Exposé

Establishing Scale Equivalence in Intercultural Marketing Research

Studies: A Structured Literature Review

submitted by: Jan-Niklas Hombach

Update: 15.09.2016
<table>
<thead>
<tr>
<th>Inhalt</th>
<th>Seite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>3</td>
</tr>
<tr>
<td>Introduction</td>
<td>4</td>
</tr>
<tr>
<td>Literature review</td>
<td>6</td>
</tr>
<tr>
<td>Theoretical background</td>
<td>10</td>
</tr>
<tr>
<td>Method</td>
<td>27</td>
</tr>
<tr>
<td>Structure</td>
<td>28</td>
</tr>
<tr>
<td>Working plan</td>
<td>29</td>
</tr>
<tr>
<td>References</td>
<td>30</td>
</tr>
</tbody>
</table>
Abstract

Titel – Establishing scale equivalence in intercultural marketing research studies: A structured literature review

Background – Conducting research in an environment where many cultures come together sometimes bears problems. Instruments implemented in research must ensure that obtained results are measured equivalent and so interpretation is not biased. Guidelines have been established using statistical techniques just like multi-group confirmatory factor analysis to test for equivalence in data.

Purpose – As literature calls for more attention on equivalence in measurement the work at hand aims to review if current intercultural marketing literature report equivalence in publications.

Method – A structured literature review is applied with content analysis as additionally citation analysis. Four high ranked marketing journals are selected looking for academic articles from the year 2000 until today as well as a second period from 2012 until today.

Keywords – intercultural marketing research, scale equivalence, comparability, multi-group confirmatory factor analysis, reflective measures, formative measures
“A critical feature of inter[cultural] marketing is that it is conducted in a multi-country, multi-cultural and multi-linguistic environment. This gives rise to a number of issues relating to the comparability of data collected in diverse research environments. A critical issue is whether the same or a similar research instrument and procedures are equally adapted or suitable, and will yield comparable results in each environment.”
(Craig & Douglas, 2000, p.41)

Introduction

In the middle of the 20th century globalisation, in political or economic aspects, takes it course to the extent as it is known today. Firms started to expand their negotiations abroad as well as free trade prospered thanks to the establishment of agreements across nations just like the European Union, NAFTA, AFTA, or Mercosur, to name just a few. On the one hand these circumstances favoured and still favour migration all over the world and so brought different countries as well as cultures closer together, although, on the other hand cultural differences still remain (Costa & Bamossy, 1995; Leung, 2008).

This fact provoked marketing academics’ interest, as well as other academic disciplines, in putting intercultural studies on the academic agenda (Douglas & Craig, 1997; Van de Vijver & Leung, 1997). The goal of marketing research, using e.g. psychological concepts, is trying to understand the mechanisms beyond numbers just as attitudes or behaviour towards products or brands (Craig & Douglas, 2000). These theoretical constructions often are measured through the implementation of multi-item scales (Wagner, Wetzels & Winklhofer, 2005). If subsequently the cultural factor takes place the focus is on
comparability of these concepts across cultures as a fundament in intercultural research (Kumar, 2000) which is only possible if obtained data are suitable, i.e. equivalent respectively scale equivalence is established.

In the past scientists were not be aware on appropriate comparison so there was claim for paying more attention on equivalence (e.g. Craig & Douglas, 2000; Douglas & Craig, 1997; Davidov, Meulemann, Cieciuch, Schmidt & Billiet, 2014) as the aim is to obtain real similarities or differences among cultures and not to get biased results from applying inadequately methods. Steenkamp and Baumgartner (1998) interpret that less attention on equivalence was due to abundance of the equivalence terminology (see also Johnson, 1998) as also less knowledge or problems with statistical testing for equivalence for academics. For these reasons guidelines of statistical testing procedures has been established to enable researchers to compare results (Diamantopoulos & Papadopoulos, 2010; Mullen, 1995; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000).

Now that such guidelines are available the purpose of the work at hand is to review current marketing literature dealing with intercultural issues if testing for data comparability is considered and reported.

Hence, the work is structured as follows: First a theoretical background is introduced to let the reader know how scale equivalence has to be established and why it is so important. This is done in a framework of the intercultural marketing research process. The focus lies on two approaches proposed in literature to assess scale equivalence in cultural comparison, reflective and formative measurement models. As Engelen and Brettel (2011) conducted a review of marketing literature counting how many marketing publications rely on cross-cultural issues the second part then is dedicated to current intercultural marketing literature through conducting a structured literature review with focus on equivalence assessment. This
is done by a content analysis and additional by a citation analysis. The work finishes then with a conclusion.

**Literature review**

The following table contains the most important academic papers or textbooks concerning the topic. Only literature is included which the author concerns as highly relevant, contributing to the structure of this work or is concerned as the most cited during the research. For more literature see references.

