Equivalence of survey data:

Relevance for international marketing

Hester van Herk

Ype H. Poortinga

Theo M.M. Verhallen

Hester van Herk is Assistant Professor of Marketing, Vrije Universiteit Amsterdam, Ype H. Poortinga is part-time Professor of Psychology, at Tilburg University, the Netherlands and the University of Leuven, Belgium. Theo M.M. Verhallen is Professor of Marketing and Marketing Research at Tilburg University.

Corresponding author: Hester van Herk. Address: Vrije Universiteit Amsterdam, Dept. IML, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands. Tel. +31 20 4446064; Fax: +31 20 4446005; E-mail address: hherk@feweb.vu.nl.

Equivalence:

The importance for international marketing

Keywords International marketing research, equivalence, bias

Abstract

The equivalence or comparability of survey data collected across countries is regarded as a basic issue in culture-comparative research. Despite the importance of bias, most culture-comparative studies in marketing and business research interpret differences at face value. Moreover, in commercial marketing research the issue is mostly neglected. In this article two general approaches found in the literature are presented, namely, an approach focussing more on the research process and an approach focussing more on measurement invariance. An integrated framework is presented as well as a case study showing it is useful to integrate both approaches to improve decision making in international marketing.

Introduction

Faced with maturing markets and stiffening competition, industries are forced to rethink their strategies. In this, internationalisation of activities is a main strategy. Several multinationals have interests in at least thirty counties (Mitra and Golder, 2002), and a company such as Unilever sells its Lipton tea in as many as 110 countries (Unilever, 2003, www.unilever.com). As companies are increasingly engaging in global trade, global marketing has become vital. Cultural, economic, legal, and geographic differences between the home market and the markets of other countries have to be taken into account. Such differences also imply that people may react differently to marketing efforts.

Unfortunately, little empirical research is available on customs, habits, attitudes, and reactions to marketing efforts in different regions. Therefore, companies tend to collect marketing information themselves (or have this done for them) in order to make well-founded decisions. The resulting growing need for international marketing research information is shown in the worldwide turnover for commercial opinion and market research; in 2001, this was 17 billion Euro, up 5.8% over the previous year (ESOMAR, 2002). In comparison with 1980, there was a seven-fold increase.

When making international comparisons data should have the same meaning across those countries, because inequivalent or biased information leads to ambiguous or even erroneous conclusions. Therefore, the equivalence or comparability of data collected across countries is regarded as a key issue (e.g., Douglas and Craig, 1983; Hui and Triandis, 1985a; Sekaran, 1983; Singh, 1995; Van de Vijver and Leung, 1997a,b). Despite its importance, the equivalence of data is usually not examined (Aulakh and Kotabe, 1993; Malhotra et al., 1996; Sin et al., 1999; Sin et al., 2001) and most culture comparative studies do not address equivalence issues. This lack of attention for issues of culture is not limited to research on

marketing; in commercial marketing studies, on which marketing decisions tend to be based, equivalence issues are also hardly addressed. The reasons for this negligence are not clear, but the analysis of equivalence in data is not a simple matter. In addition, lack of clarity in the literature has added to the complexity. In this article we try to present an integrated approach. We give an overview of terminology used in different publications and we distinguish two major approaches to equivalence. One approach focuses more on the whole research process, whereas the other approach focuses on data analysis. The objective of this article is to provide a framework for establishing equivalence that may help reduce the confusion, and better integrate measures that can be taken to avoid or deal with bias in data. First, a short overview is given of equivalence approaches in the literature. Second, we attempt to integrate the various approaches, introducing different levels of equivalence and linking these to sources of bias in the research process. Finally, we discuss what kinds of inferences are justified if there is evidence supporting various levels of equivalence.

Approaches to equivalence

In most general terms there are two approaches to equivalence. One is a psychometric approach in which characteristics of parameters in measurement models are tested for invariance across countries. If certain conditions of invariance are satisfied certain comparisons are deemed valid (e.g., Steenkamp and Baumgartner, 1998). The second approach, that started earlier in international marketing research, has been summarized by Douglas and Craig (1983). They started from a series of problems encountered in cross-cultural research for which convenient solutions were sought.

In a recent edition of their well known handbook Craig and Douglas (2000, p.141) define equivalence as: ‘Data that have, as far as possible, the same meaning or interpretation, and the

same level of accuracy, precision of measurement, or reliability in all countries and cultures'. Craig and Douglas (2000) address various issues that have to be taken into account if data are to be compared. They distinguish three forms of equivalence: construct equivalence, measurement equivalence, and equivalence in data collection techniques.

