

**A QUALITATIVE ASSESSMENT
OF THE CROSS-CULTURAL EQUIVALENCY OF
SERVICE QUALITY AS MEASURED BY THE SERVQUAL INSTRUMENT**

Gene W. Murdock, Department of Management & Marketing, University of Wyoming, P.O. Box 3275,
Laramie, WY 82071-3275, (307-766-6159 or 766-3124).

ABSTRACT

The wide acceptance in the United States of concept of service quality and its measurement via the SERVQUAL scale leads one to the question of the cross-cultural applicability of the concept and its measurement with SERVQUAL. This paper reports on the literature concerning the cross-cultural equivalence of marketing research and then assesses the equivalence of service quality and its measurement with SERVQUAL for the Chinese, Indian and Norwegian cultures. The qualitative findings (depth, group interviews and translation) suggest that both the concept of service quality and its measurement with SERVQUAL can achieve acceptable levels of conceptual, functional, category, metric or scaler, calibration and translation equivalence in the cultures studied.

A QUALITATIVE ASSESSMENT OF THE CROSS-CULTURAL EQUIVALENCY OF SERVICE QUALITY AS MEASURED BY THE SERVQUAL INSTRUMENT

Introduction

The increased emphasis on the global market place has brought with it an increased need for cross-cultural research (Albaum and Peterson, 1984; Marbeau 1990). Because we all have our own cultural bias, it is important to examine such fundamental concepts as service quality to determine if this construct that has been so well developed and tested in the United States (Zeithaml, Parasuraman and Berry, 1990) can be used in cross-cultural research.

Previous Research

It is normally assumed that some comparability of results is desirable in cross-cultural research and the search for comparability has focused on achieving equivalence (see Table 1), primarily (and perhaps wrongly) translation equivalence (Hui and Triandis 1983; Anderson 1967; Douglas and Craig 1983; Green and White 1976; Berry 1969; Sechrest, Fay and Zaidi 1972; Sekaran 1983; Bhalla and Lin 1987; Sood 1989; Barker and Kaciak 1992). However, years ago Webster (1966) cautioned against assuming the same methodology and

TABLE 1

**Types, Definitions and Examples of
Equivalencies in Cross-Cultural Research**

Type of Equivalency	Definition	Example
Construct	the understanding of psychological and marketing constructs across cultures. Three distinct parts.	the term Innovative may have a different meaning to Asians than Americans
a) conceptual	the similarity of meaning that people in different cultures place on market stimuli, product, and behavior	hot milk-based drinks are seen as relaxing in the UK but stimulating in Thailand
b) functional	similar activities that have different functions in different societies cannot be used for comparative purposes	bicycles are seen as recreational in the U.S., but as a means of transport in Europe
c) category	the way that objects, stimuli, and behavior are grouped	post graduate means masters and phd degrees in the commonwealth, but in the U.S. it refers to work done after those degrees have been achieved
Measure or Operationalization Equivalence	specifies the transition from theory to measurement	leisure is operationalized by watching golf in the U.S., but maybe not elsewhere
a) Metric or Scaler Equivalence	comparability of both scoring procedures and responses to a given measure	a 5-10 point scale may be appropriate in the U.S., while a 10-20 scale is appropriate in another country
b) Calibration Equivalence	this equivalence includes the units of measure and interpretation of colors and symbols	the color white has different meanings in different cultures
c) Translation Equivalence	verbal and non-verbal communication in a way comparable in two or more languages	
Sampling Equivalence	comparability of different samples, can be broken down into relevant respondent and representative sample	drawing samples with the same characteristics from both a French and English speaking population
a) Relevant Respondent	how relevant the sample is to the survey	if soft drink preference is tested, sample should contain soft drink consumers
b) Representative Sample	does the sample accurately represent the population under investigation	in a study of students, sample should mirror the UW population
Types of Translational Equivalency		
a) Vocabulary	equivalence in the words used in the two translations	there is no word for 'feminine' in Tagalog
b) Idiomatic	equivalence in the translation of idioms with the goal of equivalent meanings	Galit bukan is a idiom in Tagalog which means angry volcano
c) Grammatical-Syntactical	equivalence in parts of speech	verbs forms and gerunds
d) Experiential	the real things in life or things which are familiar to both cultures	florist is an American term unknown in others parts of the world
e) Conceptual	ensuring that the concepts used in the two translations are equivalent	homosexual is a concept familiar to America, but not in the Philippines