<table>
<thead>
<tr>
<th>Titel</th>
<th>Author(s)</th>
<th>Year of Publication</th>
<th>Journal / Textbook</th>
<th>Content / Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response styles in Marketing Research: A Cross-National Investigation</td>
<td>Baumgartner, Steenkamp</td>
<td>2001</td>
<td>Journal of Marketing Research</td>
<td>Summary of various response styles on scales conducting research in 11 countries</td>
</tr>
<tr>
<td>Cross-cultural psychology. Research and Applications</td>
<td>Berry, Poortinga, Breugelmans, Chasiotis, Sam</td>
<td>2011</td>
<td>Textbook</td>
<td>Overview of conducting research in cross-cultural psychology</td>
</tr>
<tr>
<td>Methodological issues in cross-cultural research: An overview and</td>
<td>Buil, de Chernatony, Martínez</td>
<td>2012</td>
<td>Journal of Targeting, Measurement and</td>
<td>Overview of cross-cultural research, focus on research process</td>
</tr>
<tr>
<td>recommendations</td>
<td>Analysis for Marketing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Byrne, Shavelson, Muthén</td>
<td>Psychological Bulletin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>International Marketing Research</td>
<td>Overview of conducting international marketing research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Craig, Douglas</td>
<td>Textbook</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessing the cross-national invariance of formative measures: Guidelines for international business researchers.</td>
<td>Guidelines conducting assessment of measurement equivalence with formative measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diamantopoulos, Papadopoulos</td>
<td>Journal of International Business Studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Index</td>
<td>Introduction to formative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diamantopoulos</td>
<td>Journal of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>Introduction to formative</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Establishing scale equivalence in intercultural marketing research studies: A structured literature review

<table>
<thead>
<tr>
<th>Construction with Formative Indicators: An Alternative to Scale Development</th>
<th>Winklohofer</th>
<th>Marketing Research</th>
<th>measurement models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessing cross-cultural marketing theory and research</td>
<td>Engelen, Brettel</td>
<td>2011</td>
<td>Journal of Business Research</td>
</tr>
<tr>
<td>Cross-Cultural Survey Methods</td>
<td>Harkness, Van de Vijver, Mohler (Eds.)</td>
<td>2003</td>
<td>Textbook</td>
</tr>
<tr>
<td>Topic</td>
<td>Author/Reference</td>
<td>Year</td>
<td>Source</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>----------------------------</td>
<td>------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Establishing scale equivalence in intercultural marketing research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>studies: A structured literature review</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Misspecification in Marketing and Consumer Research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>International Marketing Research</td>
<td>Kumar</td>
<td>2000</td>
<td>Textbook</td>
</tr>
<tr>
<td>Methods and Data Analysis for Cross-Cultural</td>
<td>Van de Vijver, Leung</td>
<td>1997</td>
<td>Textbook</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Structure of equivalence and biases</td>
</tr>
</tbody>
</table>
Establishing scale equivalence in intercultural marketing research studies: 
A structured literature review

<table>
<thead>
<tr>
<th>Research</th>
<th>Equivalence of survey data: relevance for international marketing</th>
<th>2005</th>
<th>European Journal of Marketing</th>
<th>Overview of cross-cultural research, focus on research process</th>
</tr>
</thead>
</table>

One of the major learnings when conducting the review of the relevant literature is that not only marketing but also many other academic disciplines such as psychology, sociology or business studies treat the question how to conduct correct comparison of studies in an intercultural environment and that they do so in a similar way converting it into an interdisciplinary issue. Furthermore when statistical instruments are discussed in literature using factor analysis is outstanding. Another finding is that either statistical testing procedures are discussed or the conditions of establishing equivalence in previous steps of the research process but rarely together. This has an impact on the structure of this thesis bringing both approaches together. At last, despite of the given instruments and guidelines establishing equivalence the literature mentions that in reality equivalence is often ignored which comes to the research question at hand if intercultural marketing publications report for equivalence.

**Theoretical background**

Generally spoken the research process both at the domestic as well as at the intercultural level do not differ much. But there is an outstanding difference in these two areas called culture (Craig & Douglas, 2000; Kumar, 2000). It is a factor not to be
underestimated and to be kept in mind when conducting intercultural research because “culture is an important contributor to the development and display of human behaviour” (Berry, Poortinga, Breugelmans, Chasiotis & Sam, 2011, p.2). In academic environment there are several existing definitions concerning the cultural concept differing from discipline to discipline yet no general definition is available (Longhurst, Smith, Bagnall, Crawford & Ogborn, 2008). One of the most influencing works concerning cultural issues in business studies, including marketing, and making reference in academic textbooks is Hofstede´s study (Müller & Gelbrich, 2004). Hofstede (2006) so defines culture metaphorically as a programming of the individual´s mind driven by the society he or she lives in and which thus results in shared values, symbols, rituals etc. within this particular culture.