Within construct equivalence Craig and Douglas (2000) define three aspects. (1) *Conceptual* equivalence is 'concerned with the interpretation that individuals place on objects, stimuli or behaviour, and whether these exist or are expressed in similar ways in different countries and cultures' (p. 158). (2) *Categorical* equivalence 'relates to the category in which objects or other stimuli are placed' (p. 159). Categorical equivalence refers to comparability in product category definitions, and in background or socio-demographic classes that exist between countries. This definition by Craig and Douglas arises from the practice of marketing research. Product categories need not be similar across countries. For example, beer belongs to the category soft drinks in Southern Europe, whereas beer is considered to be an alcoholic beverage in Northern Europe. Moreover, category sizes may differ; in Greece spreading on bread or toast is common, making the category big, whereas spreading is hardly done in Italy, making the category small (Van Herk and Verhallen, 1995). (3) *Functional* equivalence relates to the question whether the concepts, objects or behaviours studied have the same role or function in all countries included in the analysis. It makes quite a difference whether a bicycle is considered mainly as a means of transport (such as in the Netherlands or India) or as a product for recreational purposes (as in the USA).

Craig and Douglas (2000) take examination of equivalence as a two-step procedure: 'once construct equivalence has been examined, the next step is to consider measurement equivalence' (p. 160). They distinguish three aspects of measurement equivalence: (1)

Translation equivalence refers to the translation of the research instrument into another language so that it can be understood by respondents in different countries, and has the same meaning in each research context. (2) *Calibration* equivalence refers to equivalence with regard to units of measurement, for example, monetary units and measures of weight used in questionnaires. Moreover, it refers to the use of colours and shapes in such a way that they are interpreted the same in different countries. Finally, (3) *Metric* equivalence refers to the specific scale or scoring procedure used for assessment. In the approach by Craig and Douglas (2000) a solution is sought per problem, for example translation. There is little integration of conceptual and measurement issues.

Other research in the same tradition as Craig and Douglas can be found in the management literature with authors like, for example, Sekaran (1983), Nasif et al. (1991), and Cavusgil and Das (1997). Sekaran (1983) links equivalence to various stages in the research process. She mentions equivalence issues related to function, instrumentation, data-collection methods, sampling design, and data-analysis. As in marketing, *functional equivalence* is associated with the role of objects or behaviours in different countries. *Instrumentation equivalence* includes equivalence in translation, syntax and concepts used. With data collection Sekaran mentions the importance of equivalence in response, timing, interviewer status, and type of research (longitudinal or cross-sectional). *Sampling equivalence* covers issues such as representativeness, and matching of samples. Following Sekaran (1983), Nasif et al. (1991) identified methodological problems in the cross-cultural research process and gave suggestions for reducing those problems. They mention several issues like functional equivalence and equivalence of instrumentation and data collection, and per issue they indicate suggestions for improvement. For example, back-translation is recommended to increase translation equivalence. Building upon this work Cavusgil and Das (1997) developed

a ‘generic process model’ for cross-cultural research. In this model, including seven steps, they thoroughly describe issues to be taken into account when doing a cross-national study. As in the studies already mentioned, issues of equivalence are linked to stages in the research process (for example, equivalence of administration and equivalence of responses are linked to the phase in the research process where the instrument is developed). We like to note that in this line of research data analysis is regarded important. However, it tends to be taken as one of several aspects in the cross-cultural research process.

In the second line of research on equivalence the emphasis has been on data analysis, as the principal means of demonstrating whether or not cross-cultural data can be taken as equivalent. This research has its roots in psychology, specifically in literature on bias and measurement invariance (e.g., Horn et al., 1983; Meredith, 1993). In these studies psychometric procedures are defined for assessing whether (test) scores from different groups can be validly compared. In psychology (e.g., Little, 1997), marketing (e.g., Steenkamp and Baumgartner, 1998), as well as in management literature (e.g., Mullen, 1995; Vandenberg and Lance, 2000) the value of measurement invariance for cross-cultural research has been recognized. These authors argue that the equivalence of measures can be established by means of multi-group confirmatory factor analysis models. By adopting the procedures as outlined in, for example, Steenkamp and Baumgartner (1998) construct, metric or scalar invariance of measures can be established. That is, sequential steps in nested multi-group mean and covariance structure models can determine the extent to which constructs can be compared across groups.

To distinguish the levels of invariance Van de Vijver and Leung (1997a,b) proposed three hierarchically ordered categories: construct equivalence, measurement unit equivalence, and scalar equivalence.