**Types of Data Collection
Equivalency**

a)	Instrument	refers to the instrument design and the appropriateness of the survey technique	use of the telephone as the survey instrument may not reach representative cross-section
b)	Temporal	the timing of the research were the surveys conducted simultaneously, sequentially or independently	self explanatory
c)	Contextual/Situational	interaction between respondent and interviewer, includes interviewer ability and physical location of the interview	respondent may be more at ease in own home or office
d)	Response Style	how culture affects response style	differences in culture may evoke extreme responses or non-response behavior

Sources: Compiled from Bhatia and Lin (1987), Douglas and Craig (1983), Hui and Triandis (1985), Meyer (1978), Sechrest, Fay and Zaidi (1972)

translation equivalence yield comparable results. This theme is echoed by other authors who point to the emic/etic dilemma (Douglas and Craig 1983). The conclusions of these sometimes simultaneous pleas for equivalence and criticisms of equivalence seem to be that cross-cultural research is a series of trade-offs. Maximizing one or two forms of equivalence may tend to minimize others. Thus, there is a need to trade comparability for local adaptation (decentralization) to adequately capture cultural differences. Other approaches to achieving comparability such as multiple measures and/or statistical techniques (normalizing, standardizing) are frequently too costly or difficult to be viable (Douglas and Craig 1983; Hui and Triandis 1983).

Notwithstanding the considerable debate about whether, how, and when to achieve equivalence in cross-cultural research, this study examines the cross-cultural equivalency of the concept of service quality as operationalized by the SERVQUAL scale (Zeithaml, Parasuraman and Berry 1990). Other issues in cross-cultural research, such as validity and reliability (Parameswaran 1987; Jaffe and Nebenzahl 1984; Davis, Douglas and Silk 1981), measurement problems (Sood 1989) and the nature of the scale (Barker and Kaciak 1992) will not be dealt with directly.

Methodology

The overall methodology consists of two phases a qualitative and a quantitative phase. The first phase, the qualitative phase, was a series of depth interviews with paid student respondents representing Norwegian, Indian, and Chinese cultures. While it is acknowledged that students are different than the general population, the paying of the students for their participation was done to help offset any duress that may have been

present. Additionally, the nature of the topic (service quality) is most relevant to the segments of the cultures most likely to be represented by students, i.e., marketing researchers studying service quality would likely target upper income, well educated respondents as these are the heavy users of most for-profit services. The purpose of the depth interviews was to establish in general terms the presence or absence of each type of equivalency.

The second phase will consist of having a systematic random sample of students from each culture rate the equivalencies of the construct of service quality for specific types of services. The purpose of this phase will be to obtain a more quantitative measure of each form of equivalency. Additionally, the qualitative phase of the research revealed the need to anchor any assessment of equivalence to particular services. The quantitative phase provided specific services as anchors.

Finally, in addition to the overall methodology, a forward and backward translation along with decentering were used to assess the ability to achieve translation equivalency. While the forward and back translation approach is recommended in order to achieve translation equivalency (Sechrest, Fay and Zaidi 1972) the problems associated with using bilinguals (Anderson 1967) and the need for decentering (Choudhry 1986) were recognized. The emphasis in achieving translation equivalency was on achieving equivalent meaning rather than word for word back translation equivalency.

Results Qualitative Phase

The qualitative phase of the research provided good insight into the ability of marketers to study service quality cross-culturally. For the cultures studied (Norwegian, Chinese, Indian) there was general agreement that the construct of service quality was similar across cultures. Specifically, each of the major components of service quality as identified by Zeithaml, Parasuraman and Berry (1990): tangibles, reliability, responsiveness, assurance and empathy, were felt to exist in the respondent's cultures and were considered important components of service quality. This finding from the depth interviews supports both conceptual and construct equivalency, which are perhaps the most important equivalencies for cross-cultural researchers.