Based on this definition intercultural then is defined as the interaction between several cultures (Müller & Gelbrich, 2004).

As the focus is on comparability (Johnson, 1998; Kumar, 2000; Leung, 2008) the theoretical background concerning the establishment of comparison of intercultural research results respectively equivalence of data will be explained. Where in literature relying on equivalence focus either is on the assessment of measurement or on problems occurring during the research design, referring to as bias, (see e.g. He & Van de Vijver, 2012; Van de Vijver, 2003, 2011) this work follows the approach of Van Herk, Poortinga and Verhallen (2005) to combine the two approaches of “process-oriented” and “measurement-oriented” (see also Buil, de Cernatony & Martinez, 2012) as also entire textbooks like Craig & Douglas (2000) or Kumar (2000) following the process in form of their chapter structure.

---

1 Note that it is important to establish the right definition as academics sometimes do not when referring to intercultural issues just named as international, cross-national or cross-cultural (see for e.g. Craig & Douglas, 2011). Because the unit of analysis is often on country level (Craig & Douglas, 2000) most literature use the concept of ´nation´ when referring to as ´culture´. Even it is not the correct way this work also refers to the national focused interpretation while the treatment of several cultures into one country called subcultures or ethnicities just like Hispanics in the USA, Turkish or Russians in Germany, North Africans in France, or even Catalans in Spain is thus topic for ethnocultural marketing (see e.g. Müller & Gelbrich, 2004; Venkatesh, 1995).
Establishing scale equivalence in intercultural marketing research studies: A structured literature review

First a short overview of the entire research process in intercultural research is presented whereas in the following subchapters the equivalence established as also problems occurring in the several stages of the process are presented in detail.

The Research Process

The research process in intercultural marketing is conducted in several stages, from formulating the research question to the presentation of results where several questions have to be considered in each of the stages due to the cultural factor. This process is shown in figure 1.²

![Figure 1. The research process (based on Buil, de Chernatony & Martínez, (2012); Craig & Douglas, 2000; Ghauri & Gronhaug, 2010)](image)

The first step in the intercultural marketing research process is to define the topic, to name the problem that should be studied and therefore formulate a research question. Mainly

---

² Note that the research process contains a last stage of presenting the results. This one is not included in the present work as the focus is on the stages relying on the establishment of equivalences.
there are three types of purposes conducting an intercultural study: descriptive, comparative, and theoretical (Craig & Douglas, 2000; Kumar, 2000). A descriptive research aims at exploring concepts in foreign cultures separately. In contrast a comparative design’s goal is to compare one culture with another. The third research type then is exploring whether an established theory holds in every culture. The last one is also known as cross-cultural research (Müller & Gelbrich, 2004; Venkatesh, 1995). The present work’s focus is on the comparative and theoretical purpose. If there is few knowledge about the object’s or concept’s perception abroad then it is recommended to do a brief research of yet existing studies in considered cultures, may call this secondary data research (Craig & Douglas, 2000).

The next step is to choose the correct research design. It includes every previous step to plan the implementation of the research. One important issue is to clarify the unit of analysis across cultures called in this context culti-unit (Douglas & Craig, 1997; Naroll, 1970). The most used unit is the country as studied object (Craig & Douglas, 2000). For this reason the definition of intercultural research above mentioned refers to as research across various countries. Also the sub-unit for specific topics has to be clarified which involves more problems. For example the definition of the company’s size referring to small and medium-sized enterprises or big companies differs from country to country (Ghauri & Grønhaug, 2010) or the definition of urban and rural (Kumar, 2000).

The fact that researchers themselves belong to a particular culture also affects their point of view when conducting intercultural research. Commonly mentioned approaches to this kind of research procedure are the emic and etic approaches (see e.g. Craig & Douglas, 2000). The etic approach uses an outstanding point of view, i.e. cultural free or rather general view whereas the emic approach accounts for cultural singularities. It was criticized that most academic studies do research by a western point of view (Bond, Fu & Pasa, 2001; Harkness,
Establishing scale equivalence in intercultural marketing research studies:
A structured literature review

Mohler & Van de Vijver, 2003; Van de Vijver & Leung, 1997) which is called ethnocentrism
(Keith, 2011; Westwood, 2004). On the instrument’s level it means that instruments ones
designed in one country can show validity and reliability but that does not mean that it holds
for another country or culture by simply adopting scales (Abukar, 2015; He & Van de Vijver,
2012). To establish comparability instruments have to be either adapted or even new ones
have to be designed (Abukar, 2015; He & van de Vijver, 2012).

Further steps in the process then is the data collection where high awareness in the
sampling process is required (Craig & Douglas, 2000; Kumar, 2000)

Finishing the research process the data first needs to be analysed where measurement
procedures are used to test if the data collected holds the equivalence and comparability.
Then the obtained results have to be interpreted. If there is no equivalence established
interpretation is not possible.