Construct equivalence (or structural equivalence) is the same as 'configural invariance' a term also used (e.g., Horn and McArdle, 1992; Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000). It refers to similarity of structural psychometric properties in data from different countries (Van de Vijver and Leung, 1997a). Construct equivalence exists if equal factor structures are obtained in different cultural populations. In terms of interpretation, construct equivalence implies that the same construct is being assessed. However, scores levels may or may not be equivalent across countries.

Measurement unit equivalence is also called 'metric invariance' (Horn and McArdle, 1992; Vandenberg and Lance, 2000). It refers to a situation where the unit of measurement is equal across populations, but where the origin of the measurement scale may be different. An analogue is the measurement of temperature, where degrees Celsius and degrees Kelvin are measured in the same units, but where the zero point (offset) differs. Thus, in terms of interpretation, measurement unit equivalence does *not* imply that scores on a single variable can be compared across countries; it implies that differences between scores (or patterns of scores) can be meaningfully compared across countries.

Scalar equivalence or full-score equivalence (also called 'scalar invariance') exists if the measurement scale in addition to having measurement unit equivalence also has an equal origin across countries. Scalar equivalence is the highest level of equivalence according to Van de Vijver and Leung (1997a,b). Comparisons of scores across countries on a single variable are only meaningful if this level of equivalence has been established. If there is scalar equivalence, it can be concluded that cross-national differences in score distributions on a variable correspond to differences in the underlying constructs.

In this line of research, the level of equivalence that has been established determines which inferences can be made. For example, if a trait like innovativeness is the target of study it can be concluded that people in culture "A", are less innovative than people in culture "B", only if scalar equivalence has been established. On the other hand, if the (positive) evidence is limited to construct equivalence the only conclusion can be that the instrument used assesses innovativeness in both cultures, but it is unclear whether a higher mean score in A implies a higher level of innovativeness. It should be noted that in this line of research multi-item scales are needed; with single items multivariate procedures cannot be applied.

Bias in the research process

There can be sources of bias in every stage in the research process. To gain equivalent results in international marketing studies, attention has to be paid to a range of possible sources of bias and their impact. This implies an integration of the process-oriented approach (as by e.g., Craig and Douglas, 2000), and the measurement-oriented approach (as by e.g., Steenkamp and Baumgartner, 1998). The link pin between the two orientations in our opinion lies in the notion that sources of bias can affect (in)equivalence at different levels.

In the psychological literature (e.g., Van de Vijver and Leung, 1997a,b; Berry et al., 2002) three kinds of bias are discussed, namely construct bias, method bias, and item bias.

Construct bias is likely to be present if the construct being studied differs across countries, or if the operationalisation does not fit cultural understanding. Construct bias can, for example, be induced if behaviours are sampled that are not associated with the construct studied. The use of butter for baking in one country cannot be compared with the use of butter for spreading in another country, and as a consequence, attitudes towards butter will reflect quite different notions about the use of butter (Van Herk, et al, 1994).

Method bias refers to instances where all or most items in a questionnaire are equally affected by a factor that is independent of the construct studied (Berry et al., 2002). Method bias can be due to interviewers (interviewer-interviewee interaction), the research method (telephone, mail or personal interviewing), or background characteristics of respondents, such as age or social class (Greenleaf, 1992a).

Item bias refers to distortions in specific items in the instrument (see Van de Vijver & Tanzer, 1997). Suppose, we employ a multi-item scale on 'health consciousness' and an item is included on 'visiting a fitness club at least once a week'. With an equal average concern about 'health consciousness' in two groups, but differential availability of health clubs, the answer 'no' obviously will have a different meaning. In such instances, we say that the item is biased.

"take in Table I"

At the beginning of a research process in marketing (stage I in Table 1), the problem is formulated and the objectives of the study are defined. In a cross-national study, a common first check is to determine whether the issue to be studied is relevant across countries. This includes the concepts to be examined, and in commercial studies it also comprises the product category studied, and the function of products and consumer habits. Insight into foreign markets can be obtained from the literature, consultations with fellow researchers (cf. Craig and Douglas, 2000), colleagues who are nationals of target countries, and/or qualitative pre-studies (Malhotra et al. 1996), such as focus groups (Carson et al., 2001), and exploratory observation. For example, Barzilay et al. (1994) reported studies in Western Europe in which the behaviour of women during food preparation was videotaped to help marketers understand habits in other countries. Those habits turned out to be very different; for example, for frying potatoes a deep-frying pan with special fat was used in Germany, whereas women in Greece

used a frying pan with olive oil. It turned out that the women used similar words, but the actual behaviour regarding frying was quite different. Such differences illustrate that concepts need not be equal in meaning and/or associated behaviours. Thus, country specific (or 'emic') practices are important to understand differences between countries. It is striking to note that about 80% of studies in cross-cultural organisational research use an 'etic' (culture-common) approach (Schaffer and Riordan, 2003), and less than 15% include emic elements. In summary, during the first stage of a research project, the main way to minimize bias is through international collaboration; this provides important information on specific habits, and the suitability of methods.