While this general result held, there were specific aspects of the concept of service quality that differed somewhat from the U.S. concept. Most notably, all

three groups of foreign students felt that there were a lower set of expectations for service quality in their home country. As one of the Norwegian respondents said:

"We really don't expect much. The expectation is just for them not to be rude. The best service should probably be 'good' or 'acceptable' not 'excellent.' Although there is a Norwegian word for excellent in the dictionary, it would probably not be used in conjunction with service quality. For example at soccer games, people don't expect the waits in line to be minimized. They expect long lines."

Other expressions of this same set of lower expectations also existed in the Chinese and Indian groups' responses. For example, in the Chinese group it was mentioned that although they would like "excellent" service, you just can't get it. Lines are expected at the postal service and at train stations. You have to "know someone" to get good service in trading stocks or buying plane tickets. Bribes are commonly used in both India and China and there is thus, a direct "price" and service quality relationship. Both of these groups also expressed similar views in that anything associated with the "government" meant lower service quality. Perhaps one exception existed in the government run banks in India. Although the Indian banks provide fewer services than the U.S. counterparts, it was felt that they provided better quality service than other public sector services (transportation most notably).

An important conceptual nuance was discovered in the Indian group. It was noted that although the equipment in most post offices in India would be out-dated by U.S. standards, it would be considered up-to-date by Indian standards. This illustrates a point made in all three cultural groups. If the quality of service depends on expectations, then members of cultures who have been exposed to higher service quality in other countries will have different expectations (foreign students for example) while members of the culture who have not been so exposed will perceive the service quality differently because their base comparison point is different. Because the SERVQUAL scale measures both expectations and perceived service, the different baseline expectations may need to be identified and considered in the overall assessment of service quality. Specifically, it may be necessary to form two groups, those with relatively high expectations and those with lower expectations, and examine the reasons for the differences in expectations.

Table 2 summarizes the findings from the qualitative study of the cross-cultural equivalency of service quality among Chinese, Norwegians and Indians as represented by foreign students. In sum, it appears that functional equivalency exists in that the underlying constructs of service quality carry across the cultures. One concern here however, is that service quality may not have the same significance as it does in the U.S.. Conceptual equivalency also seems to exist in that service quality occurs and is expressed in similar terms. However, expectations may be lower (see metric or scale equivalency). Category equivalency is not a major problem as long as the distinction is made between public and private services. Given functional, conceptual and category equivalence seem to exist to an acceptable degree, construct equivalence is at least feasible for the SERVQUAL scale (Douglas and Craig 1983).

TABLE 2
Evaluation of SERVQUAL on
Construct and Measure Equivalencies

Type of Equivalency	Equivalency of SERVQUAL in Culture		
	Indian	Chinese	Norwegian
Construct			
a) Functional	High	High	High
b) Conceptual	Moderate	Moderate	High
c) Category	Moderate	Moderate	Moderate
Measure			
a) Calibration	Moderate	Moderate	Moderate
b) Translation	High	Moderate	High
c) Metric	Moderate	Moderate	Moderate

High = SERVQUAL is or can be made culturally equivalent (this does not mean made the same, decentering is assumed)

Moderate = SERVQUAL is unlikely to be culturally equivalent, but an acceptable level of equivalency can be achieved

Low = SERVQUAL cannot be made to achieve an acceptable level of equivalency

Measure equivalence (operationally defining the construct) requires calibration, translation, and metric or scale equivalence (Douglas and Craig 1983). Calibration equivalence deals with both non-verbal stimuli and measures of weight, distance, volume, money, etc. Calibration equivalence does not appear to be an issue in most aspects of the SERVQUAL scale. One problem may be in the instructions to the SERVQUAL scale. In Chinese, to select a "number" that reflects feelings on a scale does not make sense. Thus, this instruction was translated to select an "answer." From a calibration standpoint, the emphasis on intervals is lost and thus, the Likert scale is clearly ordinal for data analysis

purposes. The scale is technically ordinal in any case, but it is common practice to treat the data as interval data. This practice may be less appropriate with the Chinese version of the questionnaire.