Equivalences and biases of scales

Several terminologies due to equivalence issues have been mentioned in literature
which Johnson (1998) has reviewed and tried to summarize and categorize. As mentioned
above most literature addresses only on measurement issues by using the terminology
equivalence (see e.g. He & Van de Vijver, 2012) and focus is on the statistical testing if
validity and reliability of the instrument used is hold across cultures (Van Herk, Poortinga &
Verhallen, 2005). While in contrast issues examined in former steps as well as in
implementation rather are seen from a problem point of view, i.e. biases which might occur
(Van der Vijver 2003, 2011; Yaprak, 2003). However in the present work the author decided
to refer to this as conditions which have in some way direct or indirect influence on the
scales. It is very important during the research process establishing equivalence and resolve
biases because “any comparison stands or falls on the solution of these two issues” (Van de
Establishing scale equivalence in intercultural marketing research studies: A structured literature review

Vijver, 2003, p.154). Furthermore this work focuses on scale equivalence. The author decided for the terminology of scale equivalence as the common use of measurement equivalence is composed by various levels which in turn contribute as a whole to scales and scale development (Diamantopoulos & Winklhofer, 2001; Diamantopoulos, Riefler & Roth, 2008; Kumar, 2000; Sharma & Weathers, 2003).

Embedded into the presented framework of the research process it comes to following categorisation of scale equivalence: construct, instrument design and administration (Mullen, 1995; Van de Vijver, 2003) as shown in table 2.

Table 2. Types of equivalence

<table>
<thead>
<tr>
<th>Construct:</th>
<th>- conceptual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- functional</td>
</tr>
<tr>
<td></td>
<td>- categorical</td>
</tr>
<tr>
<td>Instrument design:</td>
<td>- Item function</td>
</tr>
<tr>
<td></td>
<td>- measurement:</td>
</tr>
<tr>
<td></td>
<td>- translation</td>
</tr>
<tr>
<td></td>
<td>- metric</td>
</tr>
<tr>
<td></td>
<td>- calibration</td>
</tr>
<tr>
<td>Administration:</td>
<td>- Sampling</td>
</tr>
<tr>
<td></td>
<td>- Interviewer effect</td>
</tr>
</tbody>
</table>

(based on Mullen, 1995; Kumar, 2000, 2009; Van de Vijver, 2003)

**The construct**

The construct is a “mental model of phenomena” (McArthur, 2007, p.28) and is directly connected to the research question as it is the object of interest. Therefore marketing academics look for the object itself just as brands or products or manage with theoretical concepts from other disciplines just like psychology (Craig & Douglas, 2000) and try to
understand constructs like behaviour towards purchase actions or the people’s attitude behind it. Attitudes, for example, on the theoretical level are composed by three elements which are affect, behaviour and cognition (Solomon, 2015). They represent a feeling about an object, a then resulting activity and the information processing about the object. As Solomon (2015) describes in the marketing context the way an attitude is formed is categorized in three dimensions (hierarchical orders) based upon the theory of consumer’s purchase process in which one of the attitude’s elements is more outstanding. Thus an attitude can be formed by, first, collecting information about the object followed by an affection towards the object and resulting in a behaviour (standard learning hierarchy: cognition, affect, behaviour). A second way is that the action occurs first (low-involvement hierarchy: behaviour, affect, cognition). Another possibility to form an attitude is beginning with the affect itself mainly based on experiences in the past (experiential hierarchy: affect, behaviour, cognition). The different types of attitude forming differ in the way a person is emotionally involved which also depends on the product type like daily live products or luxury goods.

With regard to equivalence the construct has to be perceived in all studied cultures similarly. Though there are three categories of perception on the construct level: functional, conceptual, and categorical (see e.g. Craig & Douglas; Kumar, 2000; McArthur, 2007). Functional means that people have the same sense for the function a construct is exercising. In contrast conceptual refers to the meaning itself. Categorical is understood the way constructs are categorized and if they do so in the same way across cultures.

Problems might arise when an emic approach is applied where in consequence the comparability of the construct could be erroneous (Van de Vijver, 2011).
Instrument design

The instruments respectively scales meant for the research have to be designed to be suitable in all cultures. As the constructs only exist on the theoretical level they themselves cannot be measured directly than through so called items (or indicators) which are representations of the first measured through questions in research instruments (Churchill, 1979; Harkness, Mohler & Van de Vijver, 2003). Constructs need multi-item scales as they give more evidence of the construct than do single ones by offering more information (Churchill, 1979; Craig & Douglas, 2000). The researcher has to be aware if the function of the item (Craig & Douglas, 2000) is the same in each culture thus the right election of the item is important thinking for example about formative measurement procedures where the item causes the construct (Diamantopoulos & Winkelhofer, 2001) (for detailed information about formative measures see below).