At stage II, the design stage, decisions are made concerning operationalisation of the constructs, the selection of items, and the response format. At this stage instruments (questionnaires, observation schedules) are developed and indications of construct, method, and item bias may emerge. For example, construct bias should be suspected, if a construct cannot be operationalized in a similar way in the countries studied. Again, collaboration with (preferably multi-lingual or bi-lingual) researchers across countries is vital. Another issue related to construct bias is the use of multi-item scales. Multi-item scales are required to be able to assess measurement invariance as outlined by, for example, Van de Vijver and Leung (1997). In recent academic cross-national studies, the measurement of constructs using multi-item scales, needed for psychometric analysis of equivalence, seems more common (see e.g., studies by Van Birgelen et al., 2002, and Atuahene-Gima and Li, 2002). However, for reasons of financial and time constraints, multi-items scales are scarce in commercial marketing research (Reynolds, 2000).

Other decisions made at Stage II are decisions on what response scales to use. Method bias is introduced at this stage if there is any factor in the instructions, response format of the items , or administration procedure that elicits different reactions across countries. The format of

response scales is a case in point. For example, in the United States a 5-point or a 7-point rating scale is most common, whereas in France, a 20-point scale prevails (Kotabe and Helsen, 1998). To minimise a difference in familiarity with the response scales, a good introduction with some practice items can be provided. In addition to scale use, method bias can be introduced if respondents are unfamiliar with a particular data collection method. For example, in Western countries it is common to use computerized personal interviewing or computerized telephone interviewing (CAPI method and CATI method; e.g., Malhotra and Birks, 2003), whereas this is still completely unknown in other parts of the world. Less familiarity with a research method is likely to affect results (see e.g. Serpell, 1979). The use of different methods in various countries does not alleviate problems; cultural differences in the results can then still be differences as well as due to the methods used, while several psychometric procedures to identify method bias are not anymore available. To minimise method bias, it is better to use the same method and the same response scales, and to give respondents the opportunity to practice.

Another important issue at the design phase is the translation of the instrument into other languages. The translation of one or more items can be less than optimal, because of the absence of precisely equivalent terms in each language. To minimise bias, back-translation is often recommended (e.g., Craig and Douglas, 2000). Another common method to develop a translation is the committee approach (see e.g., Van de Vijver and Tanzer, 1997); the strength of this approach is in the co-operative effort between people with different areas of expertise who together translate the instrument. Translation is paid attention to in more and more academic studies nowadays (see e.g., Sin et al., 2001), and also in commercial research it is an issue researchers are aware of (Reynolds, 2000).

At stage III, the sample composition and the sampling frame are determined. The definition of the sample may introduce bias in various ways. One strategy is to work with samples that are

representative of the target populations. In commercial surveys representative samples or samples specified by the client are preferred (Reynolds, 2000). Another strategy is to choose samples that are alike with respect to demographic characteristics. Such samples, for example students, can help to reduce bias. In academic studies, about half of the studies use matched samples (Schaffer and Riordan, 2003). To make between-country comparisons, samples should preferably show equal distribution on key demographic variables, such as age, education and income. This helps to determine whether differences found are real or measurement artefacts. If it is not possible to use similar samples, recording of background characteristics (e.g., age, education) is recommended to be able to statistically control for differences. This information can help detect differences in response styles (a type of method bias) such as yeasaying, that are known to be more prominent in people with a lower education, and a higher age (e.g., Greenleaf, 1992; Narayan and Krosnick, 1996).

At the data collection phase (IV), virtually any procedure is vulnerable to method bias. To begin with, instructions to interviewers need not always be understood in the same way. Method bias can emerge during interviews if respondents are more willing to talk about sensitive issues with certain interviewers; women may be more willing to talk about violence to females than to males. Moreover, bias may be induced by different time frames. If data are collected in one country half a year or more before this is done elsewhere, differences in fashions or in the eco-cultural environment (e.g., economic situation) may lead to different answers. This especially holds in a commercial setting, where the social context may affect variables such as buying intention. Again, method bias cannot be prevented; it can only be reduced. But the researcher is not helpless; instructions can be tried out in pilot studies; and interviewer characteristics can be recorded. The latter should be standard practice, as it is known that interviewer effects can be non-negligible (Kumar, 2000).