Translation equivalence has received much attention in the literature (Anderson 1967, Douglas and Craig 1983, Sechrest, Fay and Zaidi 1972) and it proved somewhat difficult in the case of the SERVQUAL scale. At first, the approach of iterative forward and back translation was attempted, but the problems identified by Choudhry (1986) were experienced in that forward translations that could be accurately back translated lost the original meanings and frequently required the use of words that would not be commonly used in the culture (this was true for Chinese, Indian and Norwegian translations). The decision was then made to emphasize Choudry's (1986) four measures of quality in translation:

1. Maximize clarity and comprehensiveness
2. Minimize changes in meaning
3. Emphasize ability to respond correctly
4. Pretesting (yet to be done)

Thus, the translated versions of SERVQUAL do not back translate exactly into the English version. Instead they are decentered to capture the original's meaning and to use the common terminology of the culture. Finally, these translations have not been pretested so it is assumed that they can be improved on the above criteria. Clearly, translation equivalency is possible for the SERVQUAL scale.

However, problems and concerns encountered during the translation process were numerous. One problem for all three versions was obtaining a non-hand-written copy (word processing). The Norwegian version was easiest as only a few characters are different from English and these characters are generally found in common word processing software packages. The Chinese version was more difficult because, although the conversion software existed, it was time consuming to use. The reason is that the software provides several Chinese character alternatives for each English word entered. The operator must then select the most appropriate character. This was time consuming even though a hand-written Chinese translation aided the process. The Indian version proved to be the most difficult to obtain in word processed form.

Other translation concerns include the fact that in both India and Norway there exists a national language

promoted by the government. While the governments are trying to force the national languages to be adopted (Hindi and New Norwegian) there is some backlash among the populace. Modified Hindi and plain Norwegian were used for the initial translations. If one was to attempt to administer the SERVQUAL scale to government entities (employees) it might be necessary to develop pure Hindi and New Norwegian versions. A second translation consideration in both Norway and India is that much of the target market for services in both of these countries is bilingual. This means an English version, combined English and native language version, or mixed version of the SERVQUAL questionnaire are all feasible. Cultural sensitivity, targets for research and comparability of results are all concerns when considering these additional translation options.

The Chinese translation provided some concerns unique from the Norwegian and Hindi versions. Many of the English words used in SERVQUAL such as "employee" were difficult to capture in Chinese because the employee/boss relationship is not as clear in China as it is in the U.S.. Other concerns in the Chinese version include the fact that simplified Chinese can be read by Taiwanese, but not as comfortably as by mainland China respondents. Additionally, the capturing of meaning versus word for word translation is more important in Chinese. Word for word translation to achieve forward and back translation equivalence was virtually impossible in Chinese.

Overall, initial translated versions of the SERVQUAL questionnaire were obtained and it is felt that the versions achieved translation equivalency to the extent they removed the influence of language on the results. They are, however, non-pretested versions and will certainly need pretesting and further refining. There may be a need for multiple versions of the questionnaire in the cultures studied depending on the types of respondents being studied. Additionally, the concerns of Sechrest, Fay, and Zaidi (1972) for cultural versus linguistic translation (experiential equivalence), idiomatic equivalence and grammatical-syntactical equivalence of the SERVQUAL scale need to be further researched. It appears that English could serve as a "carrier language" (Sechrest, Fay, and Zaidi 1972).

The next type of measure equivalence is metric or scale equivalence. This type of equivalence is concerned both with the specific scale used (7 point Likert and 100 point rating) and response bias. None of the groups studied in this qualitative research felt that the scales themselves were a problem. But, all three felt that there

would be a response bias, i.e., lower expectations of service quality are likely. Thus, the vocabulary of the questions may need to be changed. This is both a conceptual equivalence and metric/scale equivalence problem. The conceptual part of the problem is the lack of "excellent" service. The word excellent probably needs to be softened to "good" or "acceptable." This problem, if not corrected, may manifest itself as a metric or scaling problem in the sense that the seven point Likert scale is effectively reduced to a three or four point scale because respondents will never agree that they expect (and probably not that they received) "excellent" service. The heavy use of the word "excellent" in the SERVQUAL scale creates problems, both conceptual and metric/scaling, for the cross cultural researcher in Norway, India and China.

Summary and Conclusions

This paper reports on the initial qualitative phase of a two phase (qualitative and quantitative) study designed to assess the cross-cultural equivalency of the SERVQUAL scale.