The main attention on equivalence lies in terms of the measurement itself. Van de Vijver (2003, 2011) as well as Van de Vijver and Leung (1997) classify measurement into equivalence of the structure which relies on the relation between the construct and its items and is related to construct equivalence, the measurement unit, and scalar equivalence. Kumar (2000, 2009) as also Mullen (1995) divide the measurement equivalence into translation equivalence, metric equivalence, and calibration equivalence. The latter two overlap with the ones of the first mentioned authors as metric or unit equivalence assign similar metrics for comparing patterns of the answers given by the interviewees without the same origin while scalar equivalence or calibration equivalence includes this origin, like the same measurement of distances in meter or weights in kilogram.

In terms of translation equivalence questions developed in one language have to be translated adequately to the other cultures under study as language is one of the main representations of culture (Douglas & Craig, 1997). The most common technique in
Establishing scale equivalence in intercultural marketing research studies: A structured literature review

Translation is called back-translation where first the questions are translated and in a second step translated again to the language of origin by another translator to proof if the meaning remains the same (Harkness, 2003). Back-translation is also doubtful as a good translator uses to capture the meaning and translates freely, i.e. a bad translation would be back translated into a good one again (Abukar, 2015). In general, when translating it is useful to put in most background information as possible to overcome possible translation errors and hence the question is still understandable (Brislin, Lonner & Thorndike, 1973; Unsunier, 2011).

Furthermore, translation of scales takes place both linguistically, e.g. semantically, as also structured (concerning metric equivalence) to avoid cultural misunderstanding cause it is not enough if questions used are translated correctly but still perceived as confusing in its structure by the interviewees (Craig & Douglas, 2000; Harkness, 2003). Some cultures are familiar with a scale pointing system of 5 whereas others are familiar with 10 points or even more which consequently has to be considered in adaption of the scales (Craig & Douglas, 2000; Mullen, 1995). Also the direction of writing and reading has to be kept in mind, like Arabic language or Hebrew (Harkness, Van de Vijver & Johnson, 2003). If these factors are not considered by the researcher scores result in biased responding when different scales are applied.

Another cultural factor threatening comparison is that people within one culture tend to respond on scales the same way referring to this as response style (Baumgartner & Steenkamp, 2001). Hui and Triandis (1989) for example studied response styles of Hispanics resulting that they tend to answer towards the extremes of the scale. Harzing (2006) studied the various response styles in different countries focusing on the language used, i.e. a standardized English version implemented in all countries versus the home language. She founds that if English was used answers got better than home language where tendencies
towards the extremes were observed. Other factors are e.g. the reference-group effect studied by Heine, Lehman, Peng and Greenholtz (2002) in subjective likert scales. As different cultures has different reference points, i.e. people refer to people of their own culture, they are not comparable. For comparability they should have the same reference. Peng, Nisbett and Wong (1997) focus on the deprivation effect meaning that people tend to answer for objects that they do not possess. The authors give a solution implemented a so called scenario method where a question is put into context to make it more understandable.

**Administration**

As in the first stages of the research process also in the stage of instrument implementation equivalence as well as biases has to be dealt with. Of high importance is the sampling process asking if the sampling group is adequate for the question to be answered in the research as e.g. the so called decision makers can differ from culture to culture (Kumar, 2000). Also it has to be paid attention to the representativeness of the sampling. In developing countries selected groups might score similar due to their educational level and interaction with other cultures but do not remain the cultural traditions which are of interest when studying the impact on the perception towards an object (Berry, Poortinga, Segall, Dasen, 1992; Berry, Poortinga, Breugelmans, Chasiotis & Sam, 2011). The interviewer effect is disturbing, too (Craig & Douglas, 2000). Social desirability e.g. is one problem which can affect responses (He et al., 2015). Interviewees respond to questions to what they think the interviewer would like to hear. This should be controlled by establishing anonymous questions in instrument design as also keep distance to the respondents which means no face to face interview (Tellis & Chandrasekaran, 2010).
Establishing scale equivalence in intercultural marketing research studies: A structured literature review

Assessing scale equivalence

Once the data has been collected it has to be analysed and interpreted. To make meaningful inference establishment in scale equivalence has to be ensured as presented above. For this purpose there has been developed many techniques to test for equivalence in intercultural studies all belonging to the general latent variable approach (Davidov, Schmidt & Billiet, 2011). These ones are e.g. multi-group structural equation modeling, multilevel analysis, latent class analysis, or item response theory. Here focus is on multi-group confirmatory factor analysis (technique within the multi-group structural equation modeling) as it is the method most used when comparing results from research conducted intercultural (Engelen & Brettel, 2011) as also providing a structural standard for testing for equivalence (Van de Vijver, 2011).