During stage V, coding and editing, item bias may be introduced. Coding refers to assigning answers to response categories if open-ended questions are used, and editing refers to correcting inconsistent answers in the questionnaires. Item bias is more likely if coding and editing are done separately in each country. Thus, item bias can be decreased if there is central coordination of research activities.

At stage VI, the analysis phase, it is possible to assess the absence or presence of bias by means of statistical analysis. Procedures outlined by research on measurement invariance can be followed. In the preceding phases, one can be aware of bias (threats to equivalence), and try to minimise or avoid these, but the empirical proof of equivalence (i.e., absence of bias) usually has to come from analyses of equivalence after the data have been collected.

An illustration

In 1996, a pan-European analysis of the (male) shaving market was conducted. Countries included France, Germany, Italy, Spain and the United Kingdom. A main goal was to find similarities between these markets that could be used as a starting point for pan-European product developments and pan-European product introductions. Therefore, the comparability of market information across countries was a main issue.

As the company concerned had been active in the shaving market for almost a century, much information was available on the domain of study. For example, market shares in the various countries and shaving habits were known. In the past, qualitative exploratory studies in several countries had been done to determine the dimensions men use to describe their shaving experiences.

At stage I, extensive information was available to the researchers. They knew that men use six dimensions to describe their shaving experience. The expected similarity of these dimensions

across European countries made it worthwhile to examine whether they could be decomposed in much the same way. For example, it was expected that similar notions should exist to describe the dimension 'shaving result'. In other words, it was expected that operationalisations of the dimensions should lead to (structurally) equivalent scales.

As validated scales to measure these dimensions were not readily available, items had to be developed for each dimension (Stage II). In this process items from previous marketing research studies were used. After compiling the questionnaire items, it was decided that five-point rating scales, with the endpoints labeled 1 ('disagree strongly') to 5 ('agree strongly') should be employed. In addition, questions were developed on male shaving behaviour (e.g., shaving frequency, method used). As a next step, the questionnaires were translated from English into German, Spanish, Italian and French by bi-lingual researchers using the committee approach (see Van de Vijver and Tanzer, 1997). In all countries, the method chosen was a mail survey using a panel of a large marketing research agency. As in many commercial marketing research studies, the choice for this type of data collection method was driven by financial and time constraints.

The sample sizes were fixed at about 1000 in each country (Stage III). Representative samples of only male respondents were selected; ranging from 15 to 80 (mean 43) years of age in each country. It should be noted that this choice of representative samples may lead to method bias, because differences in demographic variables such as education and income level are known to exist between the countries studied.

Stage IV entails the data collection. Instructions were given to the respondents, who were all members of established marketing research panels. The data collection was done in the same period in all countries studied. No special precautions were taken at this stage to avoid, or control possible sources of bias.

Stage V does not apply in the present case, as there were no open-ended questions, and thus no special rules for editing and coding were needed.

Whether the data collected in the way described were equivalent had to be established afterwards through data analysis. In a sense, the proof of the pudding had to be in the eating. This last stage (VI) was done in four steps. The first step was data cleaning. Respondents with missing values on items of interest were removed from the data set. Resulting sample sizes were 985 in Germany, 890 in France, 820 in the United Kingdom, 1062 in Italy, and 790 in Spain. The partial non-responders did not differ from the rest of the sample on demographic variables. The second step was equivalence assessment. For the sake of clarity, we focus here on one construct to assess equivalence in the shaving domain, namely 'shaving result'. In the questionnaire 6 items were included that together could be used to assess this construct. The items included, for example, 'after shaving you can see there is not a single hair left uncut on your face' and 'you are closely shaven from early morning till late at night'. Next, to test for equivalence in the five countries, the program Lisrel 8.5 (Jöreskog and Sörbom, 2000) was used. It turned out that the 6 variables chosen to measure the construct 'shaving result', were not construct equivalent. Inspection showed that the poor fit was mainly due to a single item ('gives a very close shave'). After removal of this item the fit improved, and 'shaving result' was construct equivalent However, further analyses showed that there was no measurement unit equivalence. Thus, the level of equivalence that could be established was construct equivalence.

The third step in the analysis involved the interpretation of results on equivalence. The finding of construct equivalence justified the interpretation that men in all the five countries understand the same thing when 'shaving result' is talked about. However, as there was no measurement unit equivalence, let alone scalar equivalence, we could not infer whether men in, for example, Italy do experience a better shaving result than men in Germany do.