In the qualitative phase reported in this paper, three groups of foreign students (Indian, Chinese and Norwegian) were recruited (paid) to participate in a approximately two-hour discussion concerning the ability of the SERVQUAL scale to measure service quality in their cultures. While the author recognizes the limitations associated with qualitative research and the use of foreign students as representative of the cultures involved, he also feels that the results obtained represent an important assessment of the cross-culture usefulness of SERVQUAL.

The results support that within reasonable limitations construct (functional, conceptual, and categorical) equivalence and measure (calibration, translation, and metric or scale equivalence) can be achieved. The ability to achieve translation equivalence was assessed not only through group discussions, but also through an initial decentered translations of the SERVQUAL instrument.

The second phase of research will involve a more quantitative evaluation of equivalencies and will add sampling equivalence (Douglas and Craig 1983) to the types of equivalencies assessed. In conclusion the SERVQUAL instrument for measuring service quality appears to hold promise for researchers based on a qualitative assessment of construct and measure equivalence across Chinese, Indian and Norwegian

students.

REFERENCES

- Albaum, G., and R.A. Peterson (1984). Empirical Research in International Marketing: 1976-1982. Journal of International Business Studies. 15:161-173.
- Anderson, R.B.W. (1967). On the Comparability of Meaningful Stimuli in Cross-Cultural Research. Sociometry. 30:124-136.
- Barker, A.T., E. Kaciak (1992). An Examination of the Equivalence of Four Measurement Scales in Cross-Cultural Research. Journal of International Consumer Marketing. 4 (3):77-93.
- Berry, John W. (1969). On Cross-Cultural Comparability. International Journal of Psychology. 4:119-128.
- Bhalla, G., and L.U.S. Lin (1987). Cross-Cultural Marketing Research: A Discussion of Equivalence Issues and Measurement Strategies. Psychology & Marketing. 4:275-285.
- Boddewyn, J.J. (1981). Comparative Marketing: The First Twenty-Five Years. Journal of International Business Studies. 12:61-75.
- Choudhry, Y.A. (1986). Pitfalls in International Marketing Research: Are You Speaking French Like a Spanish Cow? Akron Business and Economic Review. 17:18-28.
- Davis, H.L., S.P. Douglas, and A.J. Silk (1981). Measure Unreliability: A Hidden Threat to Cross-National Marketing Research? Journal of Marketing. 45:98-109.
- Douglas, S.P., and C.S. Craig (1983). International Marketing Research. Englewood Cliffs, New Jersey: Prentice-Hall Inc.
- Elder, J.W. (1976). Comparative Cross-National Methodology. Annual Review of Sociology. 2:209-230.
- Green, R.T. and E. Langeard (1975). A Cross-National Comparison of Consumer Habits and Innovator Characteristics. Journal of Marketing. 39:34-41.
- Hui, C.H. and H.C. Triandis (1983). Multistrategy Approach to Cross-Cultural Research: The Case of Locus of Control. Journal of Cross-Cultural Psychology. 14:65-83.

Jaffe, E.D., and I.D. Nebenzahl (1984). Alternative Questionnaire Formats for Country Image Studies. Journal of Marketing Research. 21:463-71.

Marbeau, Y. (1990). Towards a Pan-European Economic Status Scale. Marketing and Research Today. 18:180-184.

Mayer, C.S. (1978). Multinational Marketing Research: The Magnifying Glass of Methodological Problems. European Research. 77-84.

Parameswaran, R. and A. Yaprak (1987). *A Cross-National Comparison of Consumer Research Measures. Journal of International Business Studies. 18:35-49.

Sechrest, L., T.L. Fay and S.M.H. Zaidi (1972). Problems of Translation in Cross-Cultural Research. Journal of Cross-Cultural Psychology. 3:41-56.

Sekaran, U. (1983). Methodological and Theoretical Issues and Advancements in Cross-Cultural Research. Journal of International Business Studies. 14:61-73.

Sood, J. (1989). Equivalent Measurement in International Marketing Research: Is it Really a Problem? Journal of International Consumer Marketing. 2:25-41.

Webster, L.L. (1966). Comparability in Multi-Country Surveys. Journal of Advertising Research. 6:14-18.

Wind, Y., and S.P. Douglas (1982). Comparative Consumer Research: The Next Frontier? Management Decision. 20:24-35.

Zeithaml, V.A., A. Parasuraman and L.L. Berry (1990). Delivery Service Quality. New York: The Free Press.