Psychology, 14,* 65–83.

Hui, C. H., and Triandis, H. C. (1985). Multidimensional scaling and item response theory. In R. Diaz-Guerrero (ed.), *Cross-cultural and national studies in social psychology.* North-Holland.

Joreskog, K. G., and Sorbom, D. (1984). *LISREL VI,* Scientific Software Inc.

Kumar, K., and Waisanen, F. B. (1979). Education, adoption of innovation, and individual modernity. *International Journal of Comparative Sociology, XX,* 3–4, September-December, 213–223.

Lin, L., McKenna, W., Rhodes, R., and Wilson S. (1980). New product analysis and testing. In K. J. Albert (ed.), *Handbook of business problem solving.* New York: McGraw-Hill, pp. 4–13 to 4–27.

Lin, L. (1984). Does intent predict behavior? Presented at AMA's 15th Annual Attitude Research Conference at Hilton Head Island.

Lord, F. M. (1977). A study of item bias, using item characteristic curve theory. In Y. H. Poortinga (ed.), *Basic problems in cross-cultural psychology.* Amsterdam: Swets & Zeitlinger.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum.

Mayberry, P. W. (1984). Analysis of cross-cultural attitudinal scale translation using maximum likelihood factor analysis. Paper presented at the meeting of the American Educational Research Association. New Orleans.

Miller, D. C. (1972). Measuring cross-cultural norms. *International Journal of Comparative Sociology, XII,* 3-4, September-December, 201-216.

Mullet, G. M., and Karson, M. J. (1985). Analysis of purchase intent scales weighted by probability of actual purchase. *Journal of Marketing Research, XXII,* February, 93-96.

Sechrest, L., Fay, T., and Zaidi, H. (1972). Problems of translation in cross-cultural research. *Journal of Cross-Cultural Psychology, 3,* 41-56.

Van de Vijier, F. J. R., and Poortinga, Y. H. (1982). Cross-cultural generalization and universality. *Journal of Cross-Cultural Psychology, 13,* 387-408.

Werner, O., and Campbel, D. R. (1970). Translating working through interpreters, and the problem of decentering. In R. Narroll and R. Cohen (eds.), *A handbook of method in cross-cultural anthropology.* New York: Natural History Press, pp. 398-420.