The fourth and final step concerned the substantive explanation of results. For international marketing purposes it was important to know why (1) the item 'gives a close shave' caused item bias, and (2) what can be concluded on the basis of the construct 'shaving result' across countries. For answers to these questions we used other information from the questionnaire, especially items on shaving behaviour. Regarding the first point it was found that 'shaving result' is positively related to shaving frequency in all countries, whereas the item 'gives a close shave' was not in some countries. Regarding the second point it could be concluded that the construct 'shaving result' had the same meaning in all countries, but between country comparisons at levels of scores (e.g., means) were not allowed. However, investigating relations of the construct 'shaving result' with other variables *within* each country was allowed. Such results can be valuable for marketing decision making. It was for example found that within all countries men scored higher on when they shaved with a blade as compared to an electric shaver. As this result was found in all countries, blade and electric shaving could be compared on 'shaving result' in the same (qualitative) way in a pan-country communication strategy. However, it should be noted that this is no quantitative comparison. It remained unclear in this study *why* there was no measurement unit equivalence.

Discussion

In analysing equivalence some researchers focus on the research process, while others are mainly concerned with analysis and interpretation of data. For greater clarity, we proposed here a differentiated view, distinguishing levels of equivalence and types of bias. Sources of bias in the research process are considered factors that decrease the level of equivalence that can be established. In our approach, equivalence is accepted if serious attempts to find inequivalence have been unsuccessful. In this we follow Van de Vijver and Leung (1997) who reserve the use of the term 'equivalence' for outcomes of formal statistical analyses. This

makes the meaning of the term 'equivalence' less extensive, and more a matter of measurement , than the way it is used by authors like Craig and Douglas (2000).

In this study we indicated sources of bias in the research process, and linked these to the level of equivalence that can be established. Construct bias is the prevalent type of bias in the first two stages in the marketing research process. This type of bias is the most serious one, because it precludes any form of comparison, making cross-national comparisons ambiguous or even erroneous. Item bias is less serious than construct bias as it only affects part of the items in the instrument. If various items are used to measure a construct, a biased item can be eliminated from the scale, and the resulting shortened scale can still be construct equivalent. Method bias affects the level of scores on all, or at least most items in a scale. If there is method bias, it is still possible to establish construct equivalence, or even measurement unit equivalence. However, scalar equivalence is ruled out. It is not easy to eliminate method bias, since separation of bias and real differences is not straightforward (Greenleaf, 1992). With respect to understanding method bias response styles offer an interesting avenue for further research (see e.g., Baumgartner & Steenkamp (2001); Smith (in press)).

Equivalence cannot be assumed; construct, measurement unit, or scalar equivalence have to be established by means of explicit procedures. In our example from commercial research only construct equivalence was found, be it after elimination of one item. This is not exceptional. In research papers in international marketing, such those by Homburg et al. (2002), and by Van Birgelen et al. (2002) construct equivalence was found, but measurement unit equivalence was not. In other studies (Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000) only partial measurement equivalence was found. Thus, even if the topic of research is well studied, the samples are matched, and questionnaires are carefully back-

translated, equivalence is not guaranteed. In articles as mentioned, the use of equivalence testing is often seen as a pre-test: if a certain level of equivalence is attained certain comparisons can be made, if not, comparisons between countries are not allowed. However, if a next level of equivalence cannot be established, analyses to find out why this is so can still provide valuable information. Analyses should not stop; investigating which items are biased and why this is the case is an interesting avenue for further research.

Our study has one important implication for the management of international companies. A large number of managerial decisions of companies is influenced by consumer perceptions and acceptance of a company's products. The findings of our study can help managers to establish the extent to which such consumer perceptions are equal across countries. That is, if the perceptions are construct equivalent, it can be concluded that they have the same meaning for people in all countries studied. Then, management can make the founded decision the same concepts can be used in a pan-country communication strategy. Being able to establish the level of equivalence therefore provides business value, as the risk of making a wrong decision decreases. International marketing research studies are expensive, and cutback in expenditure is often looked for. However, this cutback in expenditure should not be in multi-item scales. They are worth the money, because they may provide valuable information on differences and similarities between countries.