The relationship between the unobservable construct called now latent variable and the item, manifest variable or indicator, is manifested in a measurement model, i.e. the construct is operationalized by the item (Diamantopoulos, Riefler & Roth, 2008). In confirmatory factor analysis a given measurement model is assumed and the researcher observes if the hypothesized structure fits, i.e. is confirmed (see e.g. Bernstein, Garbin & Teng, 1988; Gatignon, 2010; Rencher & Christensen, 2012). It is essential to choose the right measurement model to avoid misleading indications as will be presented above (Jarvis, Mackenzie & Podsakoff, 2003).

Hence, Steenkamp and Baumgartner (1998) as well as Diamantopoulos and Papadopoulos (2010) have established guidelines to assess scale equivalence using confirmatory factor analysis and which will be presented as follows.
Establishing scale equivalence in intercultural marketing research studies: A structured literature review

Reflective measurement model

In this type of measurement model the items are reflective indicators, i.e. they are representing the effect of the latent variable (Diamantopoulos, Riefler & Roth, 2008). In marketing constructs like attitudes or purchase intention are seen as reflective (Jarvis, Mackenzie & Podsakoff, 2003). Steenkamp and Baumgartner (1998) established an assessing guideline for scale equivalence for this type of measurement model. To understand the technique first the reflective measurement model should be introduced expressed in mathematical terms as

\[ x_{ij}^g = \tau_i^g + \lambda_{ij}^g \xi_j^g + \delta_i^g \]  

(1)

with \( x_{ij}^g \) representing the item (i=1,…,p), \( \xi_j^g \) the latent variable (j=1,…,m), \( \lambda_{ij}^g \) the factor loading, i.e. the effect of \( \xi_j^g \) on \( x_{ij}^g \) by increasing its value of one unit, \( \tau_i^g \) is the intercept, and \( \delta_i^g \) the error term. This holds respectively for all cultures (g). Furthermore it is assumed that \( E(\delta_i^g) = 0 \) and \( \text{Cor}(\xi_j^g, \delta_i^g) = 0 \). Figure 2 shows equation (1) in form of a path model with arrows in reflective direction.

![Figure 2. Reflective measurement model (based on Wagner, Wetzels & Winklhofer, 2005)](image_url)
As well as equation (1) also the equations in terms of the means and variance-covariance matrix of item $x_i$ take a role in further steps and should be introduced as well. Given the means equation (2) by

$$\mu^g = \tau_{ij}^g + \lambda_{ij}^g \kappa^g$$

with $\mu$ as a $p \times 1$ vector of the item means and $\kappa$ as a $m \times 1$ vector of the latent means. The variance-covariance matrix of $x$ is represented by equation (3) is

$$\Sigma^g = \lambda_{ij}^g \Phi^g \lambda_{ij}^g' + \Theta^g$$

with $\Phi^g$ as the variance-covariance matrix of $\xi_j^g$ and $\Theta^g$ as the variance-covariance matrix of the error term $\delta_i^g$ respectively to each culture $g$.

Now to assess equivalence in the measurement of scales implies some specifications on the model. Steenkamp and Baumgartner (1998) refer to these specifications as establishment of configural equivalence, metric equivalence, scalar equivalence, factor variance equivalence as well as covariance equivalence, and error variance equivalence. Statistically spoken, with configural equivalence the authors refer to the factor structure as a whole showing equivalent patterns in each culture. That means the perception of a construct is similar in the cultures under study. Errors in response styles as mentioned above in contrast can show differences in the structure (Hui & Triandis, 1989).

Metric equivalence points on equality of factor loadings as they determine the metric by the change of one unit in $\xi$ by $\tau$ units in $x$. So the following has to be established in the models across the cultures

$$\lambda_{ij}^1 = \lambda_{ij}^2 = \lambda_{ij}^3 = \ldots = \lambda_{ij}^g$$

This could be interpreted also to the structure but not only the same pattern of factors than also the same answering structure (see also Mullen, 1995). At last, scalar equivalence then determines whether means can be compared or not represented by an introduced equality of the intercepts.
Establishing scale equivalence in intercultural marketing research studies: 
A structured literature review

\[ \tau_{ij}^{1} = \tau_{ij}^{2} = \tau_{ij}^{3} = \ldots = \tau_{ij}^{g} \]  

(5)

If equal intercepts are established scores can be compared (Meuleman & Billiet, 2011).

In addition, factor variance as well as covariance moreover ensures the comparison of correlations between any other latent variables given the factor variance equivalence

\[ \phi_{jk}^{1} = \phi_{jk}^{2} = \phi_{jk}^{3} = \ldots = \phi_{jk}^{g} \]  

(6)

with (j=1, …, m and k=1, …, [j-1]) and factor covariance equivalence with (j=1, …, m)

\[ \phi_{jj}^{1} = \phi_{jj}^{2} = \phi_{jj}^{3} = \ldots = \phi_{jj}^{g} \]  

(7)

Testing of measurement error across cultures as a last procedure offers assessment of equivalence in terms of the extent of biased measurement introducing into the model

\[ \Theta^{1} = \Theta^{2} = \Theta^{3} = \ldots = \Theta^{g} \]  

(8)

Steenkamp and Baumgartner (1998) note that the representativeness of the sampling as above mentioned is in advance assumed.