We like to conclude with a comment by Cavusgil and Das (1997, p. 74) who argued that it is : "easier to recover from lapses in data analysis than in specification error". Minimizing construct bias in the early stages of the research process is a basic prerequisite. At that stage collaboration with other researchers and marketeers can help define the marketing (research) problem. Later, to be certain that corresponding constructs were measured in all countries,

multi-item scales are required to establish equivalence. Best practices should include elements from both the conceptual and the measurement oriented approach to equivalence. Analysis of cross-cultural differences is partly an art. A researcher needs a proper grasp of the various factors that can interfere with the interpretations of findings at face value. However, the analysis of cross-cultural differences is also a science; a conceptual approach aids in a systematic procedure for assessing equivalence.

References

Athuahene-Gima, K. and Li, H. (2002), "When Does Trust Matter? Antecedents and Contingent Effects of Supervisee Trust on Performance in Selling New Products in China and the United States," *Journal of Marketing*, Vol. 66, No. 3, pp. 61-81.

Aulakh P.S. and Kotabe M. (1993), "An assessment of theoretical and methodological development in international marketing: 1980-1990," *Journal of International Marketing*, Vol. 1, No. 2, pp. 5-28.

Barzilay, J., Chada P.M., Van Herk H. and Verhallen Th.M.M., 1994, "International segmentation", In: Marketing Review 1992, C. Lambrinopoulos Ed., 207-208.

Baumgartner, H. and Steenkamp J.B.E.M. (2001), "Response styles in marketing research: A cross-national investigation," *Journal of Marketing Research,* Vol. 38, No. 2, pp. 143-156.

Berry, J.W., Y.H. Poortinga, M.H. Segall, and Dasen P.R. (2002), *Cross-cultural psychology. Research & applications.* Second Edition. Cambridge: Cambridge University Press.

Carson, D., Gilmore A., Perry C., and Gronhaug K (2001), *Qualitative Marketing Research.* SAGE Publications.

Cavusgil, S.T., and Das, A. (1997), "Methodological Issues in Empirical Cross-cultural Research: A Survey of the Management Literature and a Framework," Management International Review, Vol. 37, No. 1, pp. 71-96.

Craig C.S. and Douglas S.P. (2000), *International Marketing Research*. Second Edition, New York: Wiley.

Douglas, S.P. and Craig C.S. (1983), *International Marketing Research*. Englewood Cliffs, N.J.: Prentice Hall.

ESOMAR (2002), "ESOMAR Annual Study on the Market Research Industry 2001," ESOMAR, European Society for Opinion and Marketing Research, Amsterdam.

Greenleaf, E.A. (1992), "Improving rating scale measures by detecting and correcting bias components in some response styles," *Journal of Marketing Research*, Vol. 29, No. 2, pp. 176-188.

Homburg, C., Krohmer H., Cannon J.P. and Kiedaisch, I. (2002), "Customer Satisfaction in Transnational Buyer-Supplier Relationships," *Journal of International Marketing*, Vol. 10, No. 4, pp.1-29.

Horn, J.L., McArdle J.J., and Mason R. (1983), "When invariance is not invariant: A practical scientist's look at the ethereal concept of factor invariance," *The Southern Psychologist*, Vol. 1, pp. 179-188.

Horn, J.L. and McArdle, J.J. (1992), "A practical and theoretical guide to measurement invariance in aging research," *Experimental Aging Research*, Vol. 105, pp. 117-144.

Hui, C.H. and Triandis H.C. (1985), "Measurement in cross-cultural psychology. A review and comparison of strategies," *Journal of Cross-Cultural Psychology*, Vol. 16, No. 2, pp. 131-152.

Jöreskog K.G. and Sörbom D. (2001), *Lisrel 8 Users Reference Guide,* SSI.

Kotabe, M. and Helsen K. (1998), *Global marketing management*. New York: John Wiley & Sons, Inc.

Kumar, V. (2000) *International marketing research*. Upper Saddle River NJ: Prentice-Hall Inc.

Little, T.D. (1997), "Mean and Covariance Structues (MACS) Analyses of Cross-Cultural Data: Practical and Theoretical Issues," *Multivariate Behavioral Research*, Vol. 32, No. 1, pp. 53-76.

Malhotra, N.K., Agarwal J. and Peterson M. (1996), "Methodological issues in cross-cultural marketing research," *International Marketing Review*, Vol. 13, No. 5, pp. 7-43.

Malhotra, N.K. and Birks D.F. (2003), *Marketing Research: An Applied Orientation. Second European Edition.* Pearson Education Limited.

Meredith W. (1993), "Measurement invariance, factor analysis and factorial invariance," *Psychometrika,* Vol. 58, No. 4, pp. 525-543.

Mitra, D. and Golder, P.N. (2002), "Whose Culture Matters? Near-Market Knowledge and Its Impact on Foreign Market Entry Timing," *Journal of Marketing Research*, Vol. 39, No. 3, pp. 350-365.

Mullen, M.R. (1995), "Diagnosing measurement equivalence in cross-national research," *Journal of International Business Studies*, Third Quarter, pp. 573-596.

Nasif E.G., Al-Daeaj, H., Ebrahimi, B. and Thibodeaux, M.S. (1991), "Methodological Problems in Cross-Cultural Research: An Updated Review," *Management International Review,* Vol. 31, No. 1, pp. 79-91.

Narayan, S. and J.A. Krosnick (1996), "Education moderates some response effects in attitude measurement," *Public Opinion Quarterly*, 60, 58-88.

Reynolds, N.L. (2000), "Benchmarking international marketing research practice in UK agencies – Preliminary evidence," *Benchmarking: An International Journal*, Vol. 7, No. 5, pp. 343-359.

Schaffer B.S. and Riordan, C.M. (2003), "A Review of Cross-Cultural Methodologies for Organizational Research: A Best-Practices Approach," *Organizational Research Methods*, Vol. 6, No. 2, pp. 169-215.

Sekaran, U. (1983), "Methodological and theoretical issues and advancements in cross-national research," *Journal of International Business Studies*, Fall, pp. 61-73.

Serpell, R. (1979), "How specific are perceptual skills?," *British Journal of Psychology*, Vol. 70, pp. 365-380.

Sin, L.Y.M., Cheung G.W.H., and Lee R. (1999), "Methodology in Cross-Cultural Consumer Research: A Review and Critical Assessment", *Journal of International Consumer Marketing,* Vol. 11, No. 4, pp. 75-96.

Sin, L.Y.M., Hung, K. and Cheung G.W.H. (2001), "An Assessment of Methodological Development in Cross-Cultural Advertising Research: A Twenty-Year Review", *Journal of International Consumer Marketing,* Vol. 14, No. 2/3, pp. 153-192.

Singh, J. (1995), "Measurement issues in cross-national research," *Journal of International Business Studies*, Third Quarter, pp. 597-619.

Smith (in press), "Acquiescent Response Bias as an Aspect of Cultural Communication Style," *Journal of Cross-Cultural Psychology*.

Steenkamp, J.B.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 Birgelen, M., De Ruyter K., De Jong, A. and Wetzels, M. (2002), "Customer Evaluations of after-sales service contact modes: An empirical analysis of national culture's consequences," *International Journal of Research in Marketing*, Vol. 19, No. 1, pp. 43-64.

Vandenberg, R.J. and Lance, C.E. (2000), "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research," *Organizational Research Methods*, Vol. 3, No. 1, pp.4-69.

Van de Vijver, F.J.R. and Leung K. (1997a), *Methods and data analysis for cross-cultural research*. Sage.

Van de Vijver, F.J.R. and Leung K. (1997b), "Methods and data analysis for cross-cultural research," In: *Handbook of Cross-Cultural Psychology, 2nd Ed., Volume I. Theory and Method*, Berry, J.W., Y.H. Poortinga and J. Pandey (Eds.), 257-300. Boston: Allyn and Bacon.

Van de Vijver, F.J.R. and Tanzer, N. (1997), "Bias and Equivalence in Cross-Cultural Assessment: An Overview", *European Review of Applied Psychology*, Vol. 47, No. 4, pp. 263-279.

Van Herk H., Verhallen Th.M.M. and Barzilay J. (1994), "Methodological Issues in international segmentation," In: *Marketing: Its dynamics and challenges*. Bloemer, J., J. Lemmink and J. Kasper (Eds.), pp. 1311–1314. Maastricht NL: EMAC.

Van Herk, H. and Th.M.M. Verhallen (1995), "Response effects in international research in the food area," In: *Proceedings of the Second International Conference on "The Cultural Dimension of International Marketing'*, 392-402. Odense, Denmark.

TABLE 1 The research process and bias

	Stages in the marketing research process	Source of bias	Issues	Prevalent types of bias
I	Problem formulation	Concepts Category Function	Purpose of the study	construct
II	Research design	Operationalisation	Type of study Type of questions	construct
		Instrument design	Item selection Type of response format	item method
		Translation		item
		Method	Personal, mail, telephone	method
III	Sample selection	Sampling	Target population Sampling frame	method
IV	Data collection	Fieldwork	Procedures Interviewer selection Time frame	method
V	Data editing and coding	Editing Coding Calibration	Data editing Data coding	item
VI	Analysing and interpreting data		Statistical procedures	