The proper assessment procedure then starts with the identification of the model. This might be done by two alternative options. First a so called marker item might be detected and its intercept \( \tau_{ij}^{g} \) fixed to zero which applies for the same item in all the cultures under study (i.e. \( \tau_{m}^{g} = 0 \); with m as index for marker item). The second way is to choose one cultural group as a reference group and fix the vector of latent means \( \kappa \) in equation (2) to zero and in each culture fix one intercept per factor to be invariant, i.e. for items with invariant factor loadings.

The next step then is to test for equivalence constraining the measurement model to the conditions above explained stepwise for metric equivalence, scalar equivalence, etc.

Before starting the actual procedure advice for one prior test for the equality of covariance matrices and mean vectors is given. If equivalence holds, but the authors mentioned rarely it is the case, the collected data can be pooled without testing for each culture individually. If not then the models in each culture have to be tested for configural
equivalence first, then metric equivalence and scalar equivalence. Testing for the variance and covariance equivalences as also for error variance do not have to maintain the order then. Assessing the complete equivalence of the collected data, meaning full equivalence, is rather utopian so that the term of partial measurement equivalence is introduced (see also Byrne, Shavelson & Muthén, 1989). Meaning that if on each stage full equivalence is not hold then testing for partial equivalence has to be done first before going on to test for the next equivalence level. For testing on model fit as usually done in confirmatory factor analysis Steenkamp and Baumgartner (1998) decided to use beside the $\chi^2$-difference test other fit indices like the root mean square error of approximation (RMSEA), consistent Akaike information criterion (CAIC), Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI). Literature advices for a good model fit to use values of e.g. .08 for RMSEA, for TLI .9 or even .95, or .95 for CFI (Hu & Bentler, 1999; Vandenberg & Lance, 2000). Parameters below these values are freed from constraints and focus is only on the remained parameters until the model fits well. On the next equivalence level then focus is kept only on these parameters as well. The authors remember that such a model specification of partial equivalence has to be prepared with caution as to avoid deformation of the model. In general, Steenkamp and Baumgartner (1998) speak about a minimum condition for still talking about comparison which contains beside the marker item at least one more item that has to demonstrate same factor loading and same intercepts.

**Formative measurement model**

Often disregarded in research is the type of formative measurement where causality is different as in reflective measure from the item to the latent variable (Jarvis, Mackenzie & Podsakoff, 2003). The choice of the right measurement model is from high importance as misspecification can lead to e.g. wrong parameter estimates or wrong model fit
Establishing scale equivalence in intercultural marketing research studies: A structured literature review

Constructs just like complaint behaviour are formative (Jarvis, Mackenzie & Podsakoff, 2003).

Where formative measurement models then are more appropriate there is also call for assessing the established equivalence. Diamantopoulos and Papadopoulos (2010) assumed the situation and provided guidelines for testing equivalence in formative measurement models (see also Wagner, Wetzels & Winklhofer, 2005). Thereby the measurement model for the formative case is presented as

\[ \eta_g = \gamma_{g1} x_1 + \gamma_{g2} x_2 + \ldots + \gamma_{gi} x_{gi} + \zeta_g \]  

(9)

with \( \eta \) as the latent variable, \( x_i \) the formative item (i=1, 2, …, n), \( \gamma_i \) as a coefficient representing the amount of cause from \( x_i \) to \( \eta \) and \( \zeta \) as a disturbance term, all within a culture \( g \) (\( g = 1, 2, \ldots, m \)). Additionally assumptions of \( E(\zeta_g) = 0 \) and \( \text{Cov}(x_{gi}, \zeta_g) = 0 \) are made. Figure 3 shows the path model for equation (9).

![Figure 3. Formative measurement model (based on Wagner, Wetzels & Winklhofer, 2005)](image)

Also constraints for assumed equivalence are made in the formative model but with a different approach as other circumstances are given like for example exogenous items. These constraints are named structure equivalence, slope equivalence, and residual equivalence.
Establishing scale equivalence in intercultural marketing research studies:
A structured literature review

Structure equivalence should present the same pattern of non-zero coefficients which then implies the same perception of the construct. Slope equivalence aims at the item’s reference to the construct constraining their influence \((\gamma_{gi})\) on the construct to be equal in each culture

\[
\gamma_{i1} = \gamma_{i2} = \gamma_{i3} = \ldots = \gamma_{gi}
\]  

(10)

and finally constraining as well the variance of the disturbance term as equal to obtain residual equivalence

\[
\psi_1 = \psi_2 = \psi_3 = \ldots = \psi_g
\]  

(11)

Given these assumptions then the model has to be identified as well given some problems as by the present formative conditions the model is underidentified. For purposes of identification the solution offered by Diamantopoulos and Papadopoulos (2010) is to introduce two reflective indicators and therefore convert the current model into a so called MIMIC model (multiple indicators – multiple causes) as shown in figure 4.

![Figure 4. MIMIC model (based on Wagner, Wetzel & Winklhofer, 2005)](image)

The new introduced indicators in figure 4 expressed in equation notation are

\[
y_1 = \lambda_{g1} \eta_g + \varepsilon_{g1}
\]  

(12)
with assumption of \( E(\varepsilon_{g1}) = E(\varepsilon_{g2}) = 0 \) and \( \text{Cov}(\varepsilon_{g1};\varepsilon_{g2}) = 0 \) as well as the other conditions valid for the reflective model in each culture. Furthermore as two indicators are still underidentified a third item is introduced that serves as a marker item and whose loading is fixed to one in each culture. Subsequently the reflective indicators have to show at least metric equivalence first to go ahead with the testing procedure. If no equivalence of the two indicators is hold then there is the possibility to change them for other items which do so.

The next steps then are equal as for the reflective measurement model, testing stepwise for each equivalence assumption including test for partial equivalence. The fit indices used by Diamantopoulos and Papadopoulos (2010) are \( \chi^2 \)-difference test, RMSEA, CFI, as also non-normed fit index (NNFI), and standardized root mean square residual (SRMR) with values of .08 for SRMR and for NNFI from .9 to .95 (Hu & Bentler, 1999; Vandenberg & Lance, 2000).

**Method**

The present work aims to review if the guidelines offered by Steenkamp and Baumgartner (1998) as well as by Diamantopoulos and Papadopoulos (2010) and presented above are applied in practice and reported for in current marketing publications. In other words, one might infer: where Vandenberg and Lance (2000) conduct a literature review searching for academic disciplines carrying about the establishment of guidelines concerning the application of confirmatory factor analysis in group comparison and Engelen and Brettel (2006) reviewing how many marketing publications rely on cross-cultural issues this work then brings the two approaches together looking for practice in current intercultural marketing studies.
The method proposed therefore is conducting a structured literature review as a tool providing an optimal examination within an academic discipline (Seuring & Gold, 2015) based on the operation of Zulauf et al. (2015) including content analysis and in addition a citation analysis. The focus is on four peer-reviewed academic journals, the Journal of Marketing, the Journal of International Marketing, the Journal of Consumer Research, and the Journal of International Business Studies. Firsts are chosen by general impact in marketing research field as well as international marketing as the topic of the work at hand while in the latter ones the presented guidelines were published. Moreover, because focus is on these two guidelines the time frame is considered from 2000 respectively 2012 till today. Beside the general observation of considered equivalence aspects just like the observed cultures and the cultural background of the authors or even the right application of equivalence assessment (i.e. equivalence level) due to the respective research goal as mentioned by Steenkamp and Baumgartner (1998) could be kept in mind. Further specifications are still in process.

**Structure**

1. Introduction

2. Intercultural marketing research
   
   2.1 The research process
   
   2.2 Equivalences and biases of scales
      
      2.2.1 The construct
      
      2.2.2 Instrument design
      
      2.2.3 Administration
   
   2.3 Assessing scale equivalence
2.3.1 Reflective measurement model

2.3.2 Formative measurement model

3. A structured Literature review

3.1 Method

3.2 Results

3.2.1 Content analysis

3.2.2 Citation analysis

4. Conclusion

Working plan

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 2016</td>
<td>- Improvement of the theoretical part (searching for more literature +</td>
</tr>
<tr>
<td></td>
<td>more detailed description)</td>
</tr>
<tr>
<td></td>
<td>- Begin of part 2: literature review (criteria for election + election of</td>
</tr>
<tr>
<td></td>
<td>relevant literature)</td>
</tr>
<tr>
<td>October 2016</td>
<td>Content analysis</td>
</tr>
<tr>
<td>November 2016</td>
<td>- Citation analysis</td>
</tr>
<tr>
<td></td>
<td>- Improvement of the theoretical part</td>
</tr>
<tr>
<td>December 2016</td>
<td>- Conclusion</td>
</tr>
<tr>
<td></td>
<td>- first correction</td>
</tr>
<tr>
<td>January 2017</td>
<td>- Final correction</td>
</tr>
<tr>
<td></td>
<td>- Submission</td>
</tr>
</tbody>
</table>
Establishing scale equivalence in intercultural marketing research studies: A structured literature review

References


Establishing scale equivalence in intercultural marketing research studies: A structured literature review


Establishing scale equivalence in intercultural marketing research studies: A structured literature review


Establishing scale equivalence in intercultural marketing research studies: A structured literature